

TESIS DOCTORAL

En esta tesis se plantea un objetivo general para la obtención de un modelo eficiente aplicable a la resolución de problemas relacionados con la calidad del aire respirable, consistente en la elaboración de una herramienta de utilidad en la toma de decisiones respecto a los problemas derivados de la contaminación atmosférica. Para lograr este fin, se emplean técnicas de análisis basadas en el aprendizaje estadístico aplicadas a un conjunto de datos experimentales, específicamente óxidos de nitrógeno, monóxido de carbono, dióxido de azufre, ozono y materia particulada, registrados entre enero de 2013 y diciembre de 2015 por tres estaciones de monitoreo automatizadas, dependientes de la red de calidad de aire del Gobierno del Principado de Asturias, distribuidas por la ciudad de Oviedo. Dicha área metropolitana, es el centro administrativo de la comunidad autónoma de Asturias y tiene una densidad de población de 1185,12 habitantes por kilómetro cuadrado; emplazándose a una altitud de 232 m sobre el nivel del mar, relativamente cerca de la costa, en el norte de España.

Modelización y análisis de la calidad del aire en la ciudad de Oviedo (norte de España), mediante los enfoques PSO-SVM, red neuronal MLP y árbol de regresión M5



universidad
de león

PROGRAMA DE DOCTORADO EN INGENIERÍA DE PRODUCCIÓN
Y COMPUTACIÓN POR LA UNIVERSIDAD DE LEÓN

Código del programa: 309

TESIS DOCTORAL

**MODELIZACIÓN Y ANÁLISIS DE LA CALIDAD
DEL AIRE EN LA CIUDAD DE OVIEDO
(NORTE DE ESPAÑA), MEDIANTE LOS
ENFOQUES PSO-SVM, RED NEURONAL MLP
Y ÁRBOL DE REGRESIÓN M5**

Doctorando,

Alejandro Aurelio Rodríguez Miranda

León, 2018



TESIS DOCTORAL

Modelización y análisis de la calidad del aire en la ciudad de Oviedo (norte de España), mediante los enfoques PSO-SVM, red neuronal MLP y árbol de regresión M5

Autor:

Alejandro Aurelio Rodríguez Miranda ^a

Directores:

Dr. Antonio Bernardo Sánchez ^a

Dr. Paulino José García Nieto ^b

Tutor:

Dr. Antonio Bernardo Sánchez ^a

^a Departamento de Tecnología Minera, Topográfica y Estructuras,
Universidad de León, 24071 León, España

^b Departamento de Matemáticas, Facultad de Ciencias,
Universidad de Oviedo, 33007 Oviedo, España

Universidad de León

León, 2018

ANTONIO BERNARDO SÁNCHEZ, Profesor Titular del Departamento de Tecnología Minera, Topográfica y Estructuras de la Universidad de León, y PAULINO JOSÉ GARCÍA NIETO, Catedrático de Universidad del Departamento de Matemáticas de la Facultad de Ciencias de la Universidad de Oviedo,

CERTIFICAN que el presente trabajo de investigación titulado:

“MODELIZACIÓN Y ANÁLISIS DE LA CALIDAD DEL AIRE EN LA CIUDAD DE OVIEDO (NORTE DE ESPAÑA) MEDIANTE LOS ENFOQUES PSO-SVM, RED NEURONAL MLP Y ÁRBOLES DE REGRESIÓN M5”,

ha sido realizado por Alejandro Aurelio Rodríguez Miranda, que a la presentación de este trabajo se encuentra matriculado en el PROGRAMA DE DOCTORADO EN INGENIERÍA DE PRODUCCION Y COMPUTACIÓN POR LA UNIVERSIDAD DE LEÓN,

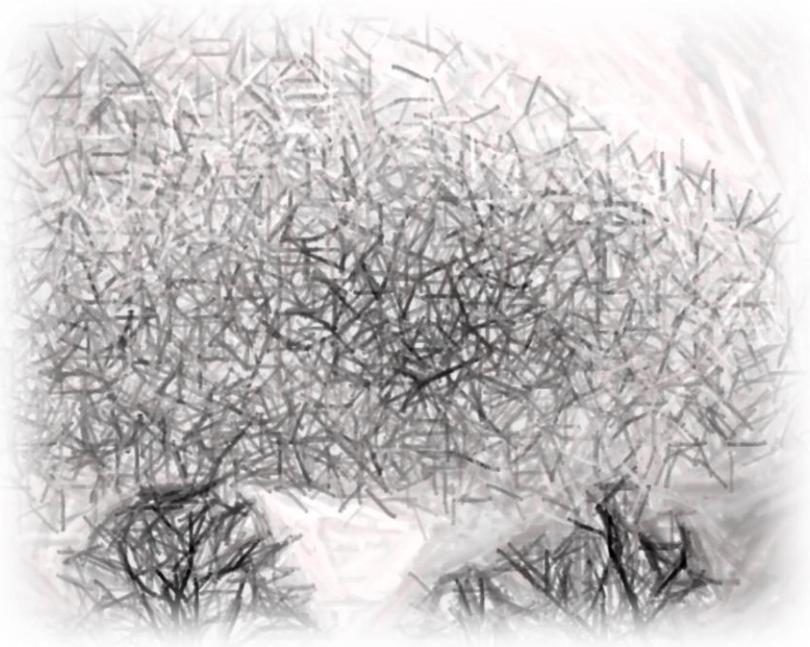
y como directores del mismo, AUTORIZAN SU PRESENTACIÓN como Tesis Doctoral para optar al grado de Doctor y a fin de que pueda ser juzgada por el tribunal correspondiente.

y para que conste,

se expide y firma la presente en León, a fecha 19 de enero de 2018.

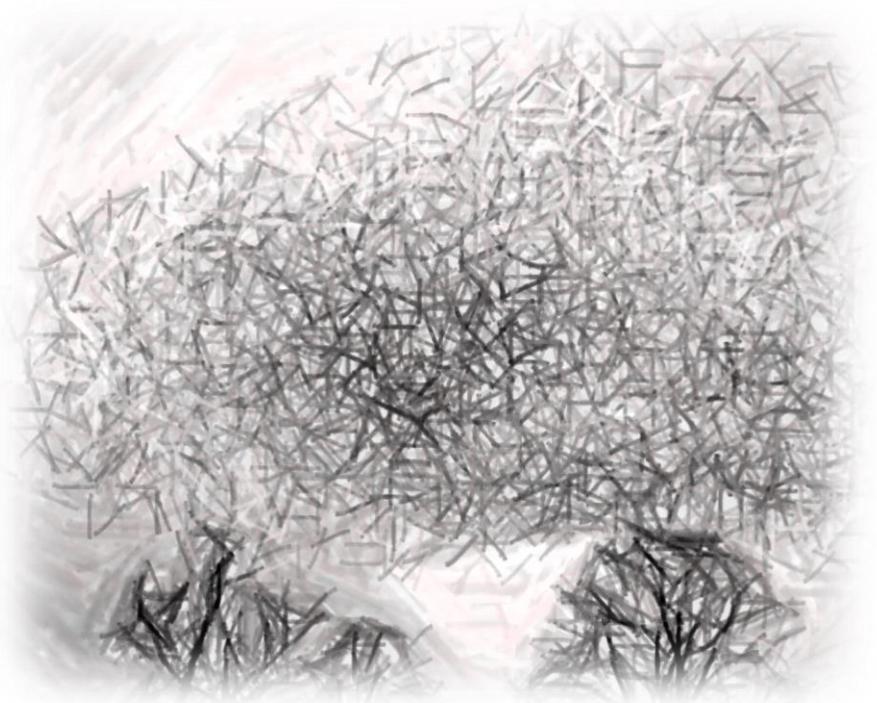
Fdo. Dr. Antonio Bernardo Sánchez

Fdo. Dr. Paulino José García Nieto



*A mis padres, a mi hermana,
a mi hija Alba y a mi esposa Dolores.*





*“El éxito no es definitivo, el fracaso no es fatídico.
Lo que cuenta es el valor para continuar.”*

(Winston Churchill)



AGRADECIMIENTOS

La vida es aquello que te va sucediendo mientras te empeñas en hacer otros planes, decía John Lennon. Siempre he pensado que se llega a algún punto por la casualidad de cruzarse con personas que, siendo o no conscientes de ello, influyen extraordinariamente en los caminos que un individuo elige. En una etapa de mi vida, en que desarrollaba una actividad que no guardaba relación con la actual, tuve la fortuna de encontrarme con una persona que influyó en mi vida, hasta el punto de considerar retomar los estudios que había abandonado y como consecuencia volver a empezar. Muchísimas gracias F. Fdez., aunque “no me escuches” debo decir que aún recuerdo tus consejos. Después de aquel suceso, fueron muchas puertas las que se me abrieron hasta llegar hoy aquí. He tenido la oportunidad de conocer a personas increíbles y de hacer muy buenos amigos durante estos años.

A vosotros Paulino y Antonio, mis directores de Tesis, tengo mucho que agradecerlos, por darme la oportunidad de participar en este trabajo de investigación (gracias por esta experiencia, ¡tremenda!, por enorme y sin duda valiosa para mí). Muchas gracias por vuestros consejos y por vuestra valiosísima ayuda en la elaboración de este documento. Debo agradecer también la ayuda prestada a los compañeros, que me proporcionaron los medios y los recursos necesarios para elaborar este trabajo. A los amigos, que en algún momento escucharon mis elucubraciones y “soportaron” teorías que no entraban seguramente dentro de sus temas favoritos de debate. A todos aquellos a quienes, seguramente, no presté toda la atención que debía; por estar enfascado en exceso en estos y otros temas. Esta tesis no habría podido finalizarse sin la paciencia y la colaboración de todas las personas que han arrimado el hombro y animado cuando más lo necesitaba, mostrándome su cariño incondicional. Por último, agradezco especialmente a mi familia, mi hija Alba y mi esposa Dolores, que son conocedoras de la evolución de mi vida y del esfuerzo realizado; sin su valioso apoyo, de ningún modo habría podido hacerlo.

RESUMEN

En la segunda edición del informe de la Organización Mundial de la Salud (OMS) de 2016, *Ambientes saludables y prevención de enfermedades: Hacia una estimación de la carga de morbilidad atribuible al medio ambiente*, se revela que desde la primera edición, publicada hace una década, las muertes por enfermedades que pueden atribuirse a la contaminación del aire (incluida la exposición al humo ajeno) han aumentado significativamente. Los problemas que se derivan de la contaminación atmosférica son conocidos, por las noticias que cada día hablan de este problema. Así, es entendible el interés creciente por el uso de modelos matemáticos, para modelizar el comportamiento de los contaminantes en la atmósfera, con el fin de mejorar la calidad del aire.

En esta tesis de investigación se elabora una herramienta, que puede ser empleada en la toma de decisiones respecto a los problemas derivados de la contaminación atmosférica. Planteando un objetivo general que pretende, ante la necesidad existente, lograr un modelo aplicable a la resolución de problemas relacionados con la calidad del aire respirable (a escala local), utilizando técnicas basadas en el aprendizaje estadístico, para el área metropolitana de Oviedo, en el norte de España. Para ello, y a partir del conjunto de datos experimentales registrados, específicamente óxidos de nitrógeno, monóxido de carbono, dióxido de azufre, ozono y materia particulada, recogidos desde 2013 a 2015 en el área urbana de Oviedo, se construyen varios modelos de regresión utilizando técnicas basadas en el aprendizaje estadístico, para predecir mediante el modelo resultante la calidad del aire en un espacio determinado. Una vez obtenidos los modelos planteados, se comparan los resultados con el fin de comprobar que el método híbrido basado en el enfoque PSO-SVM con una función de núcleo o ‘kernel’ RBF arroja mejores resultados, pudiendo modelar con precisión suficiente la calidad del aire a escala local, mejorando significativamente la capacidad de generalización alcanzable sólo con el regresor basado en SVM.

ABSTRACT

In the second edition of World Health Organization (WHO) report of 2016, “Healthy environments and disease prevention: towards an estimate of the burden of morbidity attributable to the environment”, it is revealed that since the first edition, published a decade ago, deaths from diseases that can be attributed to air pollution (including exposure to smoke from external sources) have increased significantly. The problems due to air pollution are well known, by all the news spread every day about this problem. Thus, the growing interest in the use of mathematical models is understandable, to model the behavior of pollutants in the atmosphere, in order to improve air quality.

In this research thesis a tool is developed, which can be used to make decisions in order to solve some of the air pollution problems. In this sense, a general objective is formulated to achieve a model applicable to the resolution of problems related to the quality of breathable air (at a local scale), using techniques based on statistical learning, for the metropolitan area of Oviedo, in the north of Spain. Thus, taken a experimental data set, specifically oxides of nitrogen, carbon monoxide, sulfur dioxide, ozone and particulate matter, stored from 2013 to 2015 in the urban area of Oviedo, several regression models are constructed using techniques based in statistical learning, to foretell through the resulting model the quality of air in a determined space. Once the proposed models have been obtained, the results are compared in order to verify that the hybrid method based on the PSO-SVM approach, with a kernel RBF, yields better results, being able to model with sufficient precision the air quality at local scale, significantly improving the generalization capacity achievable with the SVM-based regressor only.

ÍNDICE GENERAL

Índice de capítulos	iii
Índice de figuras.....	vi
Índice de tablas.....	x
Referencias bibliográficas	295

Índice de capítulos

Capítulo I.- Introducción	1
<i>1. Problema de investigación y objetivos de esta tesis. Noticias en prensa, marco legal y normativa.....</i>	<i>3</i>
<i>1.1. Introducción</i>	<i>3</i>
<i>1.1.1. Problema de investigación</i>	<i>5</i>
<i>1.1.2. Objetivos</i>	<i>5</i>
<i>1.1.3. Organización de la tesis</i>	<i>8</i>
<i>1.1.4. El alcance de este trabajo.....</i>	<i>11</i>
<i>1.2. Contextualización y motivación.....</i>	<i>11</i>
<i>1.2.1. Noticias publicadas en relación con la contaminación ambiental en Asturias durante el primer semestre de 2017.</i>	<i>14</i>
<i>1.3. Normativa aplicable en materia de contaminación atmosférica.....</i>	<i>26</i>
<i>1.3.1. Normativa comunitaria.....</i>	<i>26</i>
<i>1.3.2. Normativa estatal.....</i>	<i>26</i>
<i>1.3.3. Normativa autonómica</i>	<i>27</i>
Capítulo II.- Fundamentos	29
<i>2. Aprendizaje automático para el análisis de datos.....</i>	<i>31</i>
<i>2.1. Fundamentos biológicos del cerebro y transmisión de la información a través del sistema nervioso</i>	<i>31</i>
<i>2.1.1. El Cerebro</i>	<i>31</i>
<i>2.1.2. El Sistema Nervioso</i>	<i>33</i>
<i>2.2. Inteligencia Artificial (IA)</i>	<i>35</i>
<i>2.2.1. Las neuronas</i>	<i>35</i>
<i>2.2.2. Arquitectura y funcionamiento de las redes neuronales artificiales</i>	<i>37</i>
<i>2.2.3. Inteligencia artificial compleja, mediante el desarrollo 'Software' que imite el funcionamiento del cerebro</i>	<i>38</i>
<i>2.2.4. La transformación del aprendizaje profundo</i>	<i>39</i>
<i>2.3. El aprendizaje automático.....</i>	<i>43</i>
<i>2.3.1. Técnicas de Aprendizaje</i>	<i>43</i>
<i>2.3.2. Aprendizaje mediante técnicas inductivas</i>	<i>44</i>
<i>2.4. Redes neuronales. El perceptrón.....</i>	<i>47</i>

2.4.1. Arquitectura y funcionamiento del Perceptrón.....	49
2.4.2. Perceptrón simple y Adaline.....	55
2.4.3. Redes neuronales multicapa.....	58
2.5. KDD y Minería de Datos.....	65
2.5.1. KDD – ‘Descubrimiento de Conocimiento a partir de Bases de Datos’.....	65
2.5.2. Minería de Datos	66
Capítulo III.- Contaminantes del aire	71
3. Fuentes y tipos de contaminación atmosférica	73
3.1. Medioambiente y atmósfera.....	73
3.1.1. Estructura y contaminación de la atmósfera	74
3.2. Principales contaminantes del aire	75
3.2.1. Contaminantes primarios y secundarios	78
3.3. Características principales, parámetros y efectos más relevantes de los contaminantes primarios y secundarios	81
3.3.1. Material particulado (PM)	82
3.3.2. Monóxido de carbono (CO).....	89
3.3.3. Óxidos de nitrógeno (NO, NO ₂ , NO _x)	94
3.3.4. Óxidos de azufre (SO _x)	97
3.3.5. PAN o nitrato de peroxyacilo.....	99
3.3.6. Ozono (O ₃).....	102
3.3.7. Smog fotoquímico.....	110
3.3.8. Lluvia ácida	118
3.3.9. Contaminantes ‘Criterio’	127
Capítulo IV.- Materiales y métodos	137
4. Modelos de clasificación, regresión y optimización	139
4.1. Máquinas de vectores de Soporte (SVM)	139
4.1.1. Funcionamiento de las SVMs.....	141
4.1.2. Caso no lineal	146
4.1.3. La SVM de margen suave	148
4.1.4. Máquinas de Soporte Vectorial para Regresión	150
4.1.5. Truco del Kernel y tipos	159
4.1.6. Estimación de Parámetros del SVR, Validación Cruzada.....	165

<i>4.2. Árboles de decisión</i>	167
4.2.1. Complejidad o altura en árboles.	168
4.2.2. Sobrepredicción	171
4.2.3. Árboles de Clasificación y Regresión (ACR).....	173
4.2.4. Árboles Modelo M5.....	181
4.2.5. Árbol modelo M5'	189
4.2.6. M5' Rules.....	191
<i>4.3. Técnicas de Optimización.....</i>	201
4.3.1. Clasificación de las técnicas de optimización	202
4.3.2. Optimización por Enjambre de Partículas (PSO)	208
4.3.3. PSO completamente informado.....	218
<i>4.4. Método de optimización de SVM mediante PSO</i>	220
4.4.1. Enfoque basado en el PSO-SVM	221
Capítulo V.- Estado del Arte	223
<i>5. Hallazgos documentales y revisión</i>	225
<i>5.1. Evolución de la técnica.....</i>	225
<i>5.2. Situación actual del arte y revisión documental</i>	226
5.2.1. Hallazgos documentales de interés	227
<i>5.3. Origen de la investigación.....</i>	231
Capítulo VI.- Aplicación al caso de estudio	233
<i>6. Modelado y análisis de la calidad del aire en la ciudad de Oviedo (norte de España), mediante los enfoques PSO-SVM, red neuronal MLP y árbol de regresión M5. Caso de estudio.....</i>	235
<i>6.1. Planteamiento del caso de estudio.....</i>	235
<i>6.2. Introducción.....</i>	236
<i>6.3. Marco teórico y metodología del experimento.....</i>	240
6.3.1. Fuentes y tipos de contaminación del aire, en el área de localización del problema.....	240
6.3.2. Datos experimentales	242
6.3.3. Método de las máquinas de soporte vectorial (SVM)	248
6.3.4. El algoritmo de optimización de la nube de partículas (PSO).....	252
6.3.5. Método de la Red neuronal artificial MLP	255

6.3.6. Método del árbol de regresión M5	256
<i>6.4. Análisis de resultados y discusión</i>	<i>257</i>
<i>6.5. Conclusiones</i>	<i>275</i>
Capítulo VII.- Aportación y experiencias	279
<i>7. Difusión de esta tesis, contribución y posibles líneas de investigación.....</i>	<i>281</i>
<i>7.1. Contribuciones de esta tesis</i>	<i>281</i>
<i>7.1.1. Publicación en revista científica</i>	<i>282</i>
<i>7.1.2. Participación en la IX Conferencia Internacional de Aprovechamiento de Recursos Minerales CINAREM'17</i>	<i>284</i>
<i>7.2. Posibles líneas de investigación.....</i>	<i>287</i>
<i>7.2.1 Mejorar la ubicación de los ‘Árboles urbanos’ inteligentes para reducir la contaminación del aire.....</i>	<i>287</i>
<i>7.2.2. Mejorar la precisión del cálculo los niveles de emisiones contaminantes en las ciudades, aplicando Técnicas de minería de datos al Big Data obtenido de una red de dispositivos móviles.....</i>	<i>288</i>
<i>7.2.3. Generar modelos para estimar la contaminación del ambiente que nos rodea, mediante aplicación móvil (app)</i>	<i>291</i>
<i>7.2.4. Minimizar los posibles errores entre usuarios que comparten datos sobre el nivel de contaminación del aire respirable.....</i>	<i>292</i>

Índice de figuras

Figura 1.1. Emplazamiento de las tres estaciones de medición de la contaminación en Oviedo.....	3
Figura 1.2. Zonificación de la calidad del aire de España 2011. Zonas de Asturias.....	4
Figura 1.3. Emisiones de PM10 ($\mu\text{g}/\text{m}^3$).	14
Figura 1.4. a) Situación de las estaciones en niveles naranja y rojo. b) Evolución de los niveles de benceno en Trubia en los días de la noticia..	17
Figura 1.5. VISTA DE LA CONTAMINACIÓN EN OVIEDO.....	18
Figura 1.6. Concentración de partículas de dióxido de nitrógeno en España.....	25
Figura 2.1. Áreas funcionales del cerebro..	33

Figura 2.2. Principales divisiones del sistema nervioso central y Periférico.....	34
Figura 2. 3. Diagrama básico de una neurona.....	35
Figura 2.4. Partes de una neurona.....	36
Figura 2.5. Activación de un perceptrón o neurona artificial.....	39
Figura 2.6. Red neuronal artificial de tipo perceptrón multicapa.....	40
Figura 2.7. Red neuronal profunda, con tres capas ocultas.....	41
Figura 2.8. Diagrama de un perceptrón con cinco señales de entrada	49
Figura 2.9. Red de Perceptrón de una sola capa. Fuente: adaptada de Neural Network Design	50
Figura 2.10. Perceptrón de dos entradas y una sola salida. Fuente: adaptada de Neural Network Design. [Martin T. Hagan, Howard B, Demuth, Mark Beale – PWS Publishing Company].....	52
Figura 2.11. Límite de decisión para un Perceptrón de dos entradas. Fuente: Neural Network Design. [Martin T. Hagan, Howard B, Demuth, Mark Beale – PWS Publishing Company]	53
Figura 2.12. a) pares de entrada/destino para la puerta AND. b) elección de la línea que separa las dos categorías de entradas. Fuente: Neural Network Design. [Martin T. Hagan, Howard B, Demuth,Mark Beale – PWS Publishing Company].....	54
Figura 2.13. Estructura de la red neuronal artificial Adaline.....	56
Figura 2.14. Gráfica de salida para la función de umbral. Elaboración.....	57
Figura 2.15. Diferencias entre el Perceptrón simple y Adaline. Fuente: S. Raschka, Quantitative Biology and Biochemistry and Molecular Biology Ph.D.- Michigan State University.....	57
Figura 2.16. Funciones de activación sigmoidal empleadas en redes neuronales artificiales MLP. Fuente: Redes neuronales, José M ^a Valls, Universidad Carlos III de Madrid, 2007	59
Figura 2.17. Red de perceptrón multicapa (MLP), con una capa oculta.....	60
Figura 2.18. Propagación de los valores de entrada hacia delante, en la capa oculta.....	62
Figura 2.19. Función de paso.....	63
Figura 2.20. Esquema de función para cambiar gradualmente los pesos.....	63
Figura 2.21. Función sigmoidea..	64
Figura 2.22. Proceso de KDD (Knowledge Discovery from Databases).	66
Figura 3.1. Las cinco esferas del ambiente	73
Figura 3.2. Estructura vertical de la atmósfera y su equivalente adaptada a temperatura en °C.....	75
Figura 3.3. Composición de la atmósfera terrestre (tomada en diciembre de 1987).....	76
Figura 3.4. Distintas manifestaciones de los contaminantes: a) Paticulas sólidas, b) gotas de líquido, c) gases.....	76
Figura 3.5. Orígenes y tipos de contaminantes primarios	78

Figura 3.6. Distribuciones modales de tamaños de partículas.	84
Figura 3.7. Comparación entre los distintos tamaños de material particulado.	86
Figura 3.8. Capacidad de penetración de las partículas en el organismo, en función de su tamaño.	87
Figura 3.9. PM2.5, o carbón negro, o black carbon.	88
Figura 3.10. a) Emisiones PM2.5 en la UE-28: participación por sector en 2014.	89
Figura 3.11. a) Representación en 3D de una molécula de monóxido de carbono. b) Estructura de la molécula de monóxido de carbono.	90
Figura 3.12. Unión de CO a los hematíes.	92
Figura 3.13. a) Representación en 3D de una molécula de monóxido de nitrógeno	94
Figura 3.14. a) Representación en 3D de una molécula de dióxido de nitrógeno	96
Figura 3.15. a) Representación en 3D de una molécula de dióxido de azufre. b) Estructura de la molécula del dióxido de azufre.....	98
Figura 3.16.a) Representación en 3D de una molécula de PAN. b) Estructura de la molécula de PAN.	100
Figura 3.17. a) Representación 3D de una molécula de ozono b) Estructura de la molécula de ozono.....	102
Figura 3.18. Ozono (O ₃) en función de la capa terrestre.	103
Figura 3.19. Formación del Ozono Troposférico (O ₃) en las capas bajas de la atmósfera.....	104
Figura 3.20. Concentraciones de O ₃ en relación con el valor objetivo de 2014 en la UE-28.	109
Figura 3.21. Formación de “smog” fotoquímico.	111
Figura 3.22. Una ciudad contaminada.....	111
Figura 3.23. Formación de smog fotoquímico.	113
Figura 3.24. Flujo en día normal y día con inversión térmica.....	117
Figura 3.25. Origen y efectos de la lluvia ácida.	120
Figura 3.26. Escala del pH.....	122
Figura 3.27. Tipos de deposiciones	123
Figura 4.1. Plano de separación en un caso linealmente separable.	139
Figura 4.2. La imagen a) muestra los posibles hiperplanos de separación. En la imagen b) se muestra el hiperplano de separación óptimo.....	141
Figura 4.3. Conjunto de puntos linealmente separable.	142
Figura 4.4. Conjunto de función de clasificación en el caso de datos linealmente separables (izquierda) y caso linealmente no separable (derecha).	144
Figura 4.5. Conjunto de puntos linealmente separable.	145
Figura 4.6. Transformación de los datos de entrada a un espacio de mayor dimensión.....	147

Figura 4.7. Variables de Holgura ξ	148
Figura 4.8. Funciones de pérdida.....	151
Figura 4.9. Regresión lineal con error máximo ϵ	153
Figura 4.10. Función de pérdida cuadrática para SVR lineal.....	156
Figura 4.11. SRV-no lineal	160
Figura 4.12. Árboles de decisión y su analogía con los árboles vegetales.	167
Figura 4.13. Árboles con todas las hojas en el menor nivel.....	169
Figura 4.14 . Árboles incompletos de nivel 2.	169
Figura 4.15. Árboles de clasificación y regresión (ACR).....	174
Figura 4.16. Sucesos que podrían modelarse mediante un árbol de decisión.....	175
Figura 4.17. Varias funciones en un intervalo.	175
Figura 4.18. a) Ejemplo árbol y b) Representación de las distintas clases.....	178
Figura 4.19. Ejecución del árbol de modelos M5 en <i>Servo dataset (from UCI)</i>	185
Figura 4.20. Árbol modelo y reglas para el conjunto de datos de pernos.....	192
Figura 4.21. Clasificación de las técnicas de optimización.	202
Figura 4.22. Comportamiento para el aprendizaje de los elementos o partículas que constituyen el enjambre.....	209
Figura 4.23. Dirección y fuerza que empuja a la partícula.	210
Figura 4.24. Direcciones y fuerzas de empuje de varias partículas.	211
Figura 4.25. Cambio de velocidad y posición de la partícula debido a la realimentación del vecindario.	212
Figura 4.26. Visualización del efecto del peso de inercia en el PSO.....	217
Figura 6.1. Ubicación geográfica de las estaciones meteorológicas en la zona urbana de Oviedo y de la central eléctrica de Soto de Ribera	237
Figura 6.2. Tendencia mensual de las concentraciones de dióxido de nitrógeno (NO ₂), dióxido de azufre (SO ₂) y monóxido de carbono (CO) durante los años 2013, 2014 y 2015 en el área urbana de Oviedo.....	244
Figura 6.3. Tendencia mensual de las concentraciones de partículas (PM10), ozono (O ₃) y óxido nítrico (NO) durante los años 2013, 2014 y 2015, en el área urbana de Oviedo.	246
Figura 6.4. Regresión con tubo insensible para problemas unidimensionales.....	249
Figura 6.5. Diagrama de una red MLP con h neuronas en la capa oculta de neuronas en la capa de entrada y una sola neurona en la capa de salida.	255
Figura 6.6. Diagrama de flujo del nuevo modelo híbrido PSO-RBF-SVM	259
Figura 6.7. Importancia relativa de las variables de entrada para predecir el valor del dióxido de nitrógeno (NO ₂) en el área urbana de Oviedo en el modelo basado en PSO-RBF-SVM.....	263
Figura 6.8. Importancia relativa de las variables de entrada para predecir el valor del dióxido de azufre (SO ₂) en la zona urbana de Oviedo en el modelo PSO-RBF-SVM ajustado.	266

Figura 6.9. Importancia relativa de las variables de entrada para predecir el valor del material particulado (PM ₁₀) en el área urbana de Oviedo en el modelo PSO-RBF-SVM ajustado.....	269
Figura 6.10. Comparación entre los valores de NO ₂ observados y predichos por el árbol de regresión modelo M5, enfoque MLP y el modelo basado en PSO-SVM: a) árbol de regresión modelo M5; b) red MLP; y c) Modelo PSO-SVM con núcleo RBF	272
Figura 6.11. Comparación entre los valores de SO ₂ observados y predichos por el árbol de regresión modelo M5, el enfoque MLP y el modelo basado en PSO-SVM: a) árbol modelo M5; b) red MLP; Y c) Modelo PSO-SVM con núcleo RBF ...	273
Figura 6.12. Comparación entre las concentraciones de PM ₁₀ observadas y predichas por el árbol de regresión modelo M5, el enfoque MLP y el modelo basado en PSO-SVM: a) árbol modelo M5; b) red MLP; Y c) Modelo PSO-SVM con núcleo RBF	274
Figura 7.1. Ciudadanos, utilizando tecnologías colaborativas basadas en smartphones e internet, para la observación de su entorno.	293

Índice de tablas

Tabla 1.1. Valores límite PM ₁₀	13
Tabla 1.2. Valores de pico y media (24h), de distintos contaminantes, registrados al inicio del año 2017.	15
Tabla 1.3. Las cinco empresas más contaminantes de Asturias	20
Tabla 1.4. Las 22 empresas más contaminantes de Asturias.	21
Tabla 3.1. Contaminantes primarios y secundarios de la atmósfera.....	80
Tabla 3.2. Clasificación de las partículas según su tamaño y composición.	84
Tabla 3.3. Muertes prematuras derivadas de la exposición a los contaminantes indicados.....	105
Tabla 3.4. Normas de calidad del aire para O ₃ según la Directiva de la UE y las directrices de la OMS sobre calidad del aire.	107
Tabla 3.5. Consecuencias sobre la salud de los contaminantes contenidos en el smog fotoquímico.	115
Tabla 3.6. Niveles del Índice de Calidad de Aire (AQI) y sus consecuencias para la salud de las personas.	133
Tabla 3.7. Normas de aire ambiental basadas en la salud (concentración de los contaminantes según las normas de EE.UU.).	135

Tabla 4.1. Algoritmo básico de los árboles de decisión.....	170
Tabla 4.2. Resultado de nodos tomando covariable x_1 y valor de corte c	179
Tabla 4.3. Procedimiento M5.....	189
Tabla 4.4. Algoritmo del árbol modelo M5'	190
Tabla 4.5. Características de las metaheurísticas basadas en trayectoria.....	205
Tabla 4.6. Características de las metaheurísticas basadas en población.....	206
Tabla 4.7. Pseudocódigo a grandes rasgos del algoritmo PSO.	212
Tabla 6.1. Normas nacionales de calidad del aire ambiental de la US EPA.....	242
Tabla 6.2. Concentración media mensual de contaminación atmosférica en el área urbana de Oviedo entre enero de 2013 y diciembre de 2015	243
Tabla 6.3. Conjunto de variables de entrada físico-químicas utilizadas en este estudio y sus nombres junto con su media y desviación estándar.	258
Tabla 6.4. Hiperparametros óptimos del modelo ajustado PSO-RBF-SVM encontrado con la técnica de optimización de enjambre de partículas (PSO), para el dióxido de nitrógeno (NO ₂) en el área urbana de Oviedo.	260
Tabla 6.5.Los parámetros de ANN del perceptrón multicapa (MLP) para el dióxido de nitrógeno (NO ₂) en el área urbana de Oviedo	261
Tabla 6.6. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol M5 ajustado en este estudio Para el dióxido de nitrógeno (NO ₂) en la zona urbana de Oviedo	262
Tabla 6.7. Peso de las variables en el modelo de PSO-RBF-SVM ajustado para el valor del dióxido de nitrógeno (NO ₂) en el área urbana de Oviedo	263
Tabla 6.8. Hiperparametros óptimos del modelo de PSO-RBF-VM ajustado encontrado con la técnica de optimización de enjambres de partículas (PSO) para el dióxido de azufre (SO ₂) en el área urbana de Oviedo.....	264
Tabla 6.9. Parámetros de RNA del perceptrón multicapa (MLP) para el dióxido de azufre (SO ₂) en el área urbana de Oviedo.	264
Tabla 6.10. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol de regresión M5 ajustado en este estudio para el dióxido de azufre (SO ₂) en la zona urbana de Oviedo.....	265
Tabla 6.11. Peso de las variables en el modelo de PSO-RBF-SVM ajustado para el valor de dióxido de azufre (SO ₂) en el área urbana de Oviedo.	266

Tabla 6.12. Hiperparametros óptimos del modelo de PSO-RBF-SVM ajustado encontrado con la técnica de optimización de enjambre de partículas (PSO) para partículas de menos de 10 μm (PM10) en el área urbana de Oviedo.....	267
Tabla 6.13. Los parámetros de la RNA del perceptrón multicapa (MLP) para el aerosol de menos de 10 μm (PM10) en el área urbana de Oviedo.....	268
Tabla 6.14. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol de regresión M5 ajustado en este estudio Para el aerosol de menos de 10 μm (PM10) en el área urbana de Oviedo	268
Tabla 6.15. Pesos de las variables del modelo de PSO-RBF-SVM ajustado para el material particulado inferior a 10 μm (PM10) en el área urbana de Oviedo.	
.....	269

Capítulo I

INTRODUCCIÓN

Problema de investigación y objetivos de esta tesis. Noticias, marco legal y normativa

1. Problema de investigación y objetivos de esta tesis.

Noticias en prensa, marco legal y normativa

1.1. Introducción

Hace 25 años, el 8 de noviembre de 1992, en el diario La Nueva España se incluía esta noticia: “*Tres cabinas controlan la contaminación en Oviedo*”. En el texto que acompañaba este titular, se podía leer:

“Con tres cabinas situadas en puntos estratégicos de la ciudad (Figura 1.1), el Ayuntamiento hace estos días las pruebas para la puesta en marcha de la red de control de la contaminación. Las tres cabinas de la red están instaladas junto al Palacio de Deportes, en el cruce de la calle general Elorza con la avenida de Pumarín y otra junto a la plaza de toros”.

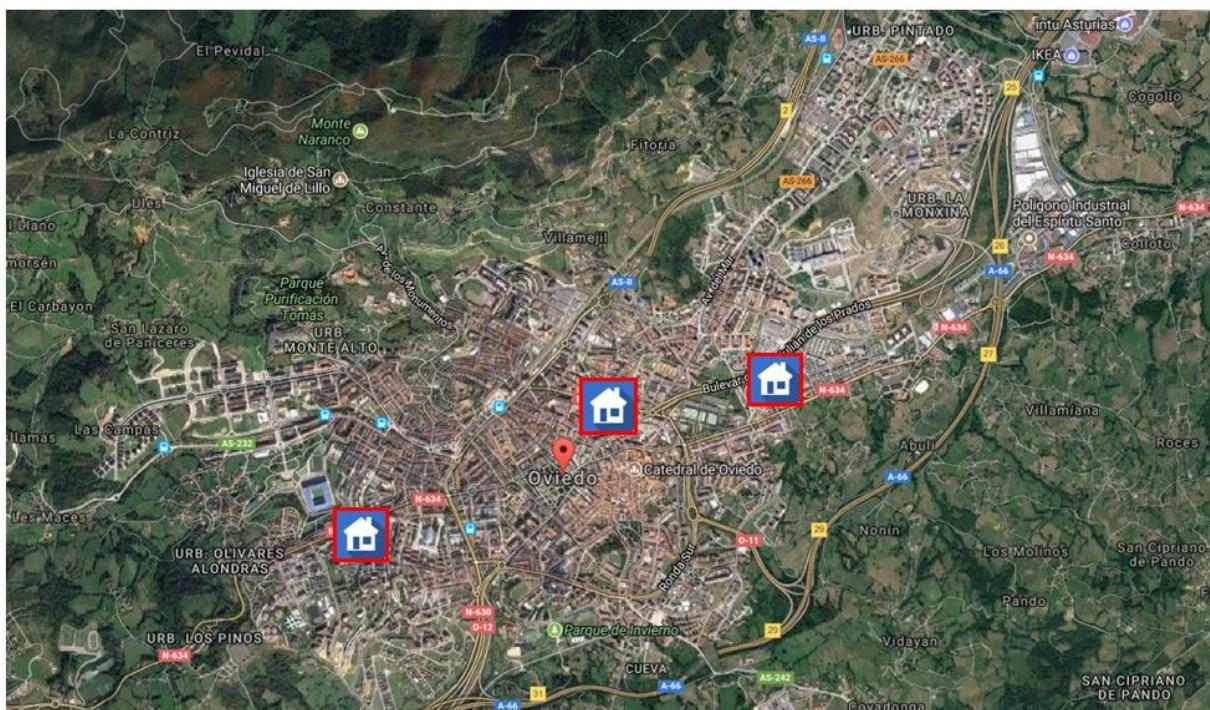


Figura 1. 1. Emplazamiento de las tres estaciones de medición de la contaminación en Oviedo.
Fuente: Elaboración propia a partir de Google maps.

Oviedo es un concejo y ciudad española, capital del Principado de Asturias. Su origen se remonta a la Alta Edad Media, y era llamada ‘Ovetum’.

Asimismo es un concejo asturiano cuya capital es la ciudad del mismo nombre, Oviedo, parroquia de dicho concejo, que comprende Oviedo capital. Es además el centro comercial, universitario, religioso y administrativo del Principado, sede de la Junta General del Principado de Asturias, de las instituciones oficiales del Principado de Asturias, de la Universidad de Oviedo, del Museo de Bellas Artes de Asturias, de los Premios Princesa de Asturias así como de la Archidiócesis de Oviedo.

El municipio de Oviedo es el segundo más poblado de la comunidad autónoma, con 220.567 (INE 2016) habitantes, ocupando el puesto número 21 entre los municipios más poblados de España y está incluido en el área metropolitana central de Asturias, que engloba a más de 860 000 personas, y en el llamado ocho asturiano [Wikipedia]. La zona ES0302, Asturias Central, es un ámbito heterogéneo ya que cuenta con varios núcleos importantes de población, englobando Oviedo, Avilés, Mieres y Langreo, una zona portuaria en Avilés y varias zonas de influencia industrial. La actual zonificación de Asturias es igual para todos los contaminantes, data de 2001 y sigue un criterio geográfico, Occidental, Central y Oriental, donde la zona central concentra el mayor número de estaciones y población. El R.D. 102/2011, relativo a la mejora de la calidad del aire, contempla que cada cinco años debe revisarse la clasificación de las zonas, recomendando que en caso de existir cambios significativos en las actividades, esta revisión se realice sin esperar a ese plazo.

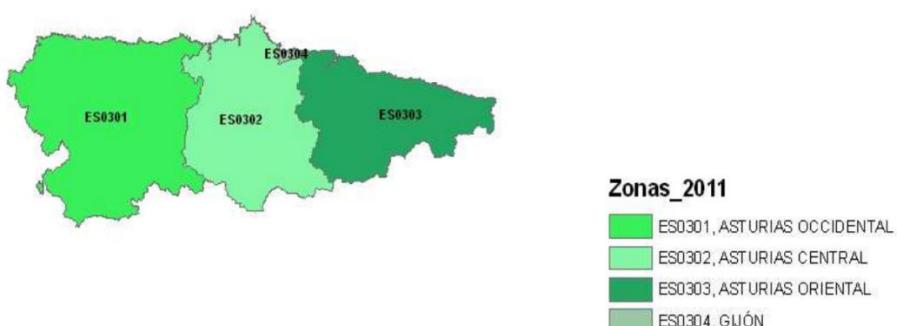


Figura 1.2. Zonificación de la calidad del aire de España 2011. Zonas de Asturias. Fuente: GOBIERNO DEL PRINCIPADO DE ASTURIAS

La preocupación motivada por la contaminación en la ciudad de Oviedo, como en otras ciudades de la zona Asturias Central, es constante; tanto por parte de las autoridades como por los grupos ecologistas, vecinos y asociaciones en general.

1.1.1. Problema de investigación

La contaminación ambiental merece mucha atención, por los efectos en la salud de las personas y la economía en general. Para tratar problemas de contaminación del aire es necesaria la intervención de un equipo multidisciplinario. En este punto estarían involucrados, biólogos, ambientalistas, matemáticos, sociólogos, médicos, ingenieros, etc. El papel que vendría a desempeñar el matemático o el ingeniero, sería el que corresponde al tratamiento de datos y a determinar en qué forma se está contaminando el ambiente. Esta tarea se realiza mediante el desarrollo de modelos matemáticos y simulaciones numéricas. Es fácil de comprender el interés creciente por el uso de modelos matemáticos, para modelizar el comportamiento de los contaminantes en la atmósfera, con el fin de mejorar la calidad del aire. En esta investigación, se propone avanzar desde un contexto más general a otro más particular, que permita comprender la influencia y la magnitud de los factores que determinan los niveles de contaminación existente en el área de estudio. En el punto 1.2, motivación y contextualización, se aportan noticias de prensa y otros datos de informes elaborados por entidades acreditadas, que justifican la elaboración de este estudio de investigación, mediante el cual se pretende crear una herramienta matemática, que permita la toma de decisiones que contribuyan a dar respuesta a un problema de contaminación atmosférica, a escala local.

1.1.2. Objetivos

En primer lugar, se plantea un OBJETIVO GENERAL que pretende dar respuesta a necesidad anteriormente descrita, con el propósito de lograr un modelo aplicable a la resolución de problemas relacionados con la calidad del aire respirable, utilizando técnicas basadas en el aprendizaje estadístico, para el área metropolitana de Oviedo, en el norte de España.

En segundo lugar, para llevar a cabo el OBJETIVO GENERAL descrito, se han planteado tres OBJETIVOS ESPECÍFICOS que se derivan de dicho objetivo, que clarifican y concretan las acciones que se deben efectuar.

Por último, se concretan los anteriores OBJETIVOS ESPECÍFICOS en una serie de operaciones detalladas en forma de OBJETIVOS OPERATIVOS, que nos van a permitir alcanzar los logros planteados en dichos objetivos generales. Seguidamente se detallan todos los objetivos descritos.

OBJETIVO GENERAL.

Desarrollar un modelo matemático para estimar las concentraciones de los agentes químicos presentes a escala local y causantes de daños en la salud de personas expuestas, el agua consumible, los cultivos y otros recursos; que pueda, en su caso, constituir una herramienta de apoyo efectivo en las acciones que deban llevar a cabo las autoridades designadas.

OBJETIVO ESPECÍFICO 1. Implementar un algoritmo basado en la optimización por nube o enjambre de partículas (en inglés, Particle Swarm Optimization – PSO) y máquinas de soporte vectorial o máquinas de vectores soporte (en inglés, Support Vector Machines-SVM), mediante un método híbrido denominado PSO-SVM.

OBJETIVO ESPECÍFICO 2. Ajustar un árbol de regresión M5 y una red neuronal de tipo perceptrón multicapa (MLP) al conjunto de datos experimentales y obtener los modelos resultantes.

OBJETIVO ESPECÍFICO 3. Contrastar la robustez del algoritmo, para comprobar si el modelo híbrido PSO-SVM propuesto es más robusto que los métodos basados en red neuronal MLP y árbol de modelos M5; resultando más estable en un amplio intervalo de situaciones y mostrando mejor desempeño.

Para alcanzar los anteriores objetivos generales se han planteado los siguientes OBJETIVOS OPERATIVOS:

OBJETIVO OPERATIVO 1. Obtención y preparación del conjunto de datos de contaminación atmosférica, a partir de los datos registrados por las estaciones de la red de monitorización de la calidad del aire del Gobierno del Principado de Asturias, entre los meses de enero de 2013 y diciembre de 2015.

OBJETIVO OPERATIVO 2.- Caracterización del origen natural o antropogénico, local o externo, de las fuentes contaminantes, así como los mecanismos que los generan.

OBJETIVO OPERATIVO 3.- Estimación de la tendencia mensual de las concentraciones de SO₂, NO, NO₂, CO, PM₁₀ y O₃, durante los años 2013, 2014 y 2015, en el área urbana de Oviedo.

OBJETIVO OPERATIVO 4.- Estudio de los métodos, herramientas e implementación de los algoritmos empleados para la resolución del problema de investigación que se plantea:

- Máquinas de soporte vectorial (SVM)
- Algoritmo de optimización por enjambre de partículas (PSO)
- Red neuronal de tipo Perceptrón Multicapa (MLP)
- Árbol de regresión modelo M5

OBJETIVO OPERATIVO 5.- Establecimiento de los medios técnicos necesarios, características del ordenador, procesador y sistema operativo; así como el paquete software más adecuado para efectuar los ensayos con la precisión deseada.

1.1.3. Organización de la tesis

El presente documento está organizado del siguiente modo:

En el **capítulo I**, se formulan los objetivos de esta investigación y el alcance de esta investigación. Se presenta la forma en la que esta tesis doctoral está organizada y se indican las aportaciones realizadas.

Seguidamente, se contextualiza el marco de desarrollo en donde se ofrecen datos de la situación ambiental de Asturias, mediante referencias a informes y estudios contrastados. Se recopilan titulares de prensa, respecto al problema de la contaminación y el estado del aire en Asturias, recopilados durante el primer semestre de 2017. Por último, se describe el marco normativo y legislativo aplicable.

En el **capítulo II**, se describen los fundamentos del análisis de datos, la inteligencia artificial y el aprendizaje automático. Se presentan las redes neuronales; así como la arquitectura y funcionamiento del perceptrón. Por último, se detalla el proceso del Descubrimiento de Conocimiento a partir de Bases de Datos (en inglés, Knowledge Discovery from Databases – KDD), y los fundamentos de la minería de datos.

En el **capítulo III**, se comentan aspectos sobre los contaminantes del aire detallando las fuentes y los tipos de contaminación atmosférica. Se definen los tipos de contaminantes primarios y secundarios; así como sus características principales, parámetros y efectos más relevantes. Por último, se refieren los contaminantes denominados ‘Criterio’, los efectos sobre la salud de las personas y los índices y herramientas empleadas por la Agencia de Protección Ambiental de Estados Unidos (en inglés, *United States Environmental Protection Agency - US EPA*), para medir la calidad del aire.

En el **capítulo IV**, se refieren los materiales y los métodos empleados en esta tesis. En primer lugar se describen las máquinas de vectores soporte (SVM) y las máquinas de vectores soporte para regresión (en inglés, Support Vector Regression – SVR). Después se describen los árboles de decisión, árboles modelo M5, árbol modelo M5' y M5 Rules.

Posteriormente se habla de las técnicas y algoritmos de optimización y del enfoque basado en el algoritmo de optimización por enjambre de partículas (en inglés, *particle swarm optimization* – PSO). Por último, se describe el método de optimización de SVM mediante PSO.

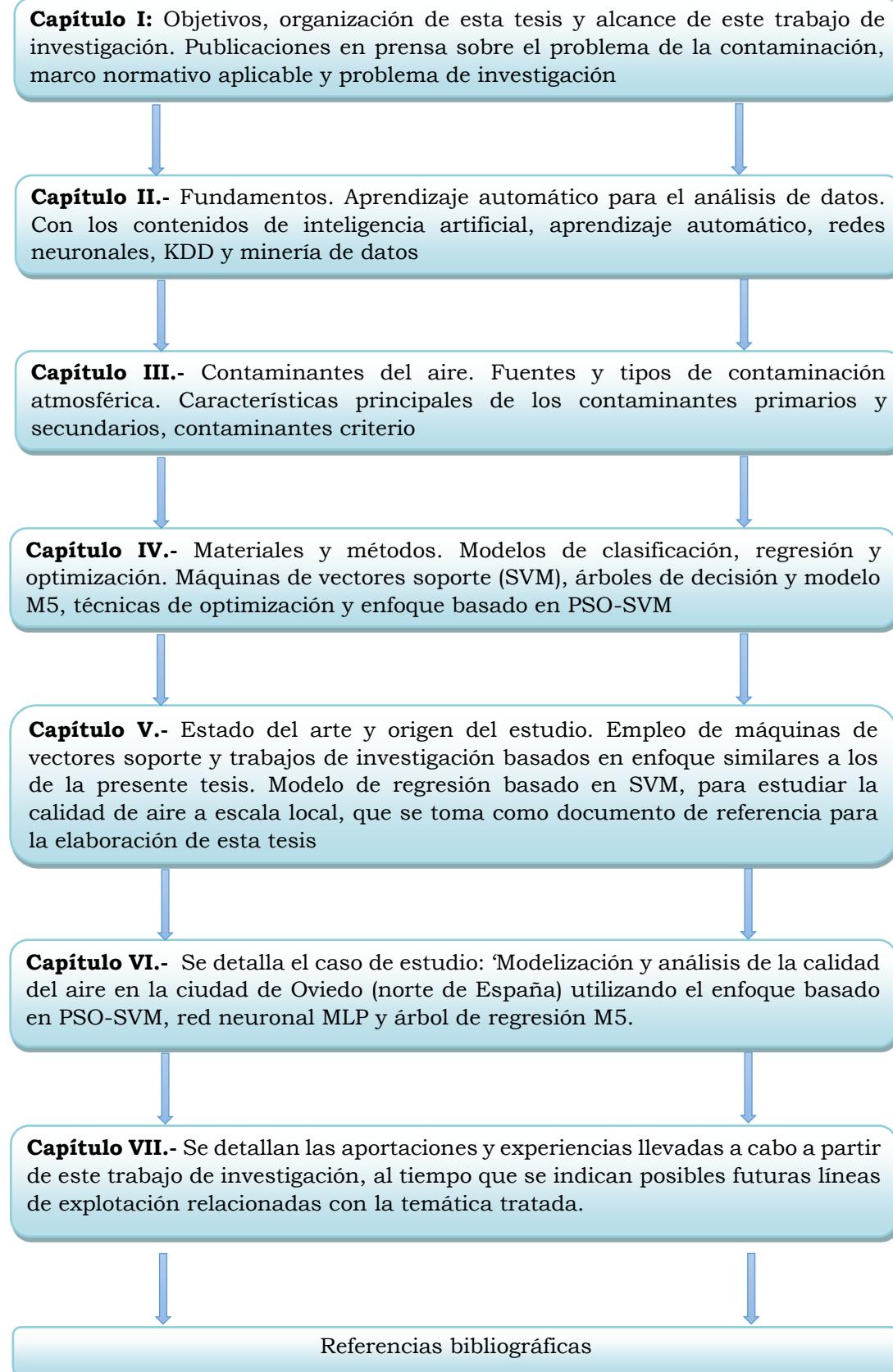
El **capítulo V** contiene los hallazgos documentales y la revisión del ‘Estado del Arte’. Se refiere la situación actual del ‘Arte’ y se describen brevemente los hallazgos documentales de interés, sobre investigaciones actuales que tienen cierta analogía con esta investigación. Por último, se comenta el origen de esta tesis a partir de un documento de investigación anterior que emplea técnicas, que dejan abiertas las líneas generales que ahora se retoman en este trabajo de investigación.

En el **capítulo VI**, se detalla el caso de estudio: ‘Modelización y análisis de la calidad del aire utilizando el enfoque basado en PSO-SVM, red neuronal MLP y árbol de modelos M5, en la ciudad de Oviedo (norte de España), que constituye el eje principal de este trabajo de investigación.

En el **capítulo VII**, se especifican las aportaciones y experiencias llevadas a cabo a partir de este trabajo, al tiempo que se indican las futuras líneas de investigación sobre la temática tratada, que se pueden abordar a partir de esta tesis doctoral; proporcionando una visión unificada del trabajo realizado y de las aplicaciones derivadas del mismo.

Como **apartado final**, se especifican las referencias bibliográficas empleadas para la elaboración de este trabajo de investigación.

Esquema general de esta tesis doctoral:



1.1.4. El alcance de este trabajo

En este apartado se expresa la intención de contribuir a la comunidad científica, con los logros que se consigan en este trabajo. Mediante la siguiente investigación se desea elaborar una herramienta matemática que posibilite que la aplicación de los resultados obtenidos sobre las concentraciones de contaminantes estudiadas, sean una valiosa herramienta para proyectos de mitigación de lluvia ácida y para la investigación de los efectos de las partículas en la salud humana, etc. Dicha herramienta, permitirá modelar con un grado de precisión adecuado el diagnóstico de la calidad del aire, a escala local, mediante un modelo híbrido basado en PSO-SVM con una función de núcleo o ‘kernel’ RBF (en inglés, radial basis function-RBF). Del mismo modo, se podrá establecer el orden de importancia de las variables de entrada implicadas en la predicción de las concentraciones de los contaminantes tratados.

Se implementarán nuevos algoritmos, que permitirán diseñar un instrumento que podría ser utilizada por las autoridades designadas, y que sin duda constituiría un apoyo efectivo para decidir las acciones que deban llevarse a cabo. En este trabajo, se contrastarán los modelos obtenidos, para así determinar la robustez del modelo propuesto.

1.2. Contextualización y motivación

Los estudios realizados por el Instituto de Salud Carlos III (ISCIII)[\[1\]](#), sitúan a Asturias entre las más castigadas por cáncer de pulmón cada año, estrechamente relacionado con el tabaco y la contaminación atmosférica. Pese a tratarse de una de las comunidades con mayor índice de contaminación, no debe determinarse una causalidad. Dicho informe detalla microdatos de mortalidad, facilitados por el Instituto Nacional de Estadística (INE), tras la firma de un acuerdo de confidencialidad. En el trabajo se dice que seis de los primeros diez municipios españoles con mayor tasa de mortalidad por cáncer de pulmón, en 2010, están en Asturias.

El promedio en España este año es de 39,34 muertos por cada 100.000 personas, siendo el valor registrado en el concejo de Aller de 119,21 muertos por cada 100.000 personas. En el citado documento según los datos manejados en 2013, Langreo fue el quinto municipio con mayor tasa de mortalidad por cáncer, falleciendo por esta enfermedad, 97,5 personas por cada 100.000 habitantes. En esta localidad se superó en 20 puntos la tasa de mortalidad para ambos sexos de Hungría, país con mayor índice según el informe Globocan, elaborado por la Organización Mundial de la Salud [2].

Este conjunto de datos comprende el período solicitado entre 2009 y 2014. Debido a la ley de secreto estadístico, el acuerdo firmado limita el estudio a aquellos municipios con más de 10.000 habitantes, por lo que el número final de ayuntamientos es de entre 743 y 749 en función del año que se observe; pese a que en España existen, según el padrón municipal a 1 de enero de 2016, 8.119 municipios. Todas las tasas de mortalidad mencionadas en el reportaje son ‘tasas crudas’, aunque el Instituto de Salud Carlos III aconseja la utilización de ‘tasas ajustadas por edad’: “El riesgo de muerte se incrementa con la edad”, explica, por lo que advierte de que las ‘tasas crudas’ pueden incrementarse simplemente porque la población envejece. Una tasa cruda (bruta), relaciona un suceso que afecta a una muestra de la población con la población total. Por ello, dichas tasas no sirven para comparar entre poblaciones con distintas estructuras; p. ej., la tasa bruta de mortalidad en 2015 de Reino Unido (9,20%) es mayor que la de Túnez (6,20%), y esto es debido al mayor índice de envejecimiento del Reino Unido; es decir que en Reino Unido hay muchos más viejos en relación a la población total. La otra gran base de datos utilizada es de acceso público, el Registro Estatal de Emisiones y Fuentes Contaminantes (PRTR) [3]. En la comarca de Avilés además de la fábrica Asturiana de Zinc, se emplaza la factoría de ArcelorMittal (antigua Ensidesa), en el concejo de Castrillón. La estación de medición de partículas contaminantes en Matadero, entre el puerto y el cuidado casco histórico, es la única de la red pública que supera el umbral de PM₁₀, uno de los agentes atmosféricos más dañinos para la salud.

El Real Decreto 102/2011, de 28 de enero, relativo a la mejora de la calidad del aire (modificado por el Real Decreto 39/2017, de 27 de enero) establece objetivos de calidad del aire, de acuerdo con el anexo III de la Ley 34/2007, y, como se puede apreciar en la Tabla 1.1, fija en 40 microgramos por metro cúbico los límites para estas sustancias, el doble de lo recomendado por la OMS. “No son cancerígenas, pero limitan la capacidad del pulmón para eliminar las sustancias cancerígenas absorbidas”, advierte la doctora Adonina Tardón (coordinadora del estudio Capua-Cáncer de pulmón en Asturias).

C. Valores límite de las partículas PM₁₀ en condiciones ambientales para la protección de la salud

	Período de promedio	Valor límite	Margen de tolerancia	Fecha de cumplimiento del valor límite
1. Valor límite diario.	24 horas.	50 µg/m ³ , que no podrán superarse en más de 35 ocasiones por año.	50% (1).	En vigor desde el 1 de enero de 2005 (2).
2. Valor límite anual.	1 año civil.	40 µg/m ³	20% (1).	En vigor desde el 1 de enero de 2005 (2).

(1) Aplicable solo mientras esté en vigor la exención de cumplimiento de los valores límite concedida de acuerdo con el artículo 23.

(2) En las zonas en las que se haya concedido exención de cumplimiento, de acuerdo con el artículo 23, el 11 de junio de 2011.

Tabla 1.1. Valores límite PM₁₀. Fuente: Anexo I del Real Decreto 102/2011, de 28 de enero, relativo a la mejora de la calidad del aire.

Según se muestra en la Figura 1.3, los niveles registrados por la estación ubicada en la factoría de ArcelorMittal, superan casi todos los días los umbrales fijados por la OMS para el tipo de partículas PM₁₀. Por otro lado la estación de Tremañes, en Arcelor, deja entrever una relajación de la contaminación en la planta y registra niveles de benceno más altos de lo permitido. La ciudad respiró diez toneladas de este carcinógeno de primer nivel, consecuencia de las emisiones de ArcelorMittal. Este compuesto químico salido de los hornos de coque de la fábrica acabó con la vida de un empleado en 2012 [4].

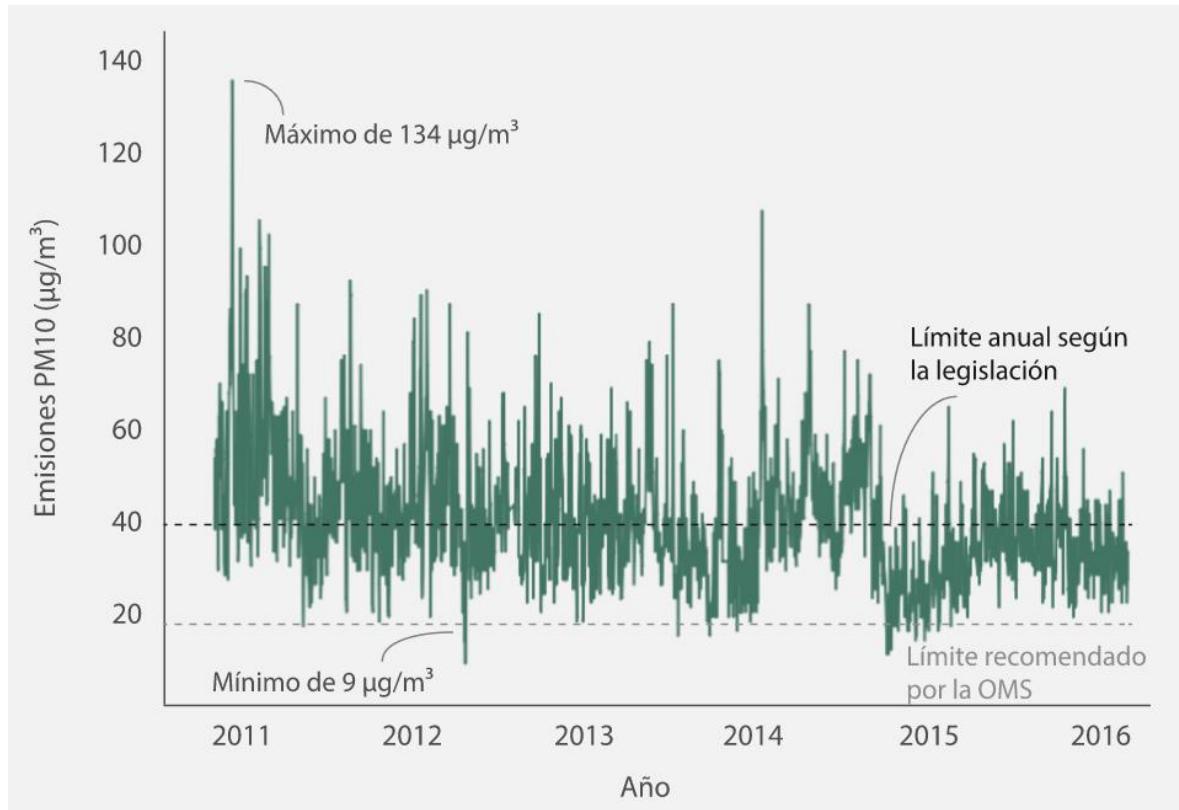


Figura 1.3. Emisiones de PM10 (µg/m³). Fuente: Red Ambiental de Asturias

1.2.1. Noticias publicadas en relación con la contaminación ambiental en Asturias durante el primer semestre de 2017.

Asturias comienza 2017 con niveles elevados de contaminación. Sólo queda esperar a que llueva, para paliar este grave problema. Los datos que se facilitan seguidamente aparecen en la web del Principado, el Puerto del Musel y el Puerto de Avilés, el día 2 de enero del año en curso. El Principado no facilita los datos en continuo de las estaciones privadas, que son la mayoría, pese a que todas ellas están monitorizadas.

Como viene siendo habitual según los datos de que se dispone, permanecen disparados los valores referidos al material particulado PM_{2,5} y PM₁₀, y hay que recordar que venimos registrando en los últimos años los peores valores de toda España, que multiplican los valores recomendados por la OMS.

Lecturas de varios contaminantes registradas el 2 de enero de 2017. Fuente: Web, Principado de Asturias.											
	<i>Langreo</i>		<i>Siero</i>		<i>Gijón</i>		<i>Oviedo</i>		<i>Castrillón</i>		
	Valor de pico $\mu\text{g}/\text{m}^3$	Valor Media 24h $\mu\text{g}/\text{m}^3$	Valor de pico $\mu\text{g}/\text{m}^3$	Valor Media 24h $\mu\text{g}/\text{m}^3$	Valor de pico $\mu\text{g}/\text{m}^3$	Valor Media 24h $\mu\text{g}/\text{m}^3$	Valor de pico $\mu\text{g}/\text{m}^3$	Valor Media 24h $\mu\text{g}/\text{m}^3$	Valor de pico $\mu\text{g}/\text{m}^3$	Valor Media 24h $\mu\text{g}/\text{m}^3$	
Partículas menores de 2,5 micras (PM_{2,5})	56	36	56	35	65	33	43	29	46	24	
Partículas menores de 10 micras (PM₁₀)	Gozón		Castrillón		Oviedo		Corvera		<i>En el caso de Corvera, no se facilitan datos de continuo. Se desconocen por tanto los datos de pico.</i>		
	89	72	194	57	94	57		52			
	Oviedo				Avilés						
Benceno	Valor de pico $\mu\text{g}/\text{m}^3$		Valor Media 24h $\mu\text{g}/\text{m}^3$		Valor de pico $\mu\text{g}/\text{m}^3$		Valor Media 24h $\mu\text{g}/\text{m}^3$				
	10		5		10		5				

Tabla 1.2. Valores de pico y media (24h), de distintos contaminantes, registrados al inicio del año 2017.
 Fuente: Elaboración propia, sobre datos de la Web del Principado de Asturias.

Hay que recordar que la Organización Mundial de la Salud – OMS, recomendó en 2005 valores límite para partículas PM_{2,5} de 10 $\mu\text{g}/\text{m}^3$ y estableció un límite de 25 $\mu\text{g}/\text{m}^3$ cada 24 horas. La normativa española, que viene siendo más tolerante, establece un límite de alerta de 25 $\mu\text{g}/\text{m}^3$ a las 24 horas de exposición para las partículas PM_{2,5}.

Por otra parte, la Organización Mundial de la Salud recomendó en 2005 unos valores límite de 20 $\mu\text{g}/\text{m}^3$ de media anual, para las PM₁₀ y un valor de alerta con una media diaria de 50 $\mu\text{g}/\text{m}^3$ para dichas partículas. La normativa española, para este tipo de material particulado, establece 40 $\mu\text{g}/\text{m}^3$ de media anual y 50 $\mu\text{g}/\text{m}^3$ para las 24 horas. Del mismo modo, el valor de referencia anual en las guías de la Organización Mundial de la Salud en el caso del benceno, está fijado en 1,7 $\mu\text{g}/\text{m}^3$ de media anual, mientras que la normativa española, mucho más tolerante, lo fija en 5 $\mu\text{g}/\text{m}^3$ de media anual.

En la línea de lo expuesto hasta ahora, para constatar el grado de preocupación que origina el problema de la contaminación atmosférica en Asturias, se exponen seguidamente los titulares que han ido apareciendo en los medios de comunicación:

¿Cómo afecta la contaminación a la salud?

Una exposición regular al aire contaminado puede alterar las defensas inmunitarias del cuerpo. La Nueva España - LNEEp/Madrid, 31.12.2016 | 18:11

El aumento de la contaminación en Asturias reaviva el debate sobre cómo combatirla.

IU urge al Principado la convocatoria de una reunión «inmediata» con los ayuntamientos más afectados y diversos agentes sociales para suscribir «un gran pacto» contra la polución. ELCOMERCIO.ES, 7 enero 2017, 18:47

Los asmáticos piden medidas "reales" contra la contaminación en Asturias.

LNE - Oviedo, M. G., 09.01.2017 | 16:36

Pacientes asmáticos exigen medidas reales contra la contaminación.

El presidente de la federación nacional y asturiana, Javier Palicio, lamenta que la administración haga «más propaganda política que realidad». El consejo más eficaz para los enfermos es «no salir de casa». EUROPA PRESS REDACCIÓN, 09/01/2017, 17:26

Trubia soporta la mayor contaminación de benceno del país, según los ecologistas.

Lugones vuelve a estar en alerta roja por partículas. LNE - Oviedo, Mónica G. Salas, J. Fernández, 09.01.2017 | 03:36

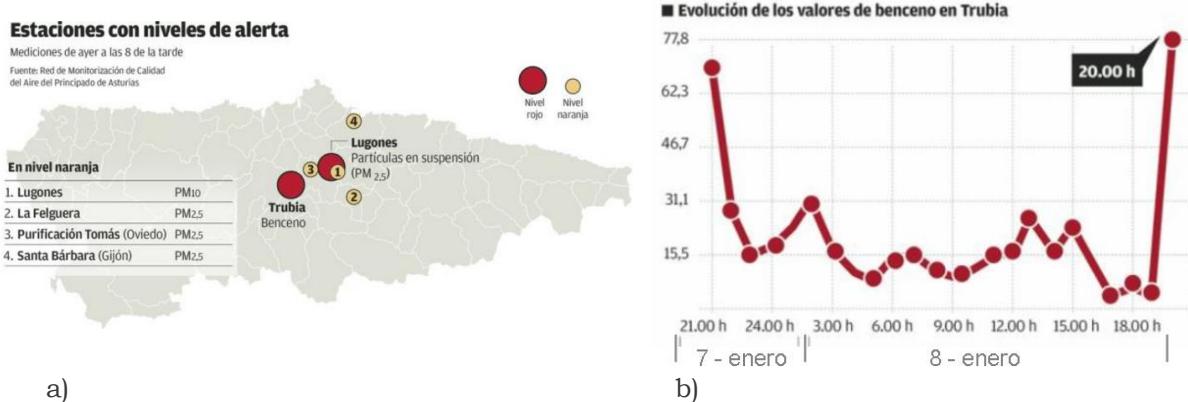


Figura 1.4. a) Situación de las estaciones en niveles naranja y rojo. b) Evolución de los niveles de benceno en Trubia en los días de la noticia. Fuente: LNE, respecto a los valores de la Red de Monitorización de Calidad de Aire en Asturias.

Adonina Tardón, dirige una unidad del Instituto Universitario de Oncología, en Oviedo. "La contaminación hace que los niños nazcan con menos peso y sufran asma".

"Tenemos pocas estaciones de calidad del aire y no todas miden los parámetros legislados; por tanto no hacemos un diagnóstico exacto". La Nueva España – LNE Oviedo, 10.01.2017 | 05:13

La situación ambiental en la región. "La polución nos ahoga", claman los vecinos de Trubia sobre el benceno.

"Si mi ático está sucio, mis pulmones también", dice Luz García sobre el "polvo" que se acumula en las terrazas. La Nueva España - LNE Trubia (Oviedo), 10.01.2017 | 03:34

"Nadie nos hace caso", dicen en Latores.

La Nueva España - LNE 10.01.2017 | 03:34

Oposición y ecologistas reclaman más control de las emisiones en las empresas.

La calidad del aire empeoró, con una decena de estaciones medidoras a niveles superiores a lo normal. ELCOMERCIO.ES, 10 enero 2017 02:28.

La lluvia no diluye la contaminación.

La Coordinadora Ecologista señala que los niveles de polución siguen disparados pese al cambio de tiempo. LA VOZ OVIEDO 10/01/2017 13:44



Figura 1.5. VISTA DE LA CONTAMINACIÓN EN OVIEDO. Fuente: La Voz de Asturias

El tráfico y la calefacción causan la contaminación, sostiene FADE.

Los niveles de partículas y de benceno remiten en todas las estaciones de calidad del aire por la persistente lluvia. LNE. Oviedo, Mónica G. Salas 11.01.2017 | 01:58

Asturias, la comunidad con mayor polución industrial: casi un quinto de la de España

Un total de veintisiete instalaciones emitieron 188.450 toneladas de contaminantes en 2015, de las que 126.212 tuvieron origen en Arcelor. LNE. Gijón, M. C. 11.01.2017 | 09:32

La lluvia puede con la contaminación.

La Fade cree «injusto» el trato a las empresas, mientras Equo y Ciudadanos llevan al Congreso iniciativas para mejorar la calidad del aire. ELCOMERCIO.ES 11 enero 2017 01:22

La polución afecta a la salud pero se desconoce su impacto real, según expertos.

RTPA 11/1/2017

La contaminación en Asturias se mantuvo en niveles «razonables» en 2016, según el Principado.

La consejera de Medio Ambiente Belén Fernández, asegura que los niveles registrados el pasado año y a comienzos de 2017 se encuentra dentro de lo que permite la legislación. EFE | OVIEDO 11 enero 2017 20:16

El área central se someterá a protocolos contra la contaminación, como Oviedo y Gijón.

Medio Ambiente sostiene que las estaciones medidoras de Asturias han registrado en 2016 valores "razonables, mejores que el año anterior". lne.es Oviedo, M. Salas/ Juan A. Ardura 12.01.2017 | 01:51

El Principado reunirá a 16 municipios para luchar contra la contaminación.

Pretende diseñar un protocolo de actuación y en el encuentro de este mes también participará la Federación Asturiana de Concejos. ELCOMERCIO.ES 12 enero 2017 05:18

¿Por qué el PSOE sigue criminalizando a los ciudadanos de la contaminación?

Los ecologistas ponen en duda el trabajo de la Consejera. AsturiasVerde.net 12 enero 2017. Fuente: Nota de prensa de la Coordinadora Ecoloxista d'Asturies

Los ecologistas critican que el Principado deja contaminar sin control a Arcelor.

Aseguran que el estado de las baterías de coque de la factoría de Avilés es lamentable. LNE 14-01-2017

Castrillón encabeza la contaminación en Asturias, dicen los ecologistas.

LNE. G.C.M. 15-01-2017. 04:18

Las plataformas contra la contaminación de Gijón y Avilés aúnan esfuerzos.

Celebran un encuentro en La Calzada al que acudirán vecinos de otros puntos de Asturias. LNE. M. C. 17.01.2017 | 03:09

La estación de Salinas que mide la contaminación no funciona, dicen los ecologistas

LNE Piedras Blancas, I. Montes, 20.01.2017 | 03:25

Belén Fernández asegura que Asturias "es pionera" a la hora de luchar contra la contaminación.

20minutos.es. EUROPA PRESS. 27.01.2017

La industria reducirá su actividad en Gijón en situaciones de alerta por contaminación.

El ejecutivo regional plantea condicionar los permisos a esta reacción. Las ocho de la mañana es la hora en la que Gijón tiene el aire más sucio

ELCOMERCIO.ES, 28 enero 2017

Estas son las 22 empresas más contaminantes de Asturias.

Un informe del Observatorio de la Sostenibilidad revela que las cinco máximas emisoras de la región acaparan el 95% del CO₂ que expulsan las estaciones fijas a la atmósfera. La Voz de Asturias, 28/01/2017, 05:00

Hidrocantábrico es la empresa con mayores emisiones (10.506.460 Tm eq C, la unidad utilizada en el estudio para medir la polución) y un 47% de contribución al total regional de fuentes fijas. Estas fuentes fijas son las que producen energía, la siderurgia, la industria del petróleo, las cementeras y otras instalaciones industriales y a ellas hay que sumar las fuentes de emisión difusas (transporte, hogares, servicios, agricultura...). El CO₂ que lanza HC a la atmósfera casi duplica al segundo productor de la región, ArcelorMittal.

Accionista principal	emisiones verificadas 2015
Hidrocantábrico	10.506.460
ArcelorMittal	5.907.100
Gas Natural Fenosa	1.786.748
Iberdrola	1.761.226
Cementos Tudela Veguin	1.276.632

Tabla 1.3. Las cinco empresas más contaminantes de Asturias. Fuente: Observatorio de la Sostenibilidad

El informe también recoge que las instalaciones de Aboño 1 y 2, son las más contaminantes de todo el país, por delante de las de Endesa en Galicia y Andalucía, y que el cuarto lugar de esa clasificación lo ocupan las plantas de Arcelor en Avilés y Gijón. En la Tabla 1.4, se muestra la lista completa de las 22 empresas más contaminantes de Asturias:

1. Hidrocantábrico	12. Fertiberia
2. Arcelormittal	13. Calera de San Cuaao
3. Gas Natural Fenosa	14. Dupont
4. Iberdrola	15. Industrial Química del Nalón
5. Cementos Tudela Veguín	16. Industrias Doy
6. EDP Cogeneración	17. Planta Cogeneración en Fábrica Cementos Tudela Veguín
7. Cogeneración de Navia	18. Asturiana de Zinc
8. Saint-Gobain	19. RHI Refractories, SL (antigua RHI Refractories España SA.)
9. Alcoa	20. Refractaria
10. Celulosas de Asturias	21. Hospital Central de Asturias
11. Cogeneración (Corporación Alimentaria Peñasanta)	22. Enagás

Tabla 1.4. Las 22 empresas más contaminantes de Asturias. Fuente: Observatorio de la Sostenibilidad

La térmica de Aboño, la que más contamina del país. Los empresarios solicitan que la reducción de la actividad en casos de alerta por polución se limite a “casos totalmente excepcionales [sic]”.

Con 7,5 millones de toneladas de CO₂, encabeza las diez instalaciones con más emisiones. En cuarto lugar están las factorías de Arcelor.
ELCOMERCIO.ES, 29 enero 2017

Asturias emitió treinta millones de toneladas contaminantes en 2015.

Las industrias de energía de la región generan casi el 74% de las emisiones.
La Nueva España – LNE, Oviedo 29.01.2017 | 03:33

IU reclama un gran pacto social y político contra la contaminación en Asturias.

La Nueva España – LNE, 02.02.2017 | 02:03

El Principado se ofrece para frenar la contaminación de Lugones.

La Nueva España – LNE, 02.02.2017 | 04:09

Los ecologistas alertan de la alta contaminación en Langreo pese a las lluvias.

La Nueva España – LNE. Langreo 07.02.2017 | 03:05

"No bajó la polución, bajó el empleo", critican los afectados por la contaminación.

La plataforma ciudadana rechaza la "autocomplacencia" de la consejera de Medio Ambiente, a la que acusa de no tomar "medidas efectivas". La Nueva España – LNE. M. C. 08.02.2017 | 03:59

La contaminación superó los límites admisibles.

La Nueva España – LNE. 08.02.2017 | 03:59

"En Oviedo sigue disparada la contaminación (17/02/17)"

Nota de prensa de la Coordinadora Ecoloxista d'Asturies

La tasa por emisiones abocará a la siderurgia a la desaparición, advierte Lakshmi Mittal.

El consejero delegado de ArcelorMittal señala al diario 'Expansión' que el mecanismo europeo para reducir el CO₂ incrementará de forma «sustancial» los costes de producción a la industria de países como España. ELCO M E RCI O.E S, 24 febrero 2017 00:47

Un estudio atribuye a la polución del aire la muerte de 300 asturianos al año.

El trabajo del Instituto de Salud Carlos III señala que Oviedo fue la capital de provincia más contaminada por las PM₁₀ en una década. ELCO M E RCI O.E S, 2 marzo 2017, 01:00

«Asturias no es un microcosmos en el que la contaminación actúa de forma diferente».

Podemos e IU acusan al consejero de Sanidad de «negar la evidencia», que insiste en que no hay estudios en la región que vinculen polución y cáncer. ELCO M E RCI O.E S, 3 marzo 2017, 04:44

La enésima nube de contaminación dispara la alarma en Gijón.

La coordinadora ecologista ha presentado denuncia en el ayuntamiento y la Consejería de Infraestructuras. LA VOZ GIJÓN 05/03/2017, 17:57

FADE carga contra las medidas de recorte de emisiones: "Ralentizan la actividad industrial".

La patronal insta a modernizar las calefacciones, aislar los edificios e impulsar el transporte colectivo. La Nueva España - LNE Oviedo, 06.03.2017 | 03:29

Oviedo es la ciudad más contaminada del norte de España, constata un estudio.

La capital asturiana encabeza el ranking de polución por partículas, que mató en nueve años a 4.136 personas, y registra altos niveles de NO₂. La Nueva España - LNE Oviedo, 06.03.2017 | 03:29

Sanidad se aparta de los criterios de la OMS para negar la relación entre contaminación y cáncer.

La autoridad sanitaria internacional considera carcinógena la mala calidad del aire desde el 2013 y advierte de su impacto en los cánceres de pulmón y vejiga. Recomienda medios eléctricos de transporte y más controles a las industrias sucias. La Voz de Asturias-OVIEDO, 06/03/2017, 05:00

El Principado no prevé hacer un estudio propio sobre contaminación y salud.

La OMS señala que los casos de cáncer de pulmón relacionados con actividades laborales representan el 10% del total y son muy inferiores a los causados por la mala calidad del aire. La Voz de Asturias-OVIEDO, 06/03/2017, 05:00

La industria deberá reducir su actividad en caso de alerta por alta contaminación.

El Principado ultima un protocolo de actuación consensuado con dieciséis concejos que se centrará en el tráfico, la calefacción y las fábricas. La Nueva España - LNE Oviedo, 08.03.2017 | 03:55

La mala calidad del aire, noveno factor de riesgo de enfermedades cardiovasculares.

La Nueva España - LNE Oviedo, 08.03.2017 | 01:23

El Principado ve desfasado el trabajo que achaca 300 muertes al año a la polución.

«Los niveles actuales no tienen nada que ver, son la mitad», afirma sobre el estudio que analizó la contaminación del aire en Oviedo de 2000 a 2009. ELCOMERCIO.ES 8 marzo 2017 03:20

El PP quiere bonificar los coches eléctricos en Avilés para reducir la contaminación.

Los populares proponen para el plan local de movilidad más líneas de bus y mejoras en la red semafórica y los pasos de cebra. La Nueva España - LNE 09.03.2017 | 03:49

Los ecologistas alertan de alta contaminación en toda la comarca avilesina.

La Nueva España - LNE 09.03.2017

- España emitió un 1,6% más de CO₂ en 2016 y la UE lo redujo. El nuestro es el sexto país que más emisiones hace en el bloque comunitario, con un 7,7% del total. Ep/Madrid 05.05.2017 | 03:41
- *Arcelor es "ineficaz" ante incidentes con contaminación, censura el Principado. La compañía, que afronta una multa de hasta dos millones por la nube tóxica, "no aborda ni lo previsto ni lo imprevisto", según Medio Ambiente. LNE 05.05.2017 | 03:07*

- La consejera de Medio Ambiente se enfrenta a una reprobación por los escapes de Arcelor. El cuestionamiento político a Belén Fernández dependerá del PP y Foro, que han polemizado reiteradamente sobre su gestión. LNE 05.05.2017 | 03:45
- Arcelor asegura que la fuga de gas ácido quince minutos y que actuó con «diligencia. La multinacional detalla en un nuevo informe las medidas de corrección ejecutadas y pide la inmediata reapertura de la planta. LNE 5 mayo 2017 | 03:58

El riesgo de morir en León por polución térmica es hasta seis veces mayor que en Madrid o Barcelona

LEONOTICIAS | A.C. @LEONOTICIAS_ 25 mayo 2017. 12:29

El riesgo de morir en Asturias por polución térmica es seis veces mayor que en Madrid

ELCOMERCIO.ES 25 mayo 2017. 02:24

El informe 'Un oscuro panorama. Los efectos en la salud de las centrales térmicas de carbón en 2014', llega a detallar el número de muertes de las que son responsables cada uno de los gases de combustión. El elemento más mortífero son las denominadas partículas en suspensión. A las 4.008 toneladas de estos compuestos que las térmicas lanzaron a la atmósfera le achacan 586 muertes prematuras, el 82% del total. El referido estudio, vincula

las micropartículas que emiten las centrales térmicas asturianas con 56 fallecimientos registrados en el 2014. Asturias concentra cinco de las dieciséis centrales térmicas que producen electricidad con la quema de carbón.

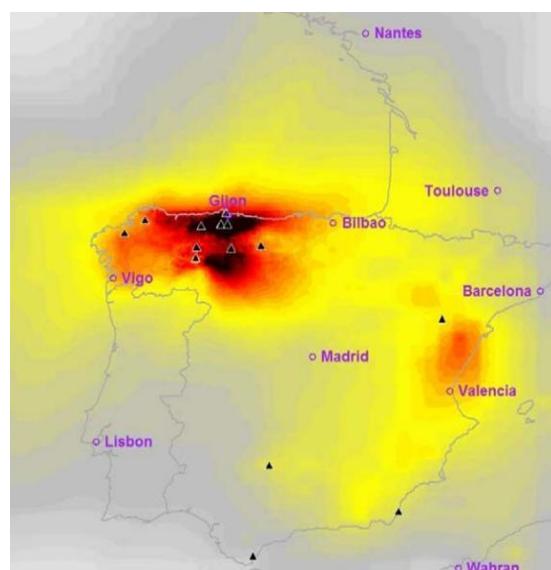


Figura 1.6. Concentración de partículas de dióxido de nitrógeno en España. Fuente: LEONOTICIAS | A.C.

El dato explica por qué si alguien analiza el perjuicio que causa a la salud este tipo de combustión, lo razonable es que la región se lleve la palma. La confirmación la aporta 'Un oscuro panorama' [5], estudio divulgado por el Instituto internacional de derecho y medio ambiente (IIDMA), organización no gubernamental con veinte años de actividad, declarada de utilidad pública, y que viene protagonizando toda una cruzada contra la hulla. El IIDMA echa mano de estudios de la OMS y de la Agencia Internacional de Energía, para detallar que el 82,6% de las 709 defunciones atribuibles a la quema de carbón lo serían por respirar más micropartículas PM_{2,5} de lo aconsejable.

1.3. Normativa aplicable en materia de contaminación atmosférica

La normativa vigente en materia de calidad del aire con carácter general es:

1.3.1. Normativa comunitaria

- Directiva 2004/107/CE del Parlamento Europeo y del Consejo, de 15 de diciembre de 2004, relativa al arsénico, el cadmio, el mercurio, el níquel y los hidrocarburos aromáticos policíclicos en el aire ambiente.
- Directiva 2008/50/CE del Parlamento Europeo y del Consejo, de 21 de mayo de 2008, relativa a la calidad del aire ambiente y a una atmósfera más limpia en Europa.
- Directiva 2010/75/UE, de 24 de noviembre de 2010, sobre las emisiones industriales
- Reglamento (UE) nº 1301/2013 del Parlamento y del Consejo de 17 de diciembre de 2013, sobre el FEDER.

1.3.2. Normativa estatal

- Real Decreto 833/1988, de 20 de julio, por el que se aprueba el Reglamento para la ejecución de la Ley 20/1986, básica de residuos tóxicos y peligrosos.

- Real Decreto 508/2007, de 20 de abril, por el que se regula el suministro de información sobre emisiones del Reglamento E-PRTR y de las Autorizaciones Ambientales Integradas.
- Ley 26/2007, de 23 de octubre, de Responsabilidad Ambiental.
- Ley 34/2007, de 15 de noviembre, de calidad del aire y protección de la atmósfera (referenciada en el punto 3.4.2. de esta tesis).
- Real Decreto 2090/2008, de 22 de diciembre, por el que se aprueba el Reglamento de desarrollo parcial de la Ley 26/2007, de 23 de octubre, de Responsabilidad Medioambiental.
- Real Decreto 100/2011, de 28 de enero, por el que se actualiza el catálogo de actividades potencialmente contaminadoras de la atmósfera y se establecen las disposiciones básicas para su aplicación.
- Real Decreto 102/2011, de 28 de enero, relativo a la mejora de la calidad del aire (modificado por RD 39/2017 y referenciado en el punto 3.4.2. de esta tesis).
- Ley 5/2013, de 11 de junio, por la que se modifican la Ley de prevención y control integrados de la contaminación y la Ley de residuos y suelos contaminados
- Real Decreto 815/2013, de 18 de octubre, por el que se aprueba el Reglamento de emisiones industriales.

1.3.3. Normativa autonómica

- Resolución de 1 de septiembre de 2009, de la Consejería de Medio Ambiente, Ordenación del Territorio e Infraestructuras, por la que se aprueba la Instrucción Técnica para la elaboración de Planes de Seguimiento de emisiones de gases de efecto invernadero en el Principado de Asturias.

- Estrategia institucional de Desarrollo Sostenible del Principado de Asturias [6], aprobada en Consejo de Gobierno de 20 de junio de 2002. Decreto 77/2012, de 14 de junio, por el que se establece la estructura orgánica básica de la Consejería de Fomento, Ordenación del Territorio y Medio Ambiente.
- Decreto 83/2013, de 1 de octubre, de primera modificación del Decreto 77/2012, de 14 de junio, por el que se establece la estructura orgánica básica de la Consejería de Fomento, Ordenación del Territorio y Medio Ambiente.
- Plan de Inspección Medioambiental del Principado de Asturias 2014-2017.
https://www.asturias.es/medioambiente/articulos/ficheros/Programa_inspeccion_2014.pdf
- Resolución de 4 de marzo de 2015, de la Consejería de Fomento, Ordenación del Territorio y Medio Ambiente. Declaración ambiental estratégica del programa operativo FEDER 2014-2020 del Principado de Asturias. Expte. IA-PP-0182/14.

Capítulo II

FUNDAMENTOS

Aprendizaje
automático para el
análisis de datos

2. Aprendizaje automático para el análisis de datos

El Aprendizaje Automático es una componente importante en campos como el análisis de datos o la minería de datos. Es además una rama de la inteligencia artificial cuyo objetivo consiste en desarrollar técnicas que permitan a las computadoras aprender. Es decir, mediante el aprendizaje automático se posibilita crear programas capaces de sistematizar comportamientos a partir de una información suministrada en forma de ejemplos.

2.1. Fundamentos biológicos del cerebro y transmisión de la información a través del sistema nervioso

El cerebro es la pieza más compleja del cuerpo humano, y realiza importantes funciones vitales necesarias para la vida. Se encuentra protegido por el cráneo y tiene la misma estructura general que el cerebro de otros mamíferos, pero es tres veces mayor con un tamaño corporal equivalente. A pesar de los numerosísimos estudios y avances que se han realizado a lo largo de los años la mecánica de funcionamiento del cerebro todavía no es del todo comprensible. En cualquier caso, los científicos trabajan para encontrar herramientas que puedan modelar su funcionamiento. Las Redes Neuronales Artificiales (RNA) están inspiradas en las redes neuronales biológicas del cerebro humano.

2.1.1. El Cerebro

Se estima que el cerebro humano contiene de 50 a 100 mil millones (10^{11}) de neuronas, de las que unos 10 mil millones (10^{10}) son células piramidales corticales. Estas células transmiten las señales a través de aproximadamente 1000 billones (10^{15}) de conexiones sinápticas [7].

Cada hemisferio cerebral tiene asignadas una serie de capacidades. El derecho tiene la capacidad para reconocer los símbolos numéricos y realizar aproximaciones o estimaciones matemáticas. El izquierdo tiene la capacidad de reconocer la escritura, esto probablemente está relacionado con su función lingüística; desde el punto de vista de procedimientos tiene la capacidad de realizar cálculos exactos como la multiplicación.

En la Figura 2.1 se pueden ver las distintas áreas funcionales del cerebro. Seguidamente se detallan las partes de que se compone el cerebro:

Cuerpo calloso. Los dos hemisferios cerebrales que dividen al cerebro, están separados por una profunda fisura, pero se mantienen unidos en su parte inferior por un haz de fibras nerviosas, de 10cm aproximadamente, llamado cuerpo calloso. Dicho cuerpo calloso permite la comunicación entre ambos hemisferios. Los hemisferios suponen cerca del 85% del peso cerebral.

Corteza cerebral. La corteza cerebral o sustancia gris, de 2 a 3 mm de espesor, está formada por capas de células amielínicas (sin vaina de mielina que las recubra). Debido a los numerosos pliegues que presenta, la superficie cerebral es unas 30 veces mayor que la superficie del cráneo. Estos pliegues forman las circunvoluciones cerebrales, surcos y fisuras y delimitan áreas con funciones determinadas, divididas en cinco lóbulos. Cuatro de los lóbulos del cerebro se denominan lóbulo frontal, parietal, temporal y occipital. El quinto lóbulo, la ínsula, no es visible desde fuera del cerebro y está localizado en el fondo de la cisura de Silvio. Los lóbulos frontal y parietal están situados delante y detrás, respectivamente, de la cisura de Rolando. La cisura parieto-occipital separa el lóbulo parietal del occipital y el lóbulo temporal se encuentra por debajo de la cisura de Silvio.

Sustancia blanca. La sustancia blanca, denominada también materia blanca, es una parte del sistema nervioso central compuesta de fibras nerviosas mielinizadas (cubiertas de mielina). Las fibras nerviosas contienen sobre todo muchos axones (un axón es la parte de la neurona encargada de la transmisión de información a otra célula nerviosa).

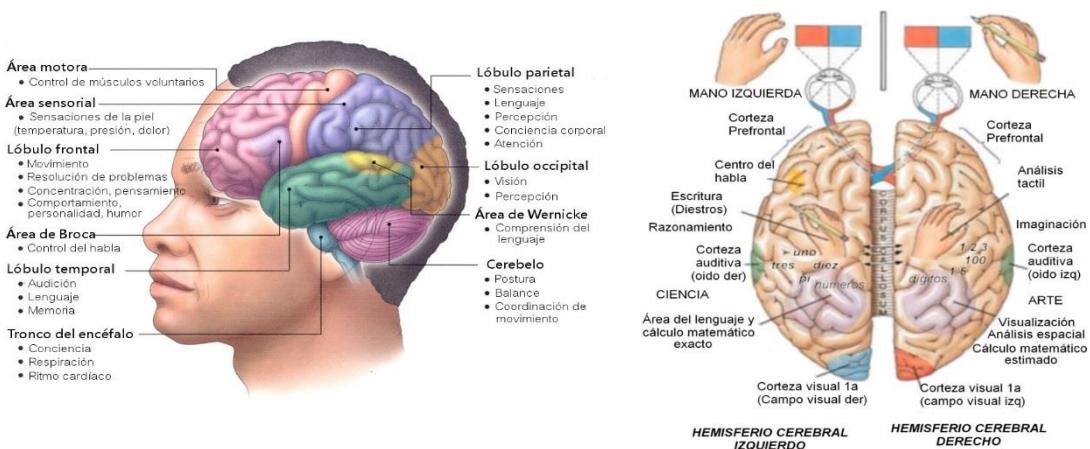


Figura 2.1. Áreas funcionales del cerebro. Fuente: <http://www.youbioit.com/es/article/informacion-compartida/20969/funciones-de-las-partes-del-cerebro>.

El cerebro humano ha inspirado modelos matemáticos y diversos algoritmos que sirven para modelizar procesos o comportamientos; no obstante, la capacidad de procesamiento y almacenamiento que puede realizar un cerebro humano supera, hoy en día, a las computadoras más eficientes.

La **paradoja de Moravec** es el descubrimiento en el campo de la inteligencia artificial y robótica de que, de forma anti-intuitiva, el pensamiento razonado humano (el pensamiento inteligente y racional) requiere de poca computación, mientras que las habilidades sensoriales y motoras, no conscientes y compartidas con otros muchos animales, requieren de grandes esfuerzos computacionales. Este principio fue postulado por Hans Moravec, Rodney Brooks, Marvin Minsky y otros en la década de los 80. Como Moravec dijo: «es fácil comparativamente conseguir que las computadoras muestren capacidades similares a las de un humano adulto en test de inteligencia, y difícil o imposible lograr que posean las habilidades perceptivas y motrices de un bebé de un año» [8].

2.1.2. El Sistema Nervioso

Podríamos decir que el sistema nervioso es el centro donde se procesan y se toman las decisiones que afectan a nuestro cuerpo. Dicho sistema recibe la información en base a unos estímulos que provienen de los receptores, ojos, nariz, boca, piel, órganos internos, etc., y, una vez procesada e interpretada, proporciona una respuesta adecuada a los estímulos recibidos.

Nuestro sistema nervioso, es uno de los más complejos del cuerpo. Está formado por un conjunto de órganos y una red de tejidos nerviosos cuya unidad básica son las neuronas. Para poder comprender su funcionalidad, consideramos que el citado sistema nervioso puede subdividirse en dos: el *sistema nervioso central (SNC)*, que está formado por el encéfalo y la médula espinal, y el *sistema nervioso periférico (SNP)*, que engloba los nervios craneales, espinales y los ganglios asociados.

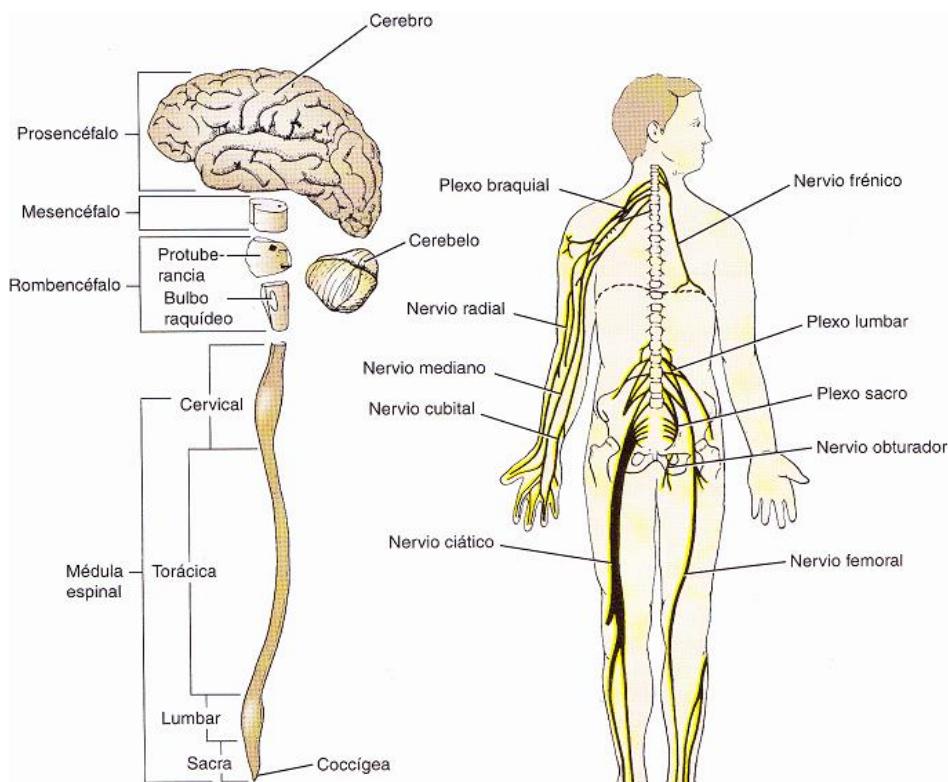


Figura 2.2. Principales divisiones del sistema nervioso central y Periférico.
(Fuente: sitio web neuroanatomia.info).

El mencionado sistema nervioso, interviene de forma directa en todas las funciones de nuestro organismo, al mismo tiempo que puede realizar tareas únicas que operan con total independencia del resto de los sistemas del cuerpo humano. El sistema nervioso tiene cuatro funciones principales:

1. Sensorial (obtener información del ambiente externo)
2. Integrador (reunir la información adquirida para interpretarla y evaluarla)
3. Efector (producir una respuesta)
4. Regulador interno (mantener una condición estable) [9]

La información que se recibe a través del sistema nervioso periférico, en forma de pulsos electroquímicos, se transmite a la medula espinal y llega hasta el centro del cerebro; para después ramificarse por toda la superficie del mismo. Estos impulsos son recibidos y procesados a razón de 100 millones por segundo, enviando otros de respuesta de vuelta a la medula espinal y de esta al sistema nervioso periférico.

2.2. Inteligencia Artificial (IA)

La inteligencia artificial (IA) facilita que las computadoras, robots y otros dispositivos, realicen trabajos que tradicionalmente han necesitado emplear la inteligencia humana. Estas tareas afectan al modo de resolver cierto tipo de problemas, así como a las capacidades necesarias para discriminar entre distintos objetos o responder a determinadas órdenes verbales. Las técnicas empleadas en inteligencia artificial (IA) imitan a los procedimientos inductivos y deductivos del cerebro, que se sirven de circuitos electrónicos que permiten la implementación de programas avanzados. El objetivo, basándose en la arquitectura de las redes neuronales humanas, es que una determinada máquina pueda realizar las mismas funciones que el cerebro humano.

2.2.1. Las neuronas

Las neuronas (del griego νεῦρον *neûron*, ‘cuerda’), son un tipo de células del sistema nervioso cuya principal función es la excitabilidad eléctrica de su membrana plasmática. Están especializadas en la recepción de estímulos y conducción del impulso nervioso (en forma de potencial de acción) entre ellas o con otros tipos celulares como, por ejemplo, las fibras musculares de la placa motora. La Figura 2.3 muestra el diagrama básico de una neurona.

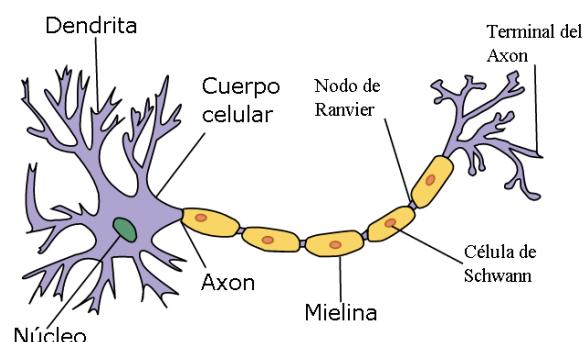


Figura 2.3. Diagrama básico de una neurona. (Fuente: https://en.wikipedia.org/wiki/Axon_hillock).

Altamente diferenciadas, la mayoría de las neuronas no se dividen una vez alcanzada su madurez; no obstante, una minoría sí lo hace. Presentan unas características morfológicas típicas que sustentan sus funciones: un cuerpo celular, denominado soma o pericarion central; una o varias prolongaciones cortas que generalmente transmiten impulsos hacia el soma celular, denominadas dendritas; y una prolongación larga, denominada axón o «cilindroide», que conduce los impulsos desde el soma hacia otra neurona u órgano diana [10].

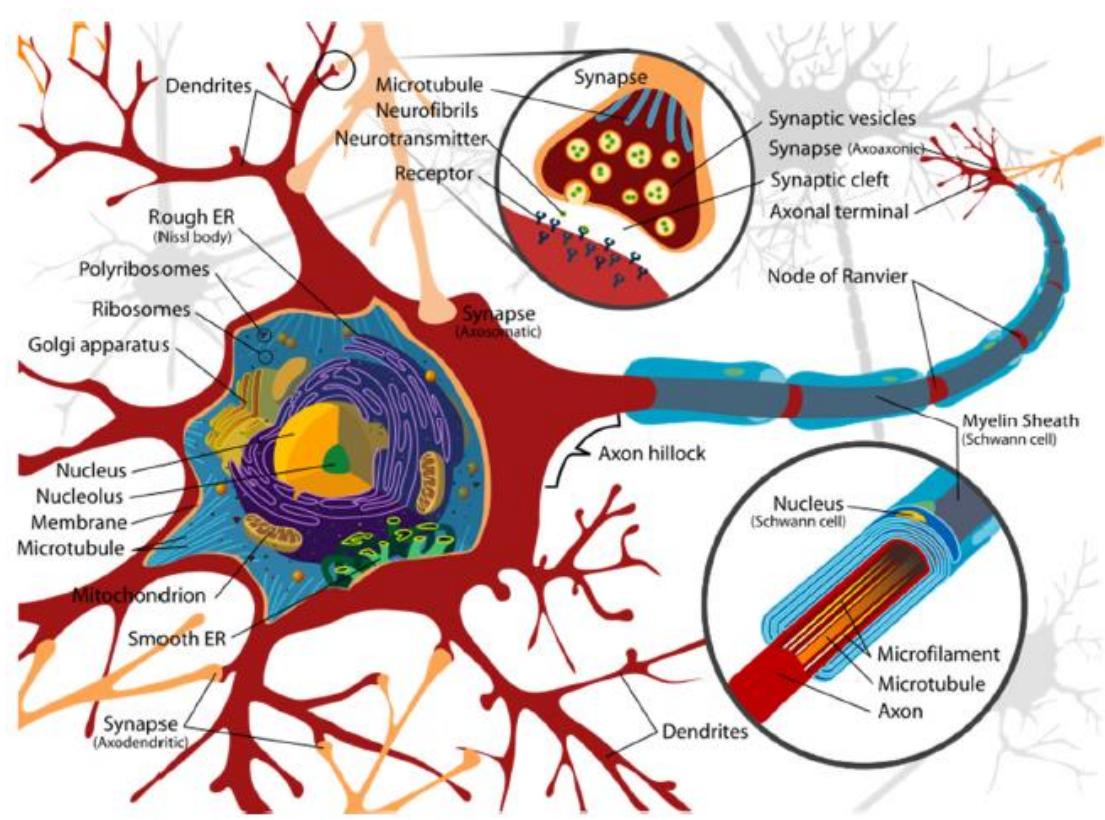


Figura 2.4. Partes de una neurona. Fuente: <https://es.wikipedia.org/wiki/Neurona>.

Las neuronas se comunican con precisión, rapidez y a larga distancia con otras células, ya sean nerviosas, musculares o glandulares. A través de las neuronas se transmiten señales eléctricas denominadas impulsos nerviosos. Estos impulsos viajan por toda la neurona hasta llegar a los botones terminales, que se pueden conectar con otra neurona, fibras musculares o glándulas. La conexión entre una neurona y otra se denomina sinapsis.

2.2.2. Arquitectura y funcionamiento de las redes neuronales artificiales

Las Redes Neuronales Artificiales (RNA) están inspiradas en las redes neuronales del cerebro humano. Están formadas por componentes que realizan funciones simples similares a las neuronas biológicas y se organizan como éstas en el cerebro. Las RNA se entrena para proporcionar una respuesta en función de sus entradas; es decir, “aprenden” mediante ejemplos y de sus propios errores. El conocimiento de una RNA se encuentra en los valores de los *pesos sinápticos* o de enlace, similar a las redes neuronales biológicas, donde el conocimiento de las mismas se encuentra en la *sinapsis* o conexiones entre ellas. Así, una RNA brinda una capacidad de procesamiento de información que es mucho mayor, que cualquier otro método de manipulación de datos. El tipo de supuestos más favorables para ser abordados mediante RNA son idénticos a los que una persona podría solucionar: los que impliquen asociación, reconocimiento de patrones y evaluación. Las RNA se diferencian de otras aplicaciones computerizadas en que no son algorítmicas; es decir, no se programan siguiendo una secuencia de instrucciones predefinida.

El cerebro está compuesto grandes cantidades de células llamadas *neuronas* con muchísimas conexiones llamadas *sinapsis* entre ellas. Tiene gran facilidad para procesar y manipular datos, en forma de pulsos electroquímicos, debido a un elevadísimo número de neuronas que cooperan entre sí. Esta gran capacidad de cómputo permite al cerebro el uso del lenguaje, la memoria y el control de los movimientos del cuerpo; dichas tareas, son difíciles de realizar para las maquinas.

Esto es aún imposible de modelar en una máquina, si consideramos que el cerebro puede emplear elementos más intangibles como la conciencia, los sentimientos y los sueños. De cualquier forma, el funcionamiento del cerebro inspira a muchos científicos en la búsqueda de herramientas que permitan modelarlo.

2.2.3. Inteligencia artificial compleja, mediante el desarrollo 'Software' que imite el funcionamiento del cerebro

Al igual que cada parte del cerebro humano se encarga de diferentes tareas (procesar el habla, distinguir los colores, los olores, las formas...), una red neuronal artificial – RNA, realizará una serie de labores diversas para cumplir su propósito. En este epígrafe vamos a tratar de determinar cuáles son las tareas que deberá poder desempeñar una RNA. Una red neuronal artificial deberá efectuar diversas operaciones de análisis y comparación de patrones predefinidos para poder establecer soluciones. Dicha red neuronal, se estructurará en diferentes capas de cálculo. La información, se recibirá por una entrada y el sistema procesará los datos, comparándolos con otros existentes, para así establecer las conclusiones. Por ejemplo, si se quiere determinar, mediante la foto de un animal, de qué animal se trata, ¿podría un ordenador saber si se trata de la foto de un gato, un lince, un perro u otro animal, por sí mismo? Mediante funciones matemáticas predefinidas, el sistema comenzará a analizar la imagen suministrada en diferentes capas: formas, colores, trazos; todos los parámetros de análisis serán comparados con otras fotos de animales. Así, se completará el estudio de todas las capas de análisis y se compararán todos los patrones estudiados, hasta llegar a una conclusión. El sistema, determinará por sí solo que se trata, por ejemplo, de un perro.

Emmanuel Mogenet, es el Jefe de Investigación de Google y responsable de un nuevo centro de investigación que se acaba de inaugurar en Zúrich, el cual está destinado a la rama de la ingeniería que trabaja en tres áreas clave:

- Máquina de Inteligencia
- Procesamiento del Lenguaje Natural y la comprensión
- La percepción de la máquina

Según las palabras del propio Emmanuel Mogenet, "un niño sabe, por ejemplo, que los coches no vuelan. Nadie se lo ha dicho ni se lo ha enseñado. Lo ha aprendido de forma indirecta en el día a día, viéndolo en la calle, estando en contacto con el entorno a través de sus sentidos. Eso es lo que llamamos 'percepción' y es lo que estamos desarrollando ahora en 'software' mediante la creación de redes neuronales artificiales".

2.2.4. La transformación del aprendizaje profundo

Citando a Wikipedia, "El aprendizaje automático es un subcampo de la informática que se desarrolló a partir del estudio de reconocimiento de patrones y la teoría del aprendizaje computacional en la inteligencia artificial. El aprendizaje automático explora el estudio y construcción de algoritmos que pueden aprender y hacer predicciones sobre los datos". Esté tipo de aprendizaje, se basa en redes neurales profundas (en inglés, *Deep Neural Nets - DNN*), que se asemejan a redes de neuronas biológicas. Los bloques de construcción de la DNN son los perceptrones, que son el equivalente a las neuronas. La información es procesada por una función de transferencia, resumiendo las señales de entrada, y una función de activación que decide si la salida está activa o no.

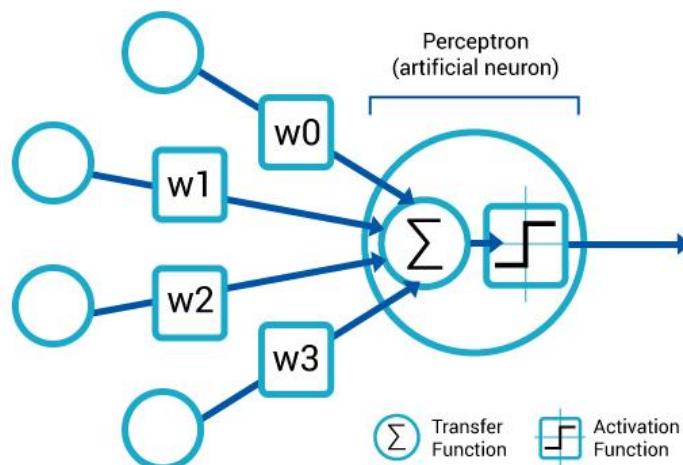


Figura 2.5. Activación de un perceptrón o neurona artificial. (Fuente: Dr. René Meyer, AMAX).

En una red neuronal artificial, un algoritmo ajusta los valores de las variables internas que conectan las neuronas digitalmente. Por ejemplo, un "cero" significa que las neuronas no están conectadas, mientras que un "uno" representa la conexión entre dos neuronas.

Con el tiempo, la red "aprende" a reconocer correctamente los patrones mediante el cálculo de los valores de los parámetros internos. En un sentido, la red escribe sus propias reglas implícitas cuando se le proporcionan los datos de entrada y de salida de las etiquetas. Una vez completada la fase de aprendizaje, el conjunto de valores y la red neuronal digitales se pueden trasladar al sistema que ejecuta la aplicación.

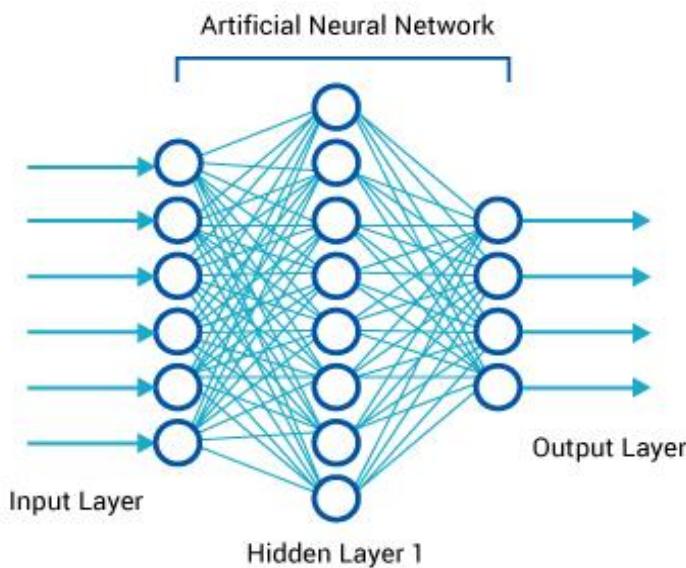


Figura 2.6. Red neuronal artificial de tipo perceptrón multicapa. (Fuente: Dr. René Meyer, AMAX).

Hasta hace poco, las redes neuronales artificiales no se implementaban de forma jerárquica y el tamaño se limitaba en función de los recursos de computación disponibles. Debido al reciente salto cuántico en el rendimiento de los ordenadores para tareas de cálculo, los datos de rendimiento que históricamente habían sido realizados por superordenadores ahora están disponibles para un amplio campo de los usuarios. Impulsado por este desarrollo, una nueva clase de algoritmos de aprendizaje profundo automático. Estos algoritmos están teniendo una popularidad creciente.

En contraste con los anteriores algoritmos de aprendizaje automático, los algoritmos de aprendizaje profundo, están constituidos por múltiples capas jerárquicas de neuronas digitales o capas de abstracción. Profundas, se refiere al gran número de capas ocultas en el DNN - *deep neural nets*.

Los beneficios de las redes neuronales jerárquicas persisten en mejorar su capacidad para reconocer los patrones resultantes y redundante en aplicaciones más fiables. Las redes neuronales artificiales imitan esa estructura del cerebro humano.

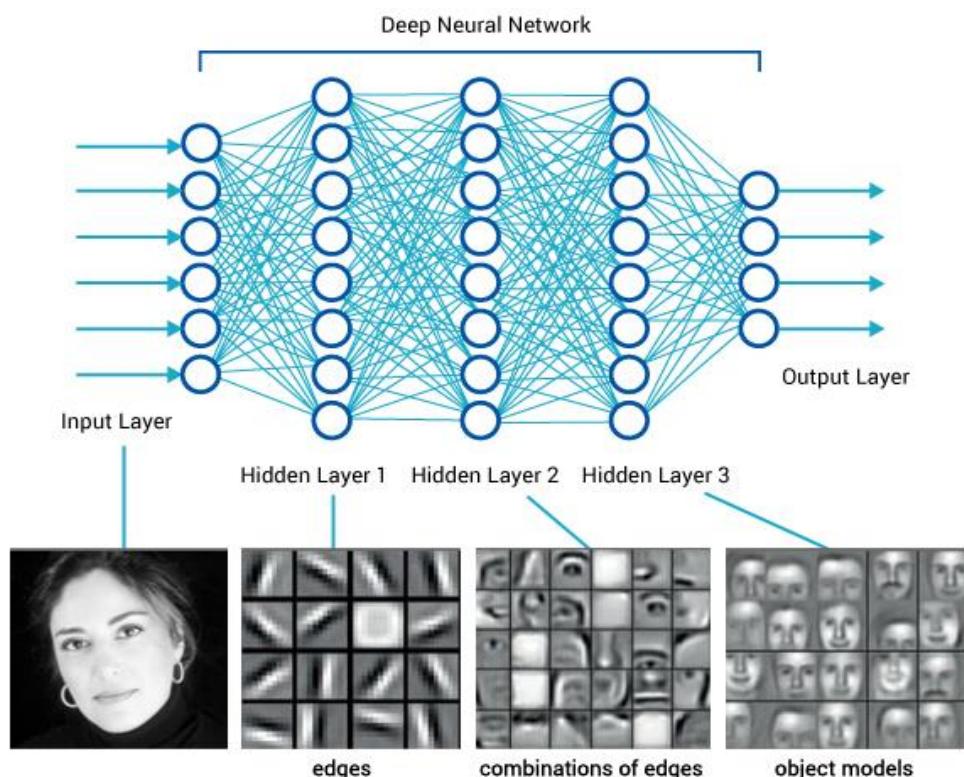


Figura 2.7. Red neuronal profunda, con tres capas ocultas. (Fuente: Dr. René Meyer, AMAX).

Greg Corrado, es un neurocientífico e investigador de Google que lidera desde Mountain View, California, buena parte del trabajo del buscador en el campo de la inteligencia artificial. Según sus propias explicaciones, nos describe cómo las redes neuronales artificiales tratan de imitar la estructura del cerebro: “En el cerebro humano hay miles de millones de neuronas individuales. Por sí solas no son muy listas, es el conjunto de neuronas trabajando entre sí lo que las hace tan poderosas”.

Según explica Tom Ruerig, ingeniero de Google, los objetivos a lograr en un futuro próximo serían similares a los que comenta en el siguiente ejemplo: "Imagina tener en un hospital una máquina a la que le envías una imagen, en alta definición, de lo que podría ser un tumor cancerígeno.

El sistema no solo la compararía en segundos con una base de millones de fotos, al mismo tiempo diagnosticaría de una forma similar a como lo hace un oncólogo, añadiendo una componente de 'intuición' basada en la experiencia profesional". El problema es que, por el momento, no lo consigue.

Por último, como reflexión sobre los peligros hacia los que estos imparables avances nos podrían llevar, cabe hacer la siguiente reflexión: ¿Se podría asegurar que las máquinas no tomarán decisiones erróneas o, peor aún, contra nosotros, sus propios creadores? En respuesta a esta pregunta, los principales investigadores a nivel mundial nos advierten de los peligros que se ciernen sobre la humanidad. "Con la inteligencia artificial nos estamos sometiendo al diablo", avisó Elon Musk, fundador de Tesla y SpaceX, en una conferencia hace ya dos años. Él y cientos de científicos e investigadores, entre ellos Stephen Hawking, Bill Gates, Jaan Tallinn o Nick Bostrom, están alertando sobre la creación de inteligencia artificial compleja. ¿Qué ocurrirá, por ejemplo, si empresas de armamento, como el gigante Lockheed Martin, comienzan a desarrollar armas con esta tecnología (si es que no lo están haciendo ya)? 1.000 investigadores, emprendedores, millonarios y científicos (con Musk entre ellos, y también el fundador de DeepMind, Demis Hassabis), firmaron en 2015 una carta para evitar que la inteligencia artificial caiga en manos de la industria armamentística. A juzgar por el pesimismo de algunos de sus firmantes, como Stephen Hawking, vamos por mal camino. En sus propias palabras: "El éxito en crear inteligencia artificial será el mayor evento en la historia de la humanidad. Desafortunadamente, puede que sea también el último".

2.3. El aprendizaje automático

El aprendizaje de máquinas o aprendizaje automático (en inglés, "Machine Learning") puede ser de dos clases: aprendizaje supervisado y aprendizaje no supervisado. Se puede pensar que para realizar la predicción, en la primera de las mencionadas clases se necesitaría la intervención humana mientras que en la segunda no; pero en realidad, estos dos conceptos estarían más relacionados con lo que se desea hacer con los datos.

2.3.1. Técnicas de Aprendizaje

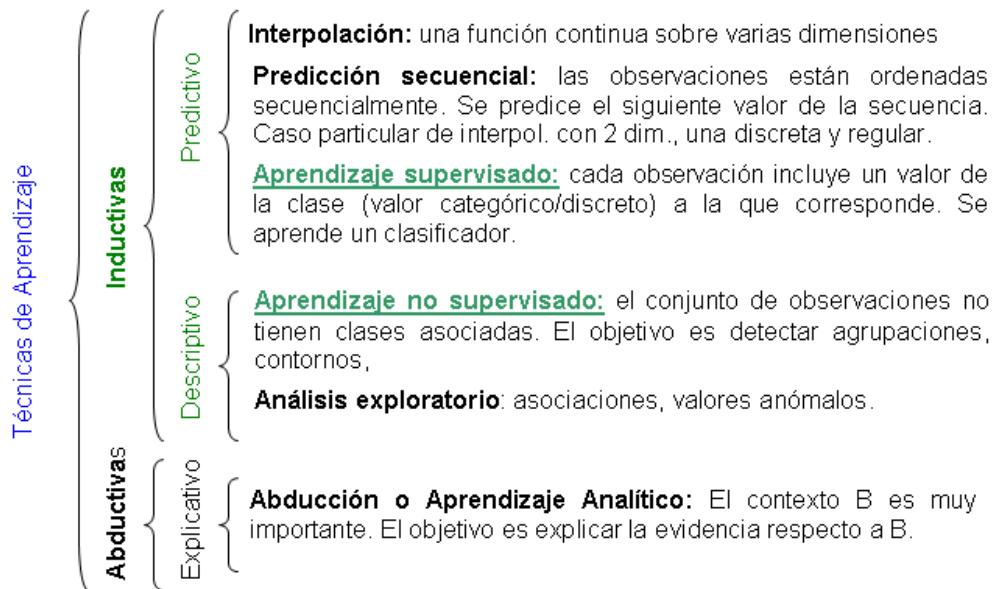
El objetivo del aprendizaje consiste en dotar a los sistemas de inteligencia artificial de la cualidad humana del aprendizaje, es decir, que dichos sistemas sean capaces de adquirir nuevos conocimientos a partir de las experiencias.

Dado un conjunto de ejemplos en la forma de pares (entradas, salidas), se puede decir que, el aprendizaje se basa en encontrar las reglas que mejor modelan la relación entre las entradas y las salidas. Es decir, que si disponemos de un conjunto de datos de la forma:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\} \quad (2.1)$$

Representando por los valores “x” las entradas, se trata de encontrar una hipótesis (función) que aproxime lo mejor posible el valor de “y” a partir de los valores de “x”. Este tipo de aprendizaje se basa en cierta manera en la memorización, pudiendo existir una limitación computacional debido a los recursos disponibles.

Si atendemos a la problemática que se presenta en la extracción del conocimiento de forma automática, podríamos clasificar las técnicas empleadas de aprendizaje según el siguiente esquema:



2.3.2. Aprendizaje mediante técnicas inductivas

Se basa en un razonamiento hipotético de casos particulares a casos generales. Creamos modelos de conceptos a partir de generalizar ejemplos simples. Buscamos patrones comunes que expliquen los ejemplos. Según el proceso se tienen los siguientes tipos de aprendizaje inductivo: Inducción supervisada e Inducción no-supervisada. En el aprendizaje por Inducción supervisada se aplican métodos como vecino más cercano (nearest neighbour), árboles de decisión (decision trees) y redes neuronales (neural networks); mientras que en el aprendizaje por Inducción no supervisada, se emplea el agrupamiento (clustering), análisis de dependencias y redes neuronales (neural networks). Dentro del *aprendizaje inductivo*, si se observa el esquema anterior, podemos ver que dicho aprendizaje inductivo se subdivide, a su vez, en predictivo y descriptivo. Perteneciendo el tipo de *aprendizaje supervisado* al tipo predictivo; mientras que el *aprendizaje no-supervisado* pertenecería al descriptivo. Seguidamente se describen las características de estos grupos y subgrupos:

PREDICTIVO: Dentro de la clasificación del aprendizaje Predictivo, estaría englobado el aprendizaje supervisado. Dependiendo de si se estima una función o una correspondencia, el aprendizaje Predictivo se podría subdividir, a su vez, en los siguientes términos:

Categorización: Entenderemos que se trata de una categorización, si se estima una correspondencia y las distintas clases se pueden solapar.

Ejemplo: determinar de entre un conjunto de temas dados, de qué temas trata una determinada página web partiendo de que cada página puede tratar varios temas.

Clasificación: Entenderemos que se trata de una clasificación, si se estima una función siendo las diferentes clases tratadas disjuntas.

Ejemplos: Determinar el grupo sanguíneo a partir de los grupos sanguíneos de los padres. Determinar si un compuesto químico es cancerígeno.

DESCRIPTIVO: Dentro de la clasificación del aprendizaje Descriptivo, estaría englobado el aprendizaje no-supervisado, del que más adelante detallaremos sus características. El aprendizaje descriptivo, consiste en extraer conclusiones a partir de un conjunto de ejemplos sin etiquetar, esto afectaría sólo a las entradas. Este tipo de aprendizaje, se basa en un análisis exploratorio que busca y asocia valores anómalos. En dicho análisis exploratorio, el conjunto de las exploraciones no tiene que tener necesariamente una o varias clases asociadas, sino que el objetivo de este método consiste en detectar asociaciones o contornos de dicho conjunto no etiquetado.

Aprendizaje supervisado

El aprendizaje supervisado se basa en un conjunto de técnicas, que nos permitirían inferir una función a partir de un conjunto de datos, empleados para el entrenamiento.

Los datos de entrenamiento son pares de vectores, de los cuales, una componente del par serían los datos de entrada y la otra los resultados esperados. Así, la salida de la función deducida sería, en los problemas de regresión un valor numérico y en los de clasificación una etiqueta de clase.

El objetivo del aprendizaje supervisado sería, por lo tanto, ser capaz de implementar una función que pueda predecir el valor correspondiente de un conjunto de datos de salida, después de haber entrenado con una serie de ejemplos, basados en datos apropiados. Para tal fin, tendría que poderse generalizar el resultado, partiendo de los datos proporcionados, sin que dichas situaciones se hayan dado previamente. Los datos que se van a emplear para el entrenamiento estarán basados en pares de modelos de entrada y de salida. El hecho de conocer al inicio del entrenamiento la salida, implica que se pueda supervisar el resultado obtenido. Es decir, en el aprendizaje supervisado, hay un “supervisor” que informa sobre lo que se debe hacer.

Aprendizaje no-supervisado

En el área del Aprendizaje automático, el aprendizaje no-supervisado se basa en encontrar estructuras desconocidas, a priori, partiendo de un conjunto de datos dado. En dicho conjunto, las estructuras específicas del conjunto son desconocidas, debido a la ausencia de un atributo que de alguna manera supervise su formación. Una de las técnicas aplicadas en el aprendizaje no-supervisado es el agrupamiento (en inglés ‘Clustering’).

El aprendizaje no-supervisado se aplica cuando se desconocen las etiquetas asociadas a los datos. La técnica del ‘Clustering’ es el proceso de agrupar los datos en clases, de tal forma que, los datos de un mismo clúster tendrán alta similitud y al mismo tiempo, serán distintos de los agrupados en otros ‘clusters’. Al hacer los ‘clusteres’, se pueden identificar regiones densas y regiones dispersas en el espacio de características, y por lo tanto, descubrir la distribución de patrones y la existencia de correlaciones entre los atributos. Aunque existen otros algoritmos específicos para ello, también se podrían detectar anomalías en un conjunto de datos dado. A diferencia de la clasificación, el ‘clustering’ o aprendizaje no supervisado no requiere clases predefinidas ni conjuntos de entrenamiento previos.

Por la razón comentada en el párrafo anterior, el aprendizaje que emplea el agrupamiento constituye un ejemplo de aprendizaje por observación, mientras que el tipo de aprendizaje en el que se emplea la clasificación constituye un aprendizaje por ejemplos. Para su aplicación se emplean métodos jerárquicos, basados en particiones, en densidad, en rejilla, en modelos y en distancia.

Un concepto importante, tanto para el aprendizaje supervisado (clasificación) como para el no supervisado (segmentación), es el concepto de *similitud*. La razón de este uso es que, intuitivamente, datos similares tendrán clases/grupos similares. Para medir la *similitud* nos basaremos en el siguiente concepto: *distancia* es inversa a *similitud*. Los métodos de *similitud* (o de distancia) se basan en almacenar los ejemplos vistos, y calcular la *similitud/distancia* del nuevo caso con el resto de ejemplos.

2.4. Redes neuronales. El perceptrón

Una *red neuronal artificial*, se basa en un sistema de interconexión de neuronas que colaboran para producir una salida. Son un paradigma de aprendizaje y procesamiento automático que se apoyan en la forma en que funciona el sistema nervioso de los animales. Fundándose en ellas, un grupo de investigadores de la Universidad Pablo de Olavide, han aplicado “técnicas basadas en la minería de datos para el descubrimiento de patrones en series temporales de origen sísmico y así poder predecir su ocurrencia”. Los científicos han empleado una serie redes neuronales artificiales para predecir terremotos en Chile y en la Península Ibérica [11]. Francisco Martínez, miembro del grupo TIC-200 Sistemas Inteligentes y Minería de Datos de la Universidad Pablo de Olavide explica además que “la metodología desarrollada podría extrapolarse a cualquier tipo de desastre natural, siempre que se aporten los datos necesarios”. En 1943, Warren McCulloch y Walter Pitts introdujeron una de las primeras neuronas artificiales. La característica principal de su modelo de neurona es que una suma ponderada de las señales de entrada se compara con un umbral para determinar la neurona de salida. Cuando la suma es mayor o igual al umbral, la salida es 1.

Cuando la suma es menor que el umbral, la salida es 0. A finales de 1950 Frank Rosenblatt y otros investigadores desarrollaron una clase de redes neuronales llamadas perceptrones. Las neuronas de estas redes eran similares a las de McCulloch y Pitts. La contribución clave de Rosenblatt fue la introducción de una regla de aprendizaje para la formación de redes perceptrón para resolver problemas de reconocimiento de patrones. Rosenblatt demostró que su regla de aprendizaje siempre convergirá a los pesos correctos de la red, si existen pesos que solucionan el problema [12].

El modelo biológico más simple de un *perceptrón* es una neurona y viceversa. Es decir, el modelo matemático más simple de una neurona es un *perceptrón*. La neurona es una célula especializada y caracterizada por poseer una cantidad indefinida de canales de entrada llamados dendritas y un canal de salida llamado axón.

Las dendritas operan como sensores que recogen información de la región donde se hallan y la derivan hacia el cuerpo de la neurona que reacciona mediante una sinapsis que envía una respuesta hacia el cerebro, esto en el caso de los seres vivos. Si considerásemos una sola neurona de manera aislada carecería de razón de ser. Sólo en la medida en que se asocie a otras neuronas, formando una red, su labor especializada se tornaría valiosa. Así, el axón de una neurona entrega su información como "señal de entrada" a una dendrita de otra neurona y así sucesivamente. El *perceptrón* que capta la señal en adelante se entiende formando una red de neuronas, sean éstas biológicas o de sustrato semiconductor (puertas lógicas).

El *perceptrón* dentro del campo de las redes neuronales tiene dos acepciones. Puede referirse a un tipo de red neuronal artificial desarrollado por Frank Rosenblatt (1959). Y dentro de la misma teoría de Frank Rosenblatt. También puede entenderse como la neurona artificial y unidad básica de inferencia en forma de discriminador lineal, es decir, un algoritmo capaz de generar un criterio para seleccionar un subgrupo, de un grupo de componentes más grande.

La limitación de este algoritmo consistiría en que si dibujáramos en un diagrama estos elementos, se deberían poder separar con un hiperplano los elementos "deseados" de los "no deseados". El perceptrón puede utilizarse con otros perceptrones u otro tipo de neurona artificial, para formar redes neuronales más complicadas.

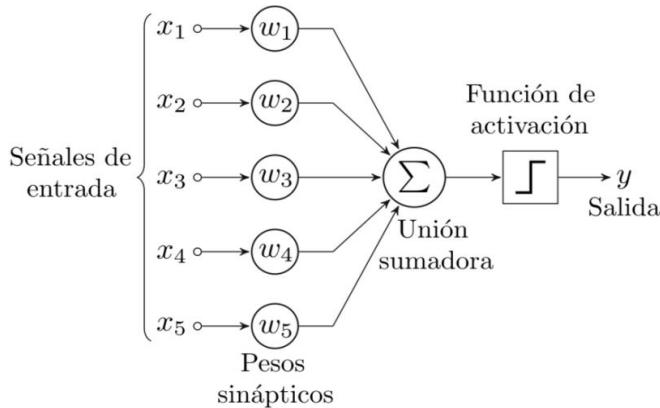


Figura 2.8. Diagrama de un perceptrón con cinco señales de entrada.

Fuente: Redes de Neuronas Artificiales, UC3M, RAI 2012.

El perceptrón emplea una matriz para representar las redes neuronales, y es un discriminador terciario que traza su entrada (x) (un vector binario), a un único valor de salida $f(x)$ (un solo valor binario) a través de dicha matriz.

$$f(x) = \begin{cases} 1 & \text{si } w \cdot x - u > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (2.2)$$

Donde “ w ” es un vector de pesos reales y “ $w \cdot x$ ” es el producto escalar (que computa una suma ponderada). “ u ” es el ‘umbral’, el cual representa el grado de inhibición de la neurona, es un término constante que no depende del valor que tome la entrada. El valor de $f(x)$ (0 o 1) se usa para clasificar “ x ” como un caso positivo o un caso negativo, en el caso de un problema de clasificación binario. El umbral puede pensarse de como compensar la función de activación, o dando un nivel bajo de actividad a la neurona del rendimiento. La suma ponderada de las entradas debe producir un valor mayor que “ u ” para cambiar la neurona de estado 0 a 1 [13].

2.4.1. Arquitectura y funcionamiento del Perceptrón

No existe, por el momento, una notación de red neural única y universalmente aceptada. La documentación sobre redes neuronales proviene de diversos campos, como la ingeniería, la física, la psicología y las matemáticas. Debido a su procedencia, muchos autores tienden a usar un vocabulario específico en relación a su especialización.

Los libros y los documentos redactados en este campo son difíciles de leer, y los conceptos parecen mucho más complejos de lo que son en realidad.

Para abordar los conceptos relacionados con esta explicación, vamos a adoptar la siguiente notación en las figuras, las ecuaciones matemáticas y el texto que se discuten:

Escalares - letras minúsculas en cursiva: a, b, c

Vectores – letras minúsculas en negrita: $\mathbf{a}, \mathbf{b}, \mathbf{c}$

Matrices – letras mayúsculas en negrita: $\mathbf{A}, \mathbf{B}, \mathbf{C}$

La Figura 2.9 ilustra un perceptrón de una sola capa, con una función de transferencia de límite duro simétrico.

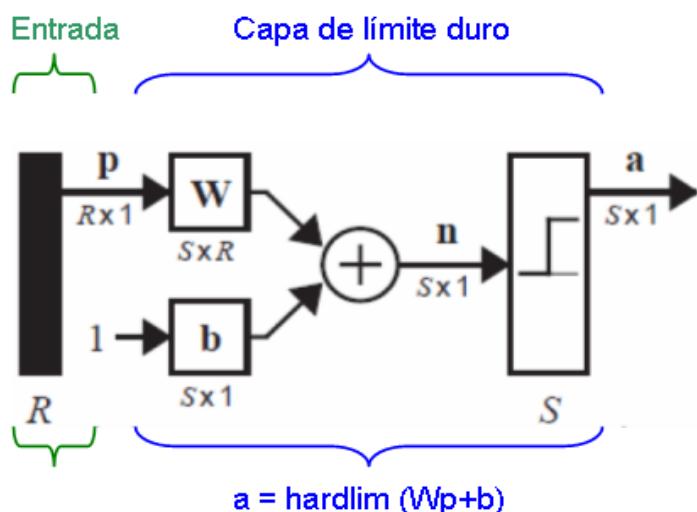


Figura 2.9. Red de Perceptrón de una sola capa. Fuente: adaptada de Neural Network Design. [Martin T. Hagan, Howard B. Demuth, Mark Beale – PWS Publishing Company]

La salida de la red viene dada por la expresión:

$$a = \text{hardlim}(Wp + b)$$

Esta expresión nos va a ser útil para desarrollar la regla de aprendizaje del Perceptrón para referenciar los elementos individuales de salida de la red. Primero, vamos a considerar la matriz de pesos de dicha red:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ w_{S,1} & w_{S,2} & \dots & w_{S,R} \end{bmatrix}. \quad (2.3)$$

Definimos un vector compuesto de la i-ésima fila de \mathbf{W} :

$$i^w = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ \vdots \\ w_{i,R} \end{bmatrix}. \quad (2.4)$$

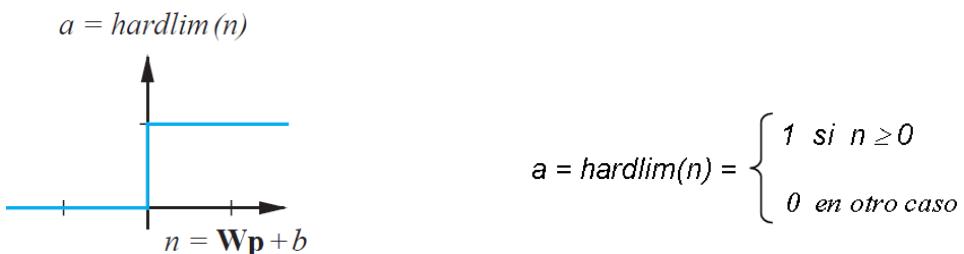
Dividimos la matriz de pesos:

$$W = \begin{bmatrix} {}_1 w^T \\ {}_2 w^T \\ \vdots \\ \vdots \\ {}_S w^T \end{bmatrix}. \quad (2.5)$$

Esto, nos permitirá escribir el i-ésimo elemento del vector de salida de la red como:

$$a_i = \text{hard lim}(n_i) = \text{hard lim}({}_i w^T p + b_i) \quad (2.6)$$

Conviene recordar, que la función de transferencia de la red (que se muestra junto a la expresión) viene dada por:



Consecuentemente, si el producto interno de la i -ésima fila de la matriz de peso con el vector de entrada es mayor o igual que $-bi$, la salida será 1, de lo contrario la salida será 0. Así, cada neurona de la red divide el espacio de entrada en dos regiones. Es útil investigar los límites entre estas regiones. Comenzaremos con el caso simple de un Perceptrón de una sola neurona con dos entradas. Dicho Perceptrón se muestra en la Figura 2.10.

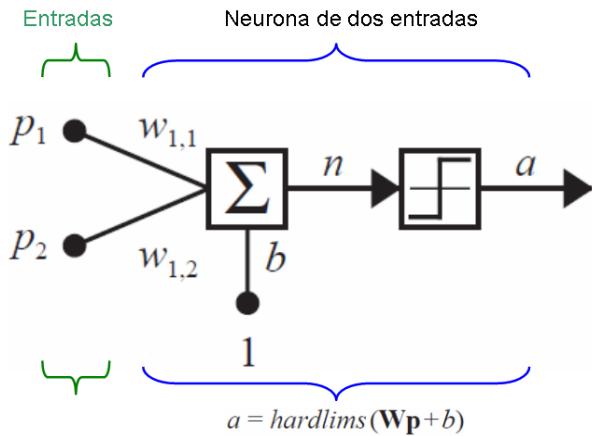


Figura 2.10. Perceptrón de dos entradas y una sola salida. Fuente: adaptada de Neural Network Design. [Martin T. Hagan, Howard B. Demuth, Mark Beale – PWS Publishing Company]

La salida de esta red viene determinada por:

$$\begin{aligned} a &= \text{hard lim}(n) = \text{hard lim}(Wp+b) \\ &= \text{hard lim}(_1w^T p + b) = \text{hard lim}(w_{1,1} p_1 + w_{1,2} p_2 + b) \end{aligned} \quad (2.7)$$

El límite de decisión está determinado por los vectores de entrada para los cuales la entrada neta n , es cero:

$$n = _1w^T p + b = w_{1,1} p_1 + w_{1,2} p_2 + b = 0 \quad (2.8)$$

Para que el ejemplo sea más concreto, asignaremos los siguientes valores a los pesos y a la variable b :

$$w_{1,1} = 1, w_{1,2} = 1, b = -1 \quad (2.9)$$

Los límites de decisión vendrán dados entonces por:

$$n = {}_1 w^T p + b = w_{1,1} p_1 + w_{1,2} p_2 + b = p_1 + p_2 - 1 = 0 \quad (2.10)$$

Esto define una línea en el espacio de entrada. En un lado de la línea, la salida de la red será 0; En la línea y en el otro lado de la línea la salida será 1. Para dibujar la línea, podemos encontrar los puntos donde intersecta los ejes p_1 y p_2 .

Para encontrar la intersección p_2 fijamos $p_1 = 0$:

$$p_2 = -\frac{b}{w_{1,2}} = -\frac{-1}{1} = 1 \quad \text{si } p_1 = 0 \quad (2.11)$$

Para encontrar la intersección p_1 fijamos $p_2 = 0$:

$$p_1 = -\frac{b}{w_{1,1}} = -\frac{-1}{1} = 1 \quad \text{si } p_2 = 0 \quad (2.12)$$

El límite de decisión resultante se ilustra en la Figura 2.11, a continuación:

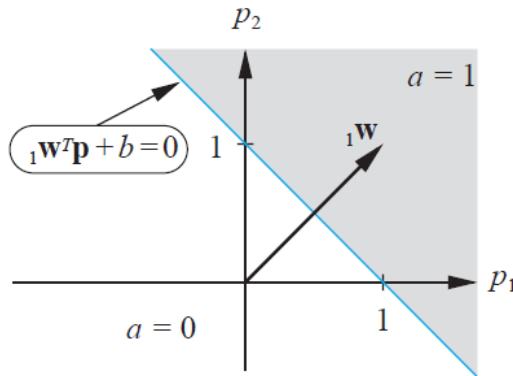


Figura 2.11. Límite de decisión para un Perceptrón de dos entradas. Fuente: Neural Network Design. [Martin T. Hagan, Howard B. Demuth, Mark Beale – PWS Publishing Company]

Para averiguar qué lado del límite corresponde a una salida de 1, sólo necesitamos probar con un punto. Para la entrada $p = [2 \ 0]^T$, la salida de red será:

$$a = \text{hard lim}({}_1 w^T p + b) = \text{hard lim}\left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} - 1\right) = 1 \quad (2.13)$$

Por lo tanto, la salida de red será 1 para la región de arriba y hacia la derecha del límite de decisión. Esta región está indicada por el área sombreada en la Figura 2.11 anterior.

Vamos a aplicar algunos de estos conceptos en el diseño de una red perceptrón para implementar una función lógica simple: la puerta AND. Los pares de entrada/destino para la puerta AND son:

$$\left\{ \begin{array}{l} p_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, t_1 = 0 \\ p_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, t_2 = 0 \end{array} \right. \left\{ \begin{array}{l} p_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, t_3 = 0 \\ p_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, t_4 = 1 \end{array} \right. \quad (2.14)$$

La Figura 2.12 ilustra el problema gráficamente. En dicha ilustración se muestra el espacio de entrada, con cada vector de entrada etiquetado de acuerdo a su objetivo. Los círculos oscuros “●” indican que el objetivo es 1, y los círculos claros “○” indican que el objetivo es 0.

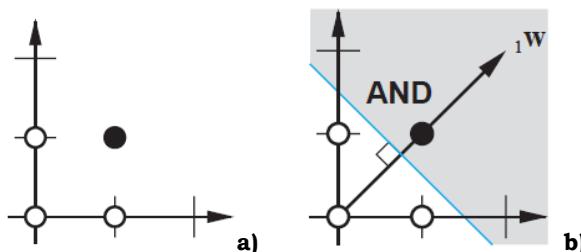


Figura 2.12. a) pares de entrada/destino para la puerta AND. b) elección de la línea que separa las dos categorías de entradas. Fuente: Neural Network Design. [Martin T. Hagan, Howard B, Demuth,Mark Beale – PWS Publishing Company]

El primer paso del diseño es seleccionar un límite de decisión. Queremos tener una línea que separe los círculos oscuros de los círculos claros. Hay un número infinito de soluciones a este problema, pero es razonable elegir la línea que está "a medio camino" entre las dos categorías de entradas, como se muestra en la Figura 2.12. A continuación, queremos elegir un vector de peso que sea ortogonal al límite de decisión. El vector de peso puede ser de cualquier longitud, por lo que hay posibilidades infinitas. Una opción es de la forma que se muestra en la Figura 2.12 b).

$$_1 W = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad (2.15)$$

Finalmente, necesitamos encontrar el sesgo, o la tendencia, b . Podemos hacerlo escogiendo un punto en el límite de decisión y satisfaciendo la siguiente ecuación. Si usamos encontramos $p=[1.5 \ 0]^T$, encontramos:

$$_1 w^T p + b = [2 \ 2] \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} + b = 3 + b = 0 \Rightarrow b = -3 \quad (2.16)$$

Ahora podemos probar la red en uno de los pares de entrada/destino. Si aplicamos \mathbf{p}_2 a la red, la salida será:

$$\begin{aligned} a &= \text{hard lim}(_1 w^T p_2 + b) = \text{hard lim}\left([2 \ 2] \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 3\right) \\ a &= \text{hard lim}(-1) = 0 \end{aligned} \quad (2.17)$$

que es igual a la salida objetivo t_2 . Deberíamos comprobar que todas las entradas están correctamente clasificadas.

2.4.2. Perceptrón simple y Adaline

Adaline (en inglés, ADaptive LINear Element) es un tipo de red neuronal artificial [B. Widrow & T. Hoff, 1960], basado en la Neurona de McCulloch-Pitts. Se compone de una sola capa de n neuronas con m entradas. Sabiendo que el Perceptrón es uno de los algoritmos de aprendizaje más antiguos y simples que existen, podemos considerar a Adaline como una mejora sobre Perceptron. La red neuronal Adaline y el Perceptron mantienen varias características en común: sirven para la clasificación binaria, tienen un límite de decisión lineal, pueden aprender de forma iterativa, ambos usan una función de umbral, etc. En cuanto a su estructura, está formada por un elemento denominado combinador adaptativo lineal (ALC) que obtiene una salida lineal (s) que pueda ser aplicada a otro elemento de conmutación bipolar, de forma que si la salida del ALC es positiva, la salida de la red ADALINE es $+1$; si la salida es negativa, entonces la salida de la red ADALINE es -1 .

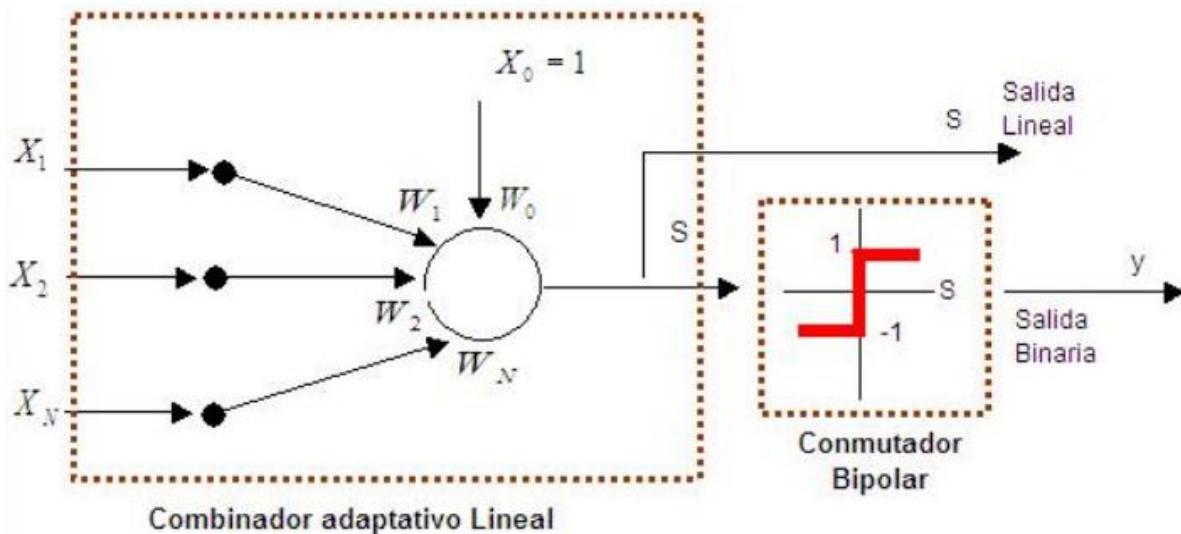


Figura 2.13. Estructura de la red neuronal artificial Adaline.

Fuente: <http://www.monografias.com/trabajos12/redneuro/redneuro2.shtml>

La red ADALINE utiliza un aprendizaje OFF LINE con supervisión denominado LMS (Least Mean Squared) o regla del mínimo cuadrado medio. También se conoce como regla delta porque trata de minimizar una delta o diferencia entre el valor observado y el deseado en la salida de la red. La salida considerada es el valor previo a la aplicación de la función de activación de la neurona. La regla aprendizaje de mínimos cuadrados es un método para hallar el vector de pesos W deseado, el cual deberá ser único y asociar con éxito cada vector del conjunto de vectores o patrones de entrada con su correspondiente valor de salida correcto (o deseado). El problema que se plantea consiste en encontrar un conjunto de pesos W que proporcione un vector de salida correcto para un único vector de entrada X.

El entrenamiento de la red consiste en adaptar los pesos a medida que se vayan presentando los patrones de entrenamiento y salidas deseadas para cada uno de ellos. Para cada combinación entrada-salida se realiza un proceso automático de pequeños ajustes en los valores de los pesos hasta que se obtienen las salidas correctas. Si definimos una función de umbral para hacer una predicción. Es decir, si z es mayor que un umbral θ , predecimos la clase 1, y 0 en caso contrario:

$$g(z) = \begin{cases} 1 & \text{si } z \geq \theta \\ -1 & \text{en otro caso} \end{cases}$$

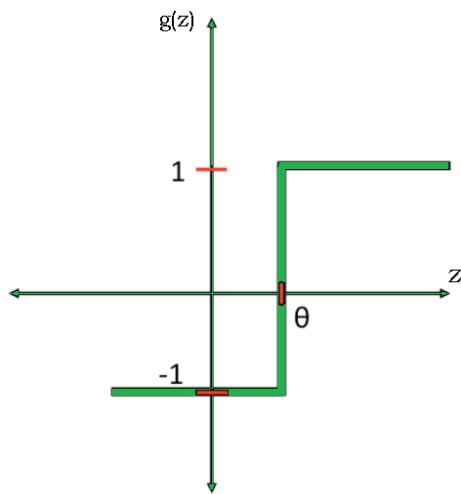


Figura 2.14. Gráfica de salida para la función de umbral. Elaboración

Las diferencias entre el Perceptrón y Adaline, se basan en que el Perceptrón usa las etiquetas de clase para aprender los coeficientes del modelo mientras que Adaline usa valores continuos pronosticados (de la entrada neta). Por tanto, en el perceptrón, como se ilustra a continuación, simplemente usamos las etiquetas de clase predichas para actualizar los pesos, y en Adaline, usamos una respuesta continua.

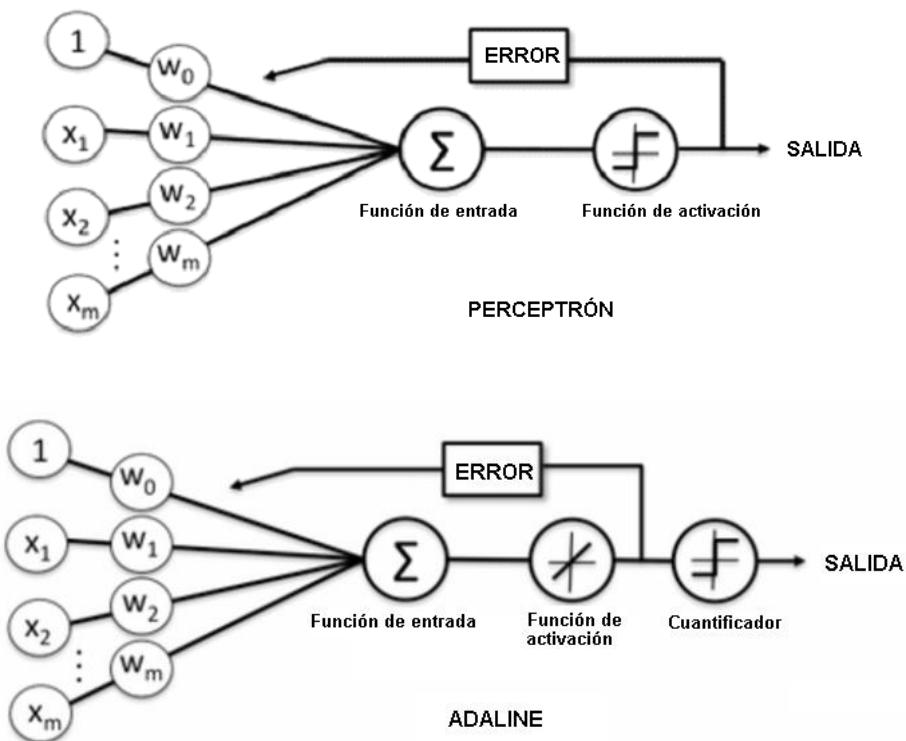


Figura 2.15. Diferencias entre el Perceptrón simple y Adaline. Fuente: S. Raschka, Quantitative Biology and Biochemistry and Molecular Biology Ph.D.- Michigan State University

2.4.3. Redes neuronales multicapa

En la Figura 2.6 se ilustra una red neuronal artificial de tipo Perceptrón Multicapa (MLP). Esta puede entenderse básicamente como una red de múltiples neuronas artificiales en múltiples capas. Aquí, la función de activación no es lineal (como en Adaline), pero utilizamos una función de activación no lineal como el sigmoide logístico (el que usamos en la regresión logística) o la tangente hiperbólica, o una función de activación lineal por partes como como la unidad rectificadora lineal (ReLU). Además, a menudo usamos una función softmax (una generalización del sigmoide logístico para problemas de clase múltiple) en la capa de salida, y una función de umbral para convertir las probabilidades predichas (por el softmax) en etiquetas de clase. Las redes neuronales artificiales pueden tener capas ocultas y diversas unidades de salida.

En este punto, se puede decir que la ventaja del MLP sobre el clásico Perceptron y Adaline es que al conectar las neuronas artificiales a través de funciones de activación no lineales, se pueden crear límites de decisión complejos, no lineales, que nos permiten abordar problemas donde las diferentes clases no son separables linealmente.

En la fase de entrenamiento de la red neuronal para la elección de los pesos iniciales se hace una asignación de pesos pequeños generados de forma aleatoria, en un rango de valores entre $-0,5$ y $0,5$ o algo similar.

Respecto a la arquitectura de la red neuronal artificial, se sabe que para la mayoría de problemas prácticos bastará con utilizar una sola capa oculta. El número de neuronas de la capa de entrada vendrá determinado por el número de variables predictoras. El número de neuronas ocultas determina la capacidad de aprendizaje de la red neuronal. Recordando el problema del sobreajuste, se debe usar el mínimo número de neuronas ocultas con las cuales la red rinda de forma adecuada. Esto se consigue evaluando el rendimiento de diferentes arquitecturas en función de los resultados obtenidos con el grupo de validación.

El valor de la tasa de aprendizaje (η) controla el tamaño del cambio de los pesos en cada iteración. Se deben evitar dos extremos: un ritmo de aprendizaje demasiado pequeño puede ocasionar una disminución importante en la velocidad de convergencia y la posibilidad de acabar atrapado en un mínimo local; en cambio, un ritmo de aprendizaje demasiado grande puede conducir a inestabilidades en la función de error, lo cual evitara que se produzca la convergencia debido a que se darán saltos en torno al mínimo sin alcanzarlo. Por tanto, se recomienda elegir un ritmo de aprendizaje lo más grande posible sin que provoque grandes oscilaciones. El factor momento (α) acelera la convergencia de los pesos. Suele tomar un valor próximo a 1 (por ejemplo, 0.9).

2.4.3.1. Función de activación de las neuronas ocultas y de salida.

Para activar las neuronas ocultas y de salida, hay dos formas básicas que cumplen esta condición: la función lineal o identidad y la función sigmoidal que puede ser logística o tangente hiperbólica.

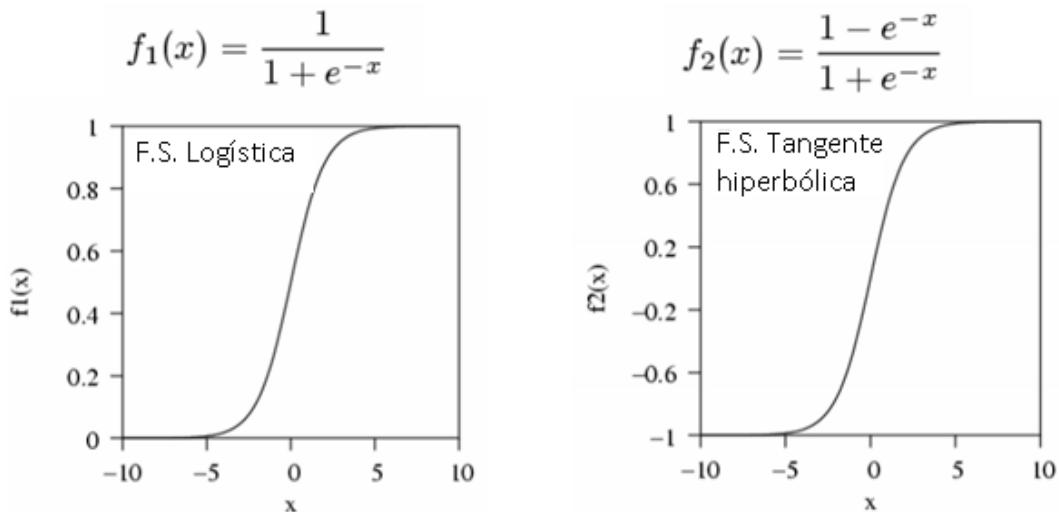


Figura 2.16. Funciones de activación sigmoidal empleadas en redes neuronales artificiales MLP.
Fuente: Redes neuronales, José M^a Valls, Universidad Carlos III de Madrid, 2007.

Para aprovechar la capacidad de las redes neuronales de aprender relaciones complejas o no lineales entre variables, se recomienda la utilización de funciones no lineales al menos en las neuronas de la capa oculta.

En general se utilizará una función sigmoidal, logística o tangente hiperbólica, como función de activación en las neuronas de la capa oculta. La elección de la función de activación en las neuronas de la capa de salida dependerá del tipo de tarea impuesto. En tareas de clasificación, las neuronas normalmente toman la función de activación sigmoidal. En cambio, en tareas de predicción o aproximación de una función, generalmente las neuronas toman la función de activación lineal.

2.4.3.1. Perceptrón multicapa (MLP).

La propagación hacia atrás de errores o retropropagación (del inglés backpropagation) es un algoritmo de aprendizaje supervisado que se usa para entrenar redes neuronales artificiales. Emplea un procedimiento para ajustar repetidamente los pesos con el fin de minimizar la diferencia entre la salida real y la salida deseada. Las capas ocultas, en una red neuronal del tipo perceptrón multicapa (MLP), son nodos de neuronas apilados entre las entradas y las salidas, que permiten a las redes neuronales artificiales ‘aprender’ características más complicadas. Las capas ocultas en una red de MLP simplemente agregan más neuronas entre las capas de entrada y salida.

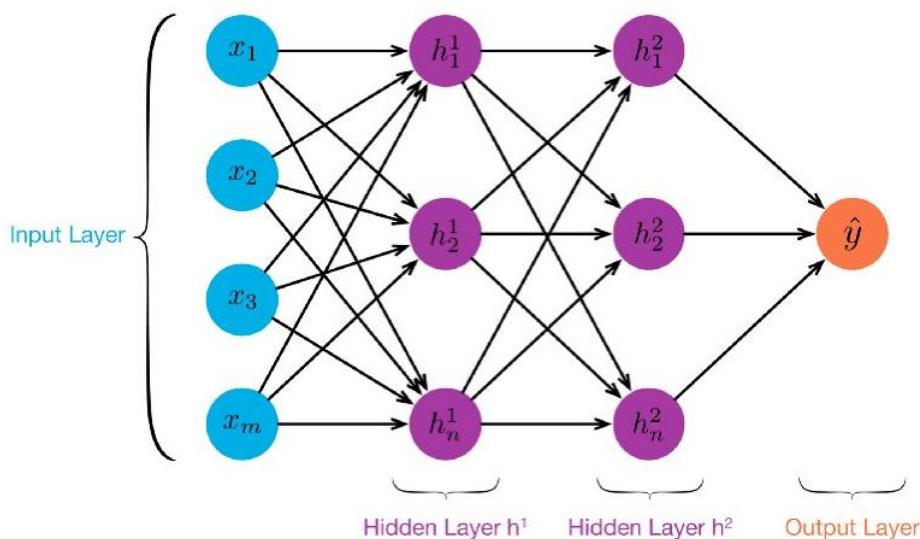


Figura 2.17. Red de perceptrón multicapa (MLP), con una capa oculta.
Fuente: Nahua Kang, Stockholm School of Economics

Los datos en la capa de entrada se etiquetan como x con los subíndices 1, 2, 3, ..., m . Las neuronas en la capa oculta se etiquetan como h con los subíndices 1, 2, 3, ..., n . Se debe notar que para la capa oculta es n y no m , ya que el número de neuronas de capa ocultas puede diferir del número en datos de entrada. Y como puede ver en la Figura 2.17, las neuronas de capa oculta también están etiquetadas con superíndice 1. Esto es para que cuando tenga varias capas ocultas, pueda identificar qué capa oculta es: primera capa oculta tiene superíndice 1, segunda capa oculta tiene superíndice 2, y así sucesivamente.

Los datos de entrada de una red neuronal MLP se corresponden con las distintas características, si tenemos m datos de entrada (x_1, x_2, \dots, x_m), llamamos a esto m características. Una característica es solo una variable que consideramos que tiene influencia en un resultado específico. Cuando multiplicamos cada una de las m características con un peso (w_1, w_2, \dots, w_m) y las sumamos todas juntas, se trata de un producto escalar:

$$W \cdot X = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_m x_m = \sum_{i=1}^m \omega_i x_i \quad (2.17)$$

De acuerdo con lo dicho en el párrafo anterior, con m funciones en la entrada X , se necesitan m pesos para realizar un producto de puntos. Del mismo modo, con n neuronas ocultas en la capa oculta, se necesitan n conjuntos de pesos (W_1, W_2, \dots, W_n) para realizar los productos.

Así, en una red neuronal del tipo MLP con 1 capa oculta se realizan los n productos para obtener la salida oculta h : (h_1, h_2, \dots, h_n). Por lo tanto, funcionaría como una red de perceptrón de una sola capa, utilizamos la salida oculta h : (h_1, h_2, \dots, h_n) como datos de entrada que tienen n características y realizamos un producto con 1 conjunto de n pesos (w_1, w_2, \dots, w_n) para obtener el resultado final.

El procedimiento de cómo los valores de entrada se propagan hacia adelante a la capa oculta, y luego de la capa oculta al resultado se puede ver en la Figura 2.18.

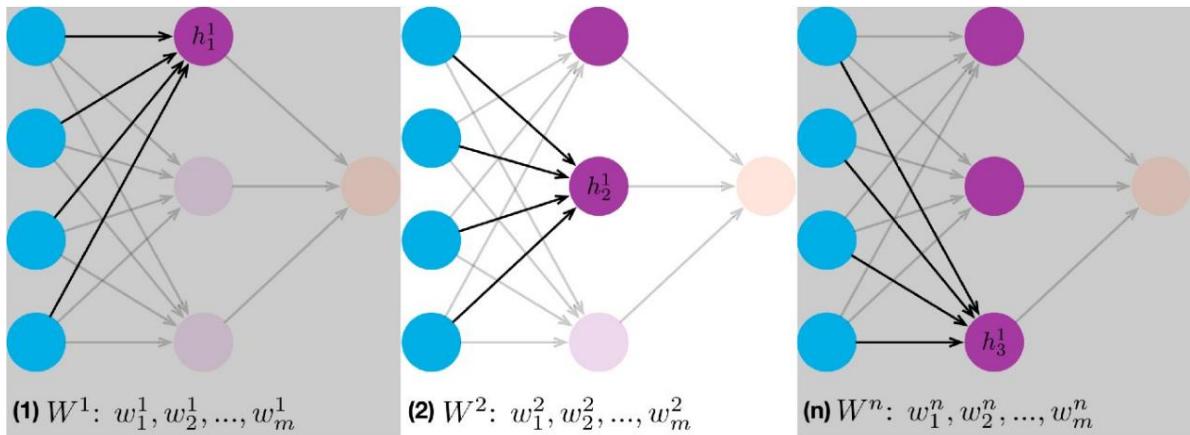


Figura 2.18. Propagación de los valores de entrada hacia delante, en la capa oculta. Fuente: Nahua Kang, Stockholm School of Economics

Una vez obtenidos los resultados de la capa oculta se calcula el resultado final, utilizando las salidas de la capa oculta como si fuesen entradas. A la vista de lo expuesto, se puede aproximar el funcionamiento de un perceptrón con múltiples capas al de un perceptrón de una sola capa. La principal diferencia es que durante el proceso intervienen muchos más pesos. Cuando se entrena redes neuronales en conjuntos de datos más grandes con muchas más características, este proceso consumirá mucha memoria en el ordenador. Esta ha sido una de las razones por las cuales el aprendizaje profundo (en inglés, deep learning) no despegó hasta los últimos años, cuando se comenzó a producir hardware mucho más evolucionado, que permite manejar las redes neuronales profundas que necesitan recursos especializados de Hardware y programas (Software) que permitan optimizar dichos recursos.

2.4.3.1.1. Funciones sigmoideas

Tenemos una red neuronal estructurada más sofisticadamente con capas ocultas. Pero no hemos resuelto el problema de activación con la función de paso. Conviene recordar que, si la función de activación es lineal, puede apilar tantas capas ocultas en la red neuronal como deseé, pero la salida final sigue siendo una combinación lineal de los datos de entrada originales. Mientras tanto, la función de paso tampoco tiene una derivada útil (su derivada es 0 en todas partes o indefinida en el punto 0 en el eje x). No se puede emplear para ‘backpropagation’.

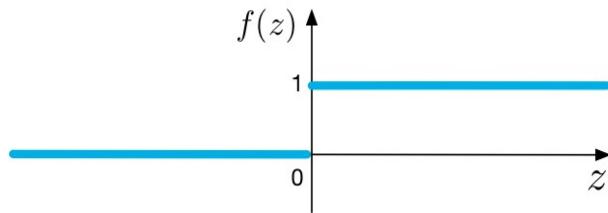


Figura 2.19. Función de paso. Fuente: Nahua Kang, Stockholm School of Economics.

El perceptrón con función de paso no tiene un comportamiento muy "estable". Se puede decir que, básicamente, un pequeño cambio en cualquier peso en la capa de entrada de nuestra red neuronal podría llevar a que una neurona voltee repentinamente de 0 a 1, lo que podría afectar nuevamente el comportamiento de la capa oculta y luego afectar el resultado final. Puesto que, queremos un algoritmo de aprendizaje que pueda mejorar nuestra red neuronal mediante el cambio gradual de los pesos, no por un salto sin respuesta o saltos repentinos, no podemos usar la función de paso para cambiar gradualmente los pesos. Para tal fin, emplearemos una función sigmoidea.

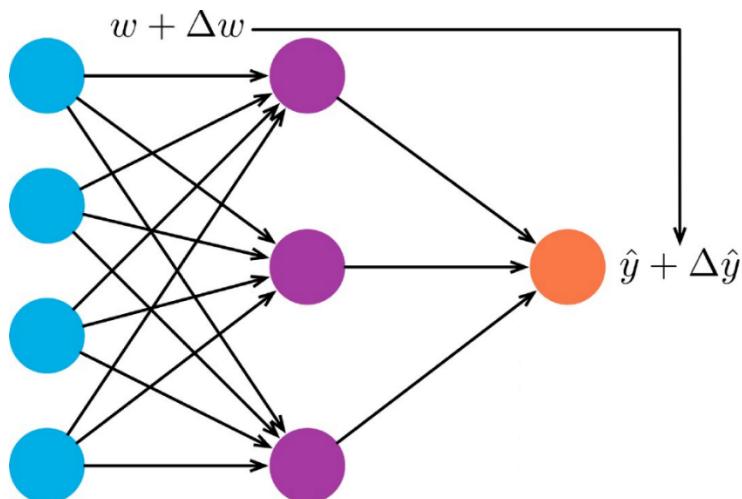


Figura 2.20. Esquema de función para cambiar gradualmente los pesos.
Fuente: Nahua Kang, Stockholm School of Economics

Sin necesidad de emplear conceptos matemáticos demasiado profundos, lo más importante a tener en cuenta sobre la función sigmoidea (Figura 2.21) es su curva, y su derivada.

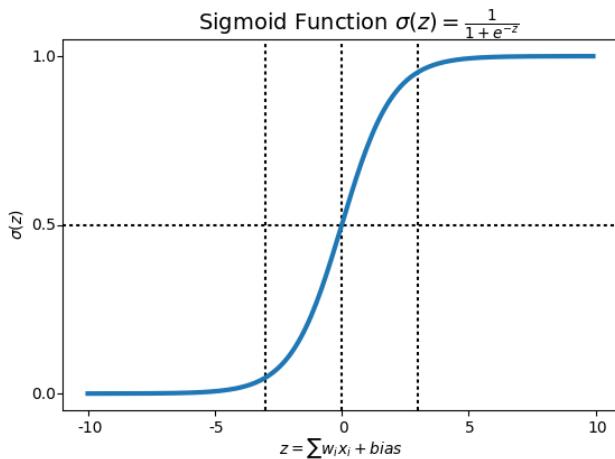


Figura 2.21. Función sigmoide. Fuente: Nahua Kang, Stockholm School of Economics.

- La función sigmoidea proporciona resultados similares a la función escalonada porque la salida está entre 0 y 1. La curva cruza 0.5 en $z = 0$, y se podrían configurar reglas para la función de activación, tales como: si la salida de la neurona sigmoidea es mayor que o igual a 0.5, emita '1'; o que si la salida es menor que 0.5, emita '0'.
- La función sigmoidea tiene una curva progresiva y suave con una derivada simple de $\sigma'(z) = \sigma(z) * (1-\sigma(z))$, que es diferenciable en todas partes en la curva.
- Si z es muy negativo, entonces la salida es aproximadamente 0; si z es muy positivo, la salida es aproximadamente 1; pero, alrededor de $z = 0$ donde z no es demasiado grande ni demasiado pequeña, entre las dos líneas verticales de cuadrículas con puntos, en la Figura 2.21 tenemos relativamente más desviación a medida que z cambia.

La función sigmoidea, a diferencia de la función de paso, introduce la no linealidad en nuestro modelo de red neuronal. No lineal significa que la salida que obtenemos de la neurona, que es el producto escalar de algunas entradas x (x_1, x_2, \dots, x_m) y pondera w (w_1, w_2, \dots, w_m) más sesgo y luego se lleva a una función sigmoidea, no puede ser representada por una combinación lineal de la entrada x (x_1, x_2, \dots, x_m). Esta función de activación no lineal, cuando es utilizada por cada neurona en una red neuronal multicapa, produce una nueva "representación" de los datos originales, y finalmente permite un límite de decisión no lineal.

2.5. KDD y Minería de Datos

El aumento del volumen de información computerizada en bases de datos digitales; así como la variedad de dicha información, ha ido aumentando de forma espectacular en nuestros tiempos. Mucha de esta información es histórica, y, representa servicios o escenarios que se han producido en el pasado y se han ido registrando a través de los sistemas informáticos disponibles. Esta información está empezando a tener un valor mucho más importante que la de un simple registro, y se está teniendo muy en cuenta debido a su utilidad para inferir información futura. La mayoría de decisiones que toman los responsables de las empresas e instituciones se basan en información de experiencias pasadas extraídas de fuentes muy diversas. A diferencia de las decisiones personales, las decisiones colectivas suelen tener consecuencias mucho más graves, especialmente económicas, y, recientemente, se deben basar en volúmenes de datos que desbordan la capacidad humana.

2.5.1. KDD – ‘Descubrimiento de Conocimiento a partir de Bases de Datos’

El ‘Descubrimiento de Conocimiento a partir de Bases de Datos’ (KDD, en inglés Knowledge Discovery from Databases), es un proceso muy amplio dentro del cual se encuadra la Minería de Datos. Dicho proceso surge de las necesidades de toma de decisión que requieren el descubrimiento de nuevos modelos, no esperados o imposibles de descubrir manualmente, a partir de una magnitud de datos muy grande. Este descubrimiento, además, deberá requerir la mínima pericia posible por parte del usuario.

El KDD, es un Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos [Fayyad et al. 1996] [14]. En la Figura 2.13, se puede apreciar un croquis del proceso KDD.

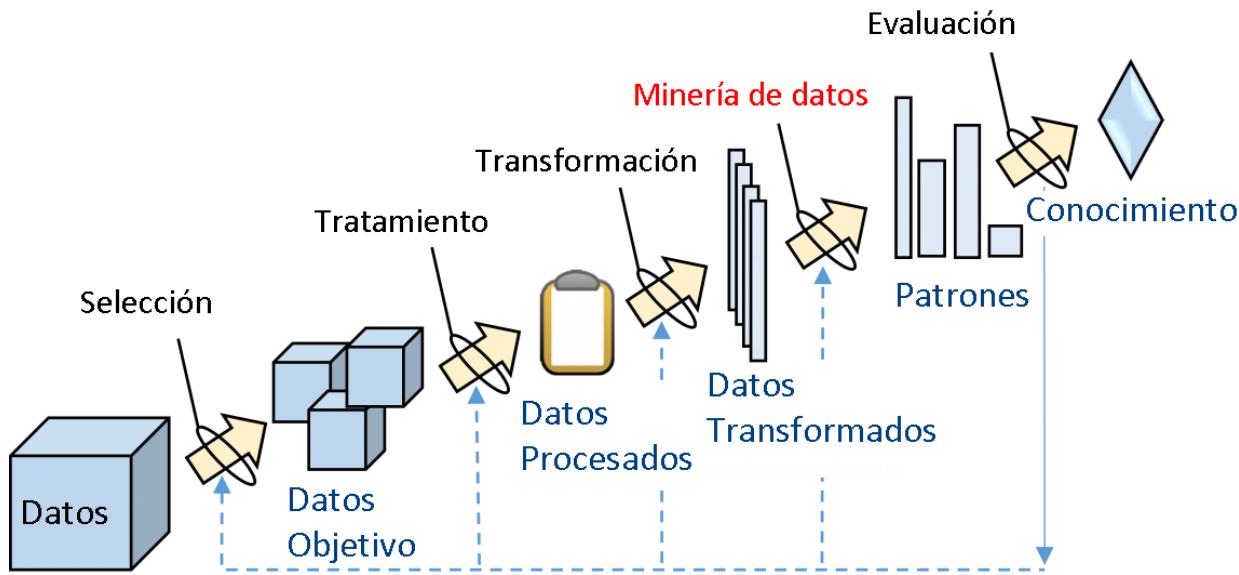


Figura 2.22. Proceso de KDD (Knowledge Discovery from Databases). Fuente: <https://behavior.lbl.gov/?q=node/11>

2.5.2. Minería de Datos

En determinadas situaciones, muchas de las decisiones importantes, los datos se convierten en conocimiento en base a métodos tradicionales; es decir, se realizan interpretaciones y análisis de forma manual. Un especialista en la materia, analiza los datos y elabora un informe u una hipótesis que refleja las tendencias o pautas derivadas de los datos analizados. Esta forma de actuar es lenta, cara y muy subjetiva; siendo imprácticable, donde la enorme abundancia de datos desborda la capacidad humana y es imprescindible la ayuda de herramientas computerizadas para poder interpretarlos. Por tanto, sin las herramientas adecuadas, se deberán tomar decisiones importantes, en lugar de sobre gran cantidad de información disponible, siguiendo la propia intuición del usuario.

La Minería de datos (en inglés '*Data mining*') es una técnica de análisis y estudio de datos que está empleándose en los últimos años, motivado fundamentalmente por la posibilidad de aplicar el potencial de procesamiento de datos de los ordenadores.

Es una técnica para el análisis y procesado de grandes volúmenes de datos con el fin de extraer información útil y patrones de fácil comprensión, que de otra forma sería imposible conseguir por los medios y herramientas estadísticas tradicionales. Seguidamente se dan unas definiciones sobre Minería de Datos, según los autores que se indican:

- Análisis de grandes volúmenes de datos para encontrar relaciones no triviales, y para resumirlos de manera que sean entendibles y útiles [Hand, Mannila y Smyth, 2001].
- Extracción de patrones y modelos interesantes, potencialmente útiles y datos en base de datos de gran tamaño [Hand, 2001].
- Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos [Witten and Frank, 2000].

Según esto el principal cometido de la Minería de Datos, consistirá en resolver problemas analizando los datos existentes en aquellas bases de datos de las que disponga la organización. Los datos por sí solos tienen valor relativo en muchos contextos, lo que resulta verdaderamente interesante es el conocimiento que puede inferirse a partir de dichos datos y, muy especialmente, la capacidad de poder emplear el conocimiento extraído. El objetivo de la minería de datos es identificar un conocimiento obtenido a partir de las bases de datos en estudio, para tomar decisiones. Para tal fin, se exploran y analizan grandes cantidades de datos con el objeto de encontrar patrones y reglas significativas; es decir, conocimiento. Los retos fundamentales de la minería de datos son dos, por un lado, trabajar con grandes volúmenes de datos procedentes de sistemas de información, con los problemas que de ello se deriven (ruido, falta de datos, intratabilidad, volatilidad de los datos...), y, por otro lado, emplear las técnicas más adecuadas para analizar los mismos y extraer conocimiento útil y novedoso.

Se puede entender que la minería de datos se encuentra encuadrada dentro de un campo multidisciplinar, que emplea técnicas estadísticas y de inteligencia artificial. Las aplicaciones susceptibles de aplicar técnicas de minería de datos se dan a continuación:

- Clasificación. Examinar las características de un nuevo objeto y asignarle una clase o categoría de acuerdo a un conjunto de tales objetos previamente definido.
- Clustering. Clasificar por "tipos" según las distintas percepciones de determinados atributos, con el fin de dividir una población en un número de grupos más homogéneos.
- Descubrimiento de Reglas de Asociación. Determinar cosas u objetos que van juntos
 - Análisis de dependencia de datos
 - Correlación y causalidad
- Descubrimiento de Patrones Secuenciales
 - Análisis de series de tiempo y asociaciones secuenciales
- Regresión. Procesar conjuntos de datos para estimar las relaciones entre sus variables.
- Tendencia y Desviación de detección. Aislar componentes que determinan el comportamiento a largo plazo de una serie y detectar las desviaciones e irregularidad en el cumplimiento de unas determinadas condiciones
- Filtros Colaborativos. Filtrar información o modelos, que emplean técnicas de colaboración entre múltiples agentes o conjuntos de referencias muy grandes para la detección y control de datos.
- Descripción de Conceptos
 - Descripción de características
 - Descripción discriminante

Como conclusión de este apartado, entendemos que el resultado de aplicar técnicas de minería de datos serán conjuntos de reglas, ecuaciones, árboles de decisión, redes neuronales, grafos probabilísticos y otros medios; los cuales podrán emplearse para responder a preguntas como ¿existe un grupo de clientes que se comporta de forma singular?, ¿cuál sería la secuencia más eficiente en el tratamiento médico de una determinada dolencia?, ¿hay alguna relación entre los factores de riesgo disponibles para evaluar la viabilidad a la hora de efectuar un seguro? ¿se podrían clasificar los mensajes de correo electrónico más susceptibles de ser spam de forma automática?

2.5.2.1. Minería de datos aplicada a la Calidad del Aire.

La calidad del aire se mide mediante redes de control, públicas y/o privadas, generalmente compuestas por varias estaciones de control, en las cuales se miden diversos contaminantes y parámetros meteorológicos de manera continua. Estos datos se almacenan y envían a un centro de control para su tratamiento, respondiendo a periodos de integración que pueden ser variables. Una red de control como la de la Comunidad de Madrid, con 23 estaciones de control, 150 parámetros contaminantes y 161 parámetros meteorológicos totales, genera al año un total de 2.724.000 datos horarios aproximadamente, algo difícil de manejar con una simple hoja de cálculo.

Se puede intuir que la Calidad del Aire es uno de los campos en los que los repositorios de datos son de difícil tratamiento por otros medios que no sean la minería de datos, y a su vez uno de los campos donde más rentable e interesante podría resultar aplicar dichas técnicas de análisis. Debido al alto peligro que puede causar la mala calidad del aire en la población de la localidad de Oviedo, es importante buscar relaciones entre las concentraciones de los contaminantes junto con las variables atmosféricas (precipitaciones, velocidad del viento, radiación solar y temperatura), que puedan presentar algún tipo de influencia sobre los efectos de los contaminantes, mediante técnicas de minería de datos.

Algunos de los contaminantes del aire, por sus efectos en la salud de la población, han sido normados y se han establecido límites máximos de concentración en el aire ambiente. Estos contaminantes son conocidos como criterio. Los contaminantes criterio incluyen el ozono (O_3), el monóxido de carbono (CO), el bióxido de azufre (SO_2), el bióxido de nitrógeno (NO_2), el plomo (Pb), las partículas suspendidas totales (PST), y las partículas suspendidas menores a 10 y a 2.5 micrómetros (PM_{10} y $PM_{2.5}$). Gracias al análisis y la preparación de datos es posible extraer un conocimiento excelente, ya que al identificar y observar los requerimientos de los datos el proceso de minería de datos es efectivo y se crean modelos cumpliendo las hipótesis planteadas.

Capítulo III

CONTAMINANTES DEL AIRE

Fuentes y tipos de
contaminación
atmosférica

3. Fuentes y tipos de contaminación atmosférica

3.1. Medioambiente y atmósfera

En la Conferencia de las Naciones Unidas sobre Medio Ambiente en Estocolmo (1972), se definió el concepto de medioambiente como: “El conjunto de componentes físicos, químicos, biológicos y sociales capaces de causar efectos directos o indirectos, en un plazo corto o largo, sobre los seres vivos y las actividades humanas”, citado en el libro “Agenda 21” de Foy (1998).

Tradicionalmente se ha considerado al agua, al aire, a la tierra y a la vida, es decir, a la Hidrosfera, a la Atmósfera, la Geósfera y la Biósfera, donde se tenía a la actividad humana como una actividad de perturbación permanente en las distintas esferas. Actualmente se ha incluido una quinta esfera denominada Antropósfera (Manahan, 2007), la que involucra a todos los seres humanos y las actividades relacionadas con su desarrollo y supervivencia (Figura 3.1).

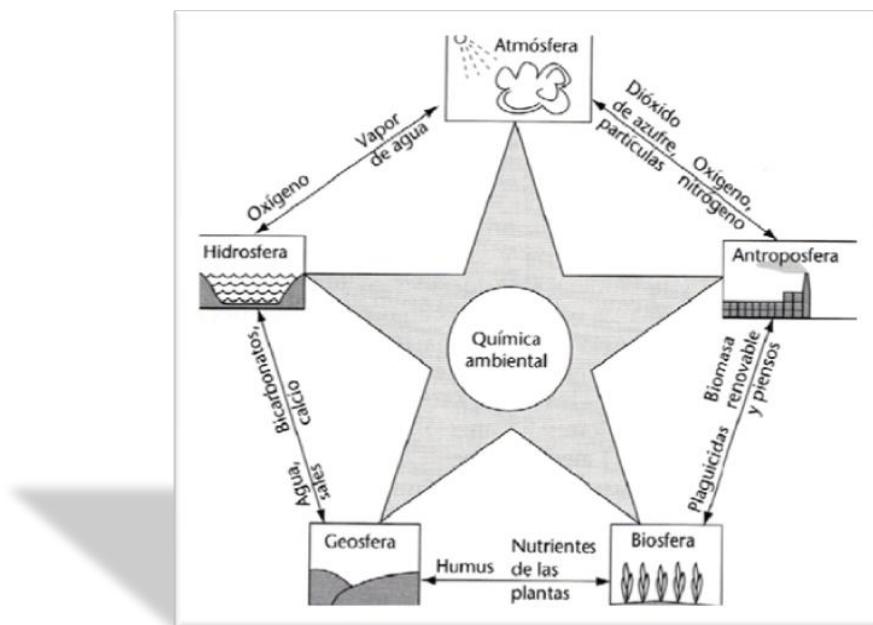


Figura 3.1. Las cinco esferas del ambiente

Fuente: Introducción a la química ambiental. Stanley E. Manahan. Editorial Reverté

En este contexto han surgido ramas de la ciencia dedicadas a estudiar en forma integral al medio ambiente: Ecología, Ecotoxicología, Fisicoquímica Ambiental, Biología Aplicada, las cuales en conjunto con la Química Ambiental y las Ciencias Sociales, evalúan permanentemente el estado poblacional, el nivel de contaminación y las metodologías de diagnóstico y remediación ambiental, tal como lo indica Manahan (2007). A todas estas disciplinas se une la ingeniería ambiental que es la rama de la ingeniería donde se estudian los problemas ambientales de forma científica e integrada, teniendo en cuenta su dimensión científica, química, ecológica, social, económica y tecnológica, con el objetivo de promover un desarrollo sostenible. Como cualquier ingeniero, el ingeniero ambiental tiene por función resolver problemas concretos recurriendo a la tecnología. Por este motivo, su mercado de trabajo es bastante heterogéneo y se distribuye entre la administración central, la administración local, empresas industriales, empresas de consultoría, empresas de servicios, organizaciones no gubernamentales, instituciones de investigación y de enseñanza superior [15].

3.1.1. Estructura y contaminación de la atmósfera

La atmósfera según su estratificación se clasifica en cuatro capas: *Troposfera, Estratosfera, Mesosfera y Termosfera*.

La troposfera es la más cercana a la superficie de la tierra y se extiende hasta los 10 a 12km, en ella se producen fenómenos meteorológicos diversos y se desarrolla la vida de todos lo que la habitan. La Estratosfera se extiende hasta los 50 km y se subdivide a su vez en dos capas, siendo la primera isoterma y aumentando la temperatura en función de la altura en la segunda. Dicho aumento de temperatura es debido a las reacciones exotérmicas del oxígeno (O_2) y del ozono (O_3), que reaccionan absorbiendo las radiaciones ultravioleta dañinas para los seres vivos. La Mesosfera se extiende desde 50 a 80 km, disminuyendo en ella la temperatura en función de la altitud. Se compone de formas ionizadas de oxígeno atómico y una baja concentración de ozono. Por último, la Termosfera se extiende desde la Mesopausa hasta el exterior.

Las primeras capas se componen de nitrógeno (N_2) y de oxígeno molecular (O_2), pero sobre los 200km predomina el oxígeno atómico (O_1 , un radical libre). En ella, la temperatura se incrementa con la altura. Se entiende por *contaminación atmosférica* la presencia de materias, sustancias o formas de energía que impliquen molestia grave, riesgo o daño para personas, medio ambiente y demás bienes.

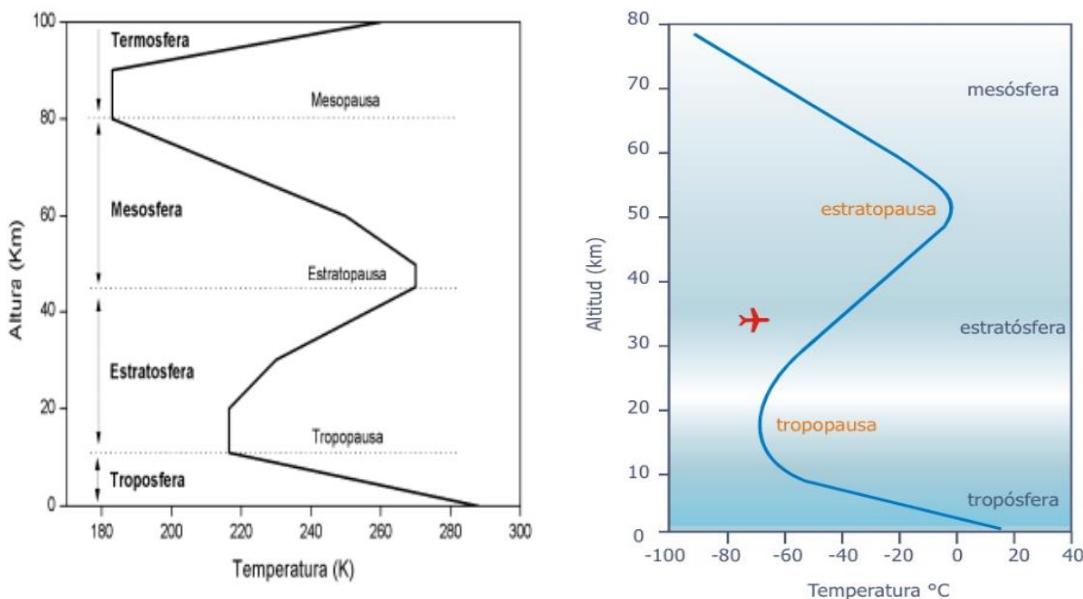


Figura 3.2. Estructura vertical de la atmósfera y su equivalente adaptada a temperatura en °C.

Fuente: Bruno Sportisse 2010

3.2. Principales contaminantes del aire

El aire puro es muy importante ya que proporciona oxígeno y otros gases que son esenciales para la vida en la Tierra. Es habitual encontrar diversos informes que muestren el incumplimiento de las respectivas normas de la calidad del aire, en las grandes urbes o espacios urbanos con alta densidad de población donde predominan el comercio, la industria y los servicios. Definimos el aire puro, como una mezcla gaseosa de nitrógeno (78%), oxígeno (21%) y otras cantidades menores de dióxido de carbono, argón, ozono y otros gases (1 %). Así, se dice que el aire está contaminado cuando existe un desequilibrio entre estos componentes y se define como una condición

atmosférica en la que ciertas sustancias presentes, podrían producir un daño medible en la salud de las personas [16].

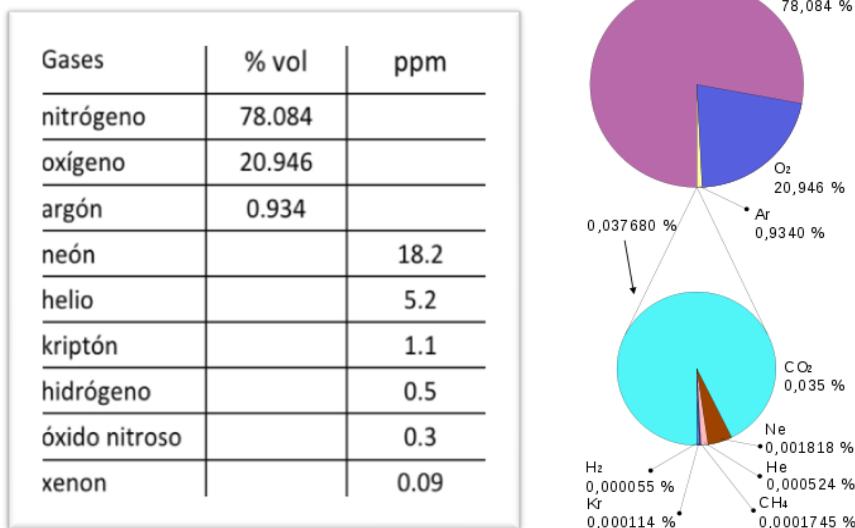


Figura 3.3. Composición de la atmósfera terrestre (tomada en diciembre de 1987).

Fuente: w:Image:Atmosphere gas proportions.gif (originally by Brockert).

El aire limpio también es vital para mantener el delicado equilibrio de la vida en la Tierra. No obstante, la calidad del aire puede resultar afectada por la contaminación.

Cuando ciertos gases y partículas se acumulan en la atmósfera a niveles tales que pueden causar daño a nuestra salud, decimos que el aire está contaminado, causando problemas respiratorios, e incluso pudiendo provocar muertes prematuras, además de dañar el ambiente que nos rodea. Estos gases y partículas, denominados contaminantes, provienen de fuentes artificiales, como la quema de combustibles fósiles de carbón, el petróleo, la gasolina o el diesel, pero también pueden provenir de fuentes naturales como las erupciones volcánicas e incendios forestales. Dichos contaminantes pueden estar constituidos como partículas sólidas, gotas de líquido o gases:

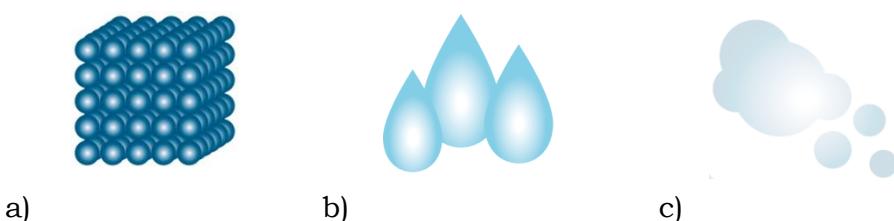


Figura 3.4. Distintas manifestaciones de los contaminantes: a) Partículas sólidas, b) gotas de líquido, c) gases. Fuente: Elaboración a partir de gráficos de internet.

Se puede establecer una clasificación de los contaminantes presentes en la atmósfera, pudiendo ser dichos contaminantes: primarios o secundarios. Los contaminantes primarios, son aquellos que permanecen en la misma forma que fueron emitidos, como el material particulado, el monóxido de carbono (CO), el dióxido de carbono (CO₂), los óxidos de nitrógeno (NO_x), los óxidos de azufre (SO_x) y los hidrocarburos. Los contaminantes secundarios son aquellos en los que se han producido cambios químicos o son producto de las reacciones químicas de dos o más contaminantes primarios, entre ellos podríamos citar el ácido sulfúrico (H₂SO₄), el ácido nítrico (HNO₃), el ozono (O₃), el smog fotoquímico y los compuestos orgánicos volátiles (COVs).

Atendiendo a las fuentes que pueden producir la contaminación atmosférica, se clasifican en fijas y móviles. Dentro del grupo de las fuentes contaminantes fijas, tenemos las emisiones procedentes de los procesos industriales, las canteras, las empresas mineras, las centrales térmicas, y otros procesos de fabricación y producción similares. En cuanto a las fuentes contaminantes móviles, se engloban todas aquellas procedentes de vehículos de todo tipo, los aviones, buques y similares. En base a lo anterior, se puede inferir que la contaminación del aire se volverá mucho más crítica en las grandes urbes, donde el crecimiento acelerado de las industrias, del parque automovilístico, las construcciones y la gestión de los residuos, entre otros, aumentan potencialmente estos prejuicios [17]. De este modo la Organización Mundial de la Salud (OMS), ha alertado sobre los peligros a los que están expuesta una gran parte de la población mundial como son las personas que residen en Ciudad de México, Los Ángeles, Santiago de Chile, entre otras, que están expuestas a niveles de contaminación atmosférica que excede los límites planteados por dicha organización (OMS 2002, Korc 2002). Este mismo organismo a alertado que existen evidencias fundadas de las consecuencias negativas debidas a la exposición prolongada a contaminantes atmosféricos, cuando éstos superan ciertos límites de concentración y de tiempo [18].

Por otra parte, también se constata que las condiciones geográficas y climáticas afectan a la dispersión de los gases y de las partículas generadas en los procesos industriales, las viviendas y transporte, afectando al estado de la atmósfera de las ciudades. Así, la temperatura ambiental, la humedad, las lluvias, la dirección y velocidad del viento o las altas y bajas presiones; todo ello, teniendo en cuenta la orografía del terreno objeto de estudio, condicionan la concentración final de los contaminantes.

3.2.1. Contaminantes primarios y secundarios

Los *contaminantes primarios*, son todos los que proceden directamente de las fuentes de emisión, por ejemplo: Plomo (Pb), Monóxido de Carbono (CO), Óxidos de azufre (SO_x), Óxidos de Nitrógeno (NO_x), Hidrocarburos (HC), material particulado, entre otros. Se emiten directamente a la atmósfera [19] como el dióxido de azufre SO₂, que daña directamente la vegetación y es irritante para los pulmones.

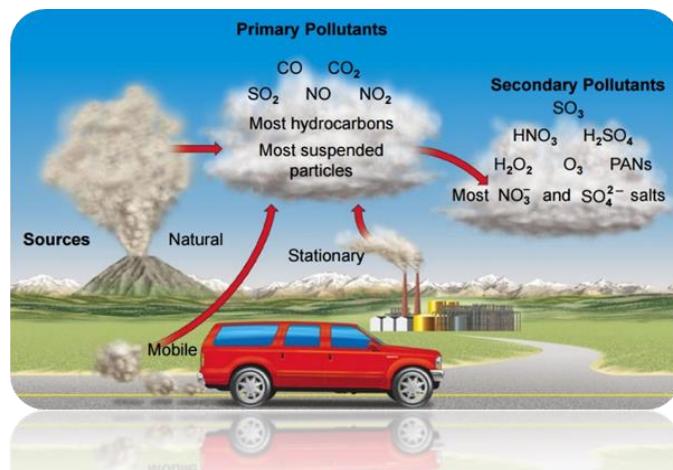


Figura 3.5. Orígenes y tipos de contaminantes primarios y secundarios. Fuente: Brooks/Cole Thomson, 2005.

Los *contaminantes secundarios*, se originan en el aire por la interacción entre dos o más contaminantes primarios, o por sus reacciones con los componentes naturales de la atmósfera. Este tipo de contaminantes se crean mediante procesos químicos atmosféricos que actúan sobre los contaminantes primarios, o, sobre especies no contaminantes en la atmósfera. Por ejemplo: el Ozono troposférico (O₃), que se forma a partir del oxígeno O₂ mediante reacción fotoquímica en presencia de NO_x y compuestos orgánicos volátiles; los Hidrocarburos (HC); los Sulfatos (SO₄);

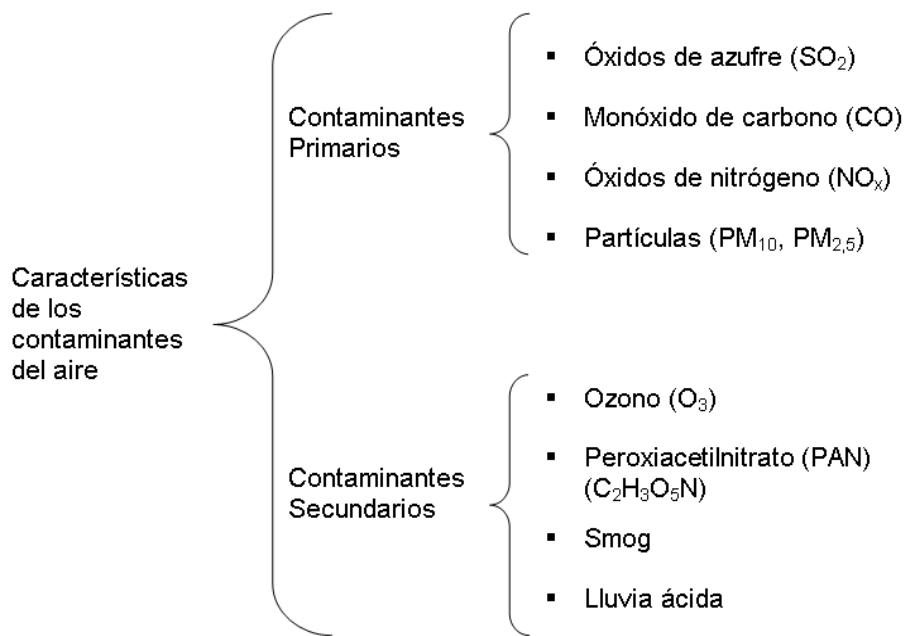
los Nitratos (NO_3) que son unos de los más frecuentes contaminantes de aguas subterráneas en áreas rurales; el Dióxido de Nitrógeno NO_2 , que se forma al oxidarse el contaminante primario NO; el Ácido Sulfúrico (H_2SO_4), formado a partir de la oxidación del SO_2 ; el material particulado (PM); El Peroxiacetilnitrato (PAN), que combinado con el ozono O_3 , óxidos de nitrógeno, los compuestos orgánicos volátiles (COV) y la acción de la radiación solar, provoca el fenómeno del ‘smog fotoquímico’ o niebla tóxica, que es el tipo más comúnmente reconocido de contaminación del aire. La niebla tóxica generalmente se refiere a una condición producida por la acción de la luz solar sobre los gases de escape de automotores, fábricas, edificios, casas, etc.

<u>TIPOS DE CONTAMINANTES.</u>	
<p><u>CONTAMINANTES PRIMARIOS:</u> Son vertidos directamente a la atmósfera por alguna fuente de emisión como chimeneas, automóviles, entre otros.</p>	<p><u>CONTAMINANTES SECUNDARIOS:</u> Son originados en el aire por la transformación y reacciones químicas que sufren los contaminantes primarios.</p>
<p><u>Óxidos de azufre (SO_x):</u> Se forman por la combustión del azufre presente en el carbón y el petróleo. Los SO_x forman con la humedad ambiente aerosoles, incrementando el poder corrosivo de la atmósfera, disminuyendo la visibilidad y provocando la lluvia ácida.</p> <p><u>Monóxido de carbono (CO):</u> Es el contaminante más abundante en la capa inferior de la atmósfera. Se produce por la combustión incompleta de compuestos de carbono. Es un gas inestable que se oxida generando dióxido de carbono (CO₂). Alrededor del 70 por ciento del CO provienen de los vehículos.</p>	<p><u>Ozono (O₃):</u> Como se mencionó anteriormente el O₃ forma parte de la composición de la atmósfera, sin embargo a baja altura (O₃ troposférico) resulta perjudicial por su carácter oxidante, reactivo, corrosivo y tóxico, por lo que reacciona con rapidez generando compuestos secundarios.</p> <p><u>Lluvia ácida.</u> Es el término que se ha usado para describir el proceso por el cual ciertos ácidos se forman en la atmósfera a partir de contaminantes y luego se precipitan a la tierra. El SO₂ (dióxido de azufre) y los NOX, causan la lluvia ácida.</p>

<p>Óxidos de nitrógeno (NOx). Se producen en la combustión de productos fósiles, destacando los vehículos, carbón y quemadas de madera. La producción de fertilizantes y explosivos, tabaco y calderas generan emisiones importantes de NOx. El monóxido de nitrógeno (NO) y el dióxido de nitrógeno (NO₂) requieren especial atención. El NO se oxida formando NO₂, mientras que el NO₂ es precursor del esmog fotoquímico.</p> <p>Partículas (PM). Es material respirable presente en la atmósfera en forma sólida o líquida (polvo, cenizas, hollín, partículas metálicas, cemento y polen, entre otras). De acuerdo con su tamaño se pueden dividir en dos grupos principales: las de diámetro aerodinámico igual o inferior a los 10 µm (PM₁₀) y las de fracción respirable más pequeña (PM_{2,5}).</p> <p>Hidrocarburos (HC). Las emisiones de HC están asociadas a la mala combustión de derivados del petróleo. Las fuentes más importantes de emisión son el transporte por carretera, los disolventes, pinturas, vertederos y la producción de energía. Los de mayor interés son los compuestos orgánicos volátiles (COV), dioxinas, furanos, bifenilos policlorados (PCB) y los hidrocarburos policíclicos aromáticos (PAH).</p>	<p>Estas sustancias en presencia de agua, O₂ y otros compuestos químicos forman ácido sulfúrico (H₂SO₄) y ácido nítrico (HNO₃) respectivamente, que se precipitan a la tierra en forma líquida cuando se presentan lluvias o en forma seca en presencia de nevadas o neblinas. El pH de la lluvia normal es de alrededor de 6, mientras que la lluvia ácida presenta un pH menor a 5.</p> <p>Contaminación fotoquímica. La constituyen la luz solar y sustancias susceptibles de ser oxidadas. El esmog fotoquímico es una mezcla de contaminantes que se forman por reacciones producidas por la luz solar al incidir sobre los contaminantes primarios.</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 3.1. Contaminantes primarios y secundarios de la atmósfera. Elaboración propia, a partir del informe de la Secretaría de Medioambiente y Recursos Naturales – SEMARNAT (Méjico).

3.3. Características principales, parámetros y efectos más relevantes de los contaminantes primarios y secundarios



En España en la actualidad hay en funcionamiento diversas redes de vigilancia de la contaminación atmosférica, instaladas en las diferentes Comunidades Autónomas y que efectúan medidas de una variada gama de contaminantes que abarcan desde los óxidos de azufre y nitrógeno hasta hidrocarburos, con sistemas de captación de partículas, monóxido de carbono, ozono, metales pesados, etc. [20].

Los contaminantes gaseosos más comunes son el dióxido de carbono, el monóxido de carbono, los hidrocarburos, los óxidos de nitrógeno, los óxidos de azufre y el ozono. Estos compuestos químicos son producidos por distintas fuentes, pero la principal fuente artificial es la quema de combustible fósil. Los contaminantes gaseosos del aire provienen de volcanes, e industrias. El tipo más comúnmente reconocido de contaminación del aire es la niebla tóxica (smog). La niebla tóxica, como más adelante se detalla, se refiere a una serie de reacciones producidas por la acción de la luz solar sobre los gases de escape de automotores y fábricas. Seguidamente vamos a describir las características, los parámetros y algunos de los efectos de los contaminantes habituales del aire.

3.3.1. Material particulado (PM)

A partir de la década de 1990, se han venido realizado numerosos estudios epidemiológicos (Dockery et al., 1993; Dockery y Pope, 1996; Hoek et al., 2002; Pope et al., 2002), que demuestran los efectos adversos para la salud derivados de la exposición, puntual o prolongada, a los niveles elevados de material particulado atmosférico (MPA). En otros estudios más recientes (información y enlaces en la dirección, <https://www.epa.gov/environmental-topics/air-topics>), se apunta hacia las partículas de menor diámetro (partículas finas, con diámetro < 2.5 μm – PM_{2,5}) como causantes de la mayoría de las afecciones respiratorias sufridas por los habitantes de las ciudades industrializadas. El 1 de diciembre de 2016, la Agencia de Protección Ambiental de Estados Unidos USEPA, propuso determinaciones para 11 áreas de incumplimiento, clasificados como moderados, para las Normas Nacionales Calidad del Aire Ambiental de 2006, 24 horas partículas finas. Sobre la base de los requisitos en virtud de la Ley de Aire Limpio, se exigió a estas áreas para cumplir la norma del 31 de diciembre de 2015 (<https://www.gpo.gov/fdsys/pkg/FR-2016-12-16/pdf/2016-30174.pdf>, con acceso en diciembre de 2017).

El término *aerosol* o *partícula* se utiliza indistintamente, considerándose los aerosoles atmosféricos como «*dispersiones de sustancias sólidas o líquidas en el aire*». Las propiedades de los aerosoles que más afectan a la contaminación atmosférica son el tamaño de partículas, la forma y la composición química. Las partículas del aerosol ambiental pueden ser de origen primario o secundario. Los aerosoles primarios son formados por la emisión directa de las partículas desde su fuente y los secundarios se forman a partir de reacciones atmosféricas a partir de precursores gaseosos [21]. Los aerosoles orgánicos secundarios son un producto de la oxidación atmosférica de algunos de los gases orgánicos reactivos presentes en ella por uno de los tres gases electrofílicos presentes en trazas en la atmósfera: el radical hidroxilo (OH[·]), el ozono (O₃), y el radical nitrato (NO₃[·]) [22].

Los gases orgánicos reactivos son emitidos durante el quemado de madera y combustibles fósiles, uso de solventes, emisiones de los vegetales y por los océanos [Duce et al., 1983], estos gases incluyen a compuestos como alquenos, hidrocarburos aromáticos y fenoles (Jacobson et al., 2000). Tanto los primarios como los secundarios pueden formarse como consecuencia de las actividades naturales o antrópicas. Las fuentes naturales son muy importantes, especialmente las que exponen una superficie grande como pueden ser los océanos o desiertos y generalmente superan a las fuentes antrópicas a excepción de ciudades muy urbanizadas con diversos procesos de combustión dominantes, de fuentes fijas como las industriales y las calefacciones domésticas y móviles como el tránsito vehicular [23].

Las partículas no son esféricas ni todas de igual densidad, sino que responden a un alto grado de variabilidad de tamaño y de anisotropía en todas sus demás propiedades físicas; a fines práctico-comparativos se les asigna un diámetro equivalente (D_e) que corresponde al diámetro que tendría la partícula esférica con el mismo volumen que la partícula en estudio, además y más importante aún es el diámetro aerodinámico equivalente (D_a) (Bird et al., 1982). Se pueden distinguir, distintos modos de distribuciones de tamaño [24].

- Partículas ultra-finas, moda de nucleación ($<0,02 \mu\text{m}$)
- Moda Aitken ($0,02-0,1 \mu\text{m}$)
- Moda de acumulación ($0,1-1 \mu\text{m}$)
- Moda gruesa ($>1 \mu\text{m}$ en términos de aerosoles atmosféricos)

Atendiendo a distintos tipos de clasificación, a continuación, en la Tabla 3.2 se puede ver la clasificación de las partículas según su tamaño y composición, así como las distribuciones modales de tamaños de partículas en la Figura 3.6.

<u>Denominación</u>	<u>Composición</u>
Núcleos de Aitken	Partículas con menos de 1 micra de diámetro.
Partículas medias (en suspensión)	Partículas con un diámetro comprendido entre 1 y 10 micras.
Partículas sedimentables	Partículas con diámetro superior a 10 micras.
Polvos	Partículas sólidas de origen mineral o materia sólida dispersada por el aire.
Humos industriales	Partículas sólidas o líquidas debidas a la volatilización de metales, seguida o no de su oxidación por el aire o condensación de vapores.
Humos de combustión	Humos debidos a proceso de combustión, constituidos por partículas de carbono y de hidrocarburos no quemados y cenizas volantes.

Tabla 3.2. Clasificación de las partículas según su tamaño y composición.

Fuente: Adaptada de José E. Marcano <http://www.jmarcano.com/recursos/contamin/catmosf2a.html>

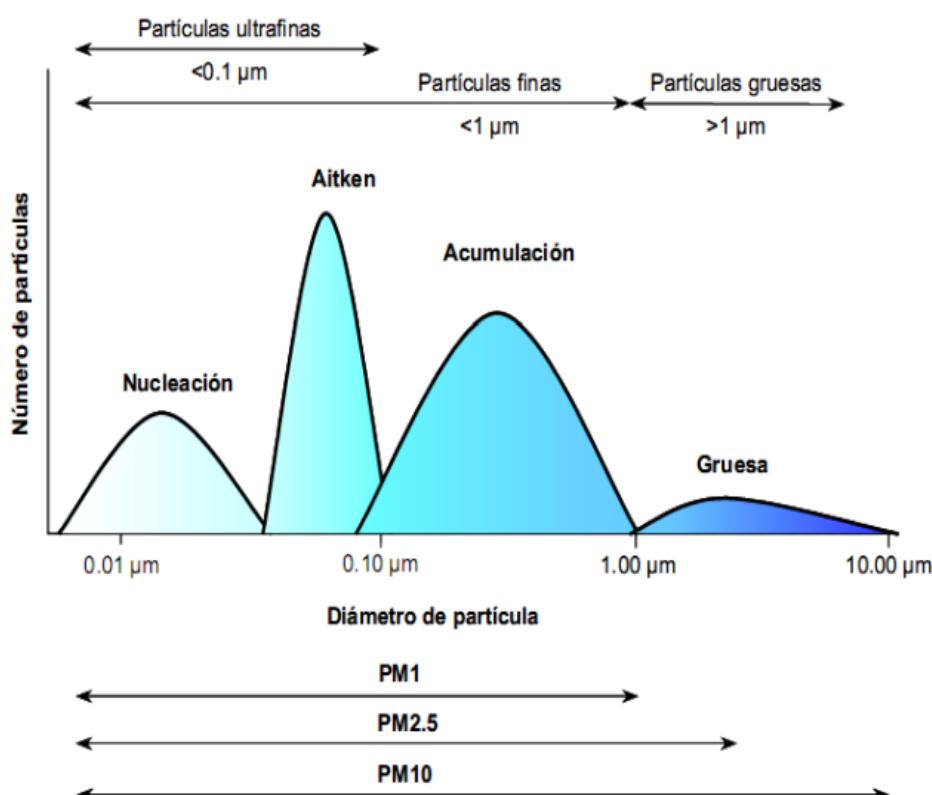


Figura 3.6. Distribuciones modales de tamaños de partículas.

Fuente: USEPA 1996, 2004, y Querol et al., 2001.

El tamaño de las partículas puede oscilar entre 1 y 1000 micras. Las partículas de tamaño inferior a 1 micra circulan por la atmósfera desplazándose de forma aleatoria pudiendo originar agregados de mayor tamaño, como consecuencia de choques que se producen entre ellas, en un proceso denominado *coagulación*. Las partículas de tamaño comprendido entre 1 y 10 micras forman suspensiones mecánicamente estables, y se denominan «*materia en suspensión*», recorriendo grandes distancias por la acción del viento. Las partículas mayores de 10 micras, permanecen en suspensión durante periodos de tiempo relativamente cortos, por lo que se las conoce como «*materia sedimentable*»; sus efectos, son más acusados en las proximidades de las fuentes que las emiten. El tamaño de las partículas es un factor muy importante en la determinación tanto de los efectos que producen como de las áreas afectadas; ya que, establece su tiempo de permanencia en la atmósfera y la facilidad con que se introducen en las vías respiratorias profundas.

El material particulado (PM), incluye las partículas sólidas y líquidas del aire, de los contaminantes primarios como el polvo y el hollín, y las partículas líquidas producidas por la condensación de los vapores de los secundarios. Las normas sobre material particulado, incluyen todas aquellas partículas de diámetro menor a $10\mu\text{m}$ (PM_{10}). En la segunda mitad de la década de los 90's, se incluyeron en la normativa las partículas menores de $2,5\mu\text{m}$ ($\text{PM}_{2,5}$) [25]. Las partículas con diámetros entre $2,5$ y $40\mu\text{m}$, se forman por la trituración mecánica y la abrasión de superficies. El movimiento de dichas partículas viene determinado por el viento y el tráfico, permaneciendo en suspensión y dispersándose. Las partículas menores de $2,5\mu\text{m}$ ($\text{PM}_{2,5}$), son consideradas finas, y, penetran en los pulmones profundamente. La procedencia de estas partículas tiene su origen en las fuentes de combustión existentes y en la condensación de los materiales volatilizados, denominándose dichas partículas material particulado primario; así como, en los gases que reaccionan en el aire formando las denominadas partículas secundarias [26]. En la Figura 3.7, se puede apreciar la comparación entre los distintos tamaños de material particulado.

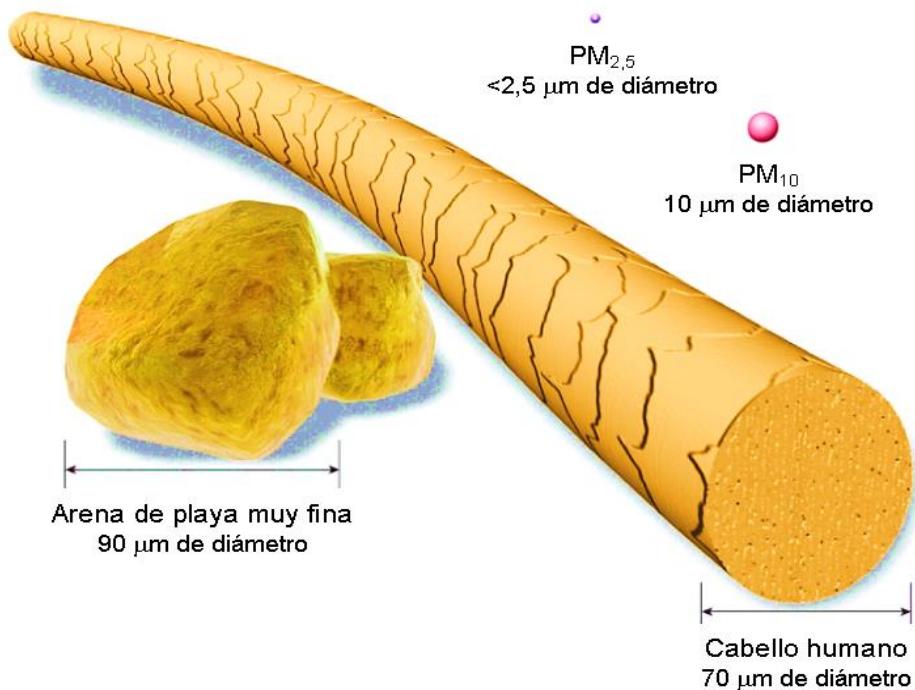


Figura 3.7. Comparación entre los distintos tamaños de material particulado.

Fuente: Pearson Prentice Hall, inc. 2010.

Las partículas (PM₁₀), y menores, pueden ser inhaladas penetrando en el sistema respiratorio, causando así efectos nocivos a la salud de las personas; mientras que las de diámetros mayores, quedan atrapadas en las fosas nasales. El material particulado que una persona puede inhalar depende de la respiración de esa persona; es decir, a mayor frecuencia y profundidad de la respiración mayor cantidad de partículas inhaladas. Las partículas de mayor diámetro se depositan en las fosas nasales y laringe; mientras que las de diámetros entre 5 y 10 μm lo hacen en los bronquios, llegando las más pequeñas hasta los bronquiolos y alveolos. Las partículas que se depositan en los pulmones pueden ser eliminadas mediante la actividad mucociliar y por los macrófagos, en un periodo de tiempo que va de semanas hasta años [27]. Las partículas muy pequeñas, emitidas en cantidades de trillones por vehículos diesel, aviones, plantas generadoras y otras fuentes, podrían viajar miles de kilómetros permaneciendo en suspensión durante semanas o meses.

La composición química de las partículas indica su fuente de procedencia, en las partículas finas ($PM_{2,5}$) hay sulfatos, hidrocarburos aromáticos policíclicos, plomo arsénico, berilio, cadmio, amonio, mercurio, sulfatos y nitratos. Las partículas de diámetro entre (PM_{10}) y ($PM_{2,5}$), están compuestas por materiales cristalinos como sílice, compuestos de hierro y aluminio. La masa de las partículas ultrafinas puede estar constituida o portar metales pesados, hidrocarburos aromáticos policíclicos u otros elementos nocivos para la salud, que pudieran causar daños y lesiones a largo plazo [28]. En cuanto al impacto sobre la salud que presenta la exposición al material particulado PM susceptible de ser respirado se deberán tener en cuenta, entre otros factores, el tamaño de los aerosoles; ya que, la capacidad de penetración de los mismos en las vías respiratorias, aumenta al disminuir su tamaño.

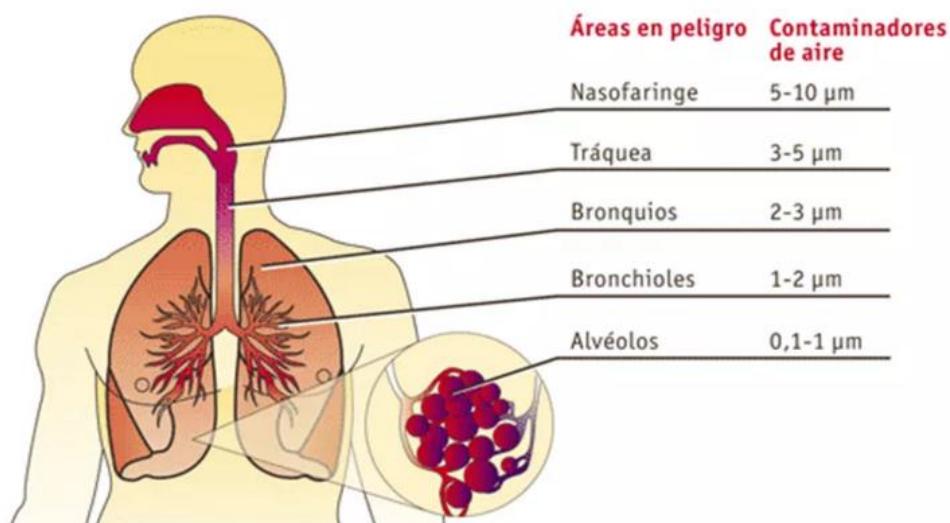


Figura 3.8. Capacidad de penetración de las partículas en el organismo, en función de su tamaño.

Fuente: BIOSAIT EUROPE SL.

Como se aprecia en la Figura 3.8, las partículas PM_{10} son retenidas en la región extra-torácica, mientras que la fracción de aerosoles con tamaño entre 2.5 y 10 μm ($PM_{10-PM_{2,5}}$), pueden alcanzar la zona traqueo-bronquial. Para materiales particulados cuyas fracciones de tamaño son menores a 2.5 μm , se tendrá en cuenta que pueden filtrarse a través de los alvéolos pulmonares y llegar al torrente sanguíneo.

Las partículas, de 2.5 micras y menores, denominadas partículas finas por su muy pequeño tamaño, penetran profundamente en los alvéolos pulmonares y causan daños permanentes, así como una reducción de la capacidad respiratoria. Las citadas partículas, constituyen lo que se denomina carbón negro (BC o black carbón) y puede absorber un millón de veces más energía que el CO₂. El BC es producido por la combustión de combustibles fósiles, en especial el Diesel, que es el mayor responsable de su emisión, biocombustibles y biomasa y por lo tanto el mayor responsable del calentamiento global.



Figura 3.9. PM_{2.5}, o carbón negro, o black carbon.

Fuente: Dr. Gonzalo E. Díaz Murillo - Experto en Telemedicina y Ultrasonido

Los niveles normativos de calidad del aire referentes al PM se pueden consultar en la legislación europea en la Directiva 1999/30/CE y en la Directiva 2008/50/CE. La transposición de las mismas queda recogida en el Real Decreto 102/2011 de 28 de enero, relativo a la mejora de la calidad del aire, modificado recientemente por el Real Decreto 39/2017, de 27 de enero, en la dirección: <https://www.boe.es/boe/dias/2017/01/28/pdfs/BOE-A-2017-914.pdf>, se puede consultar la información al respecto.

A la vista de los datos anteriores, se puede entender que las partículas más finas multipliquen tienen efectos potencialmente más nocivos sobre la salud que las partículas con diámetros superiores a 10µm. Los estudios de epidemiología demuestran la correlación positiva entre la concentración de la fracción más fina de las partículas y los casos de mortalidad en las ciudades.

En los inventarios europeos, los datos relativos a las emisiones por combustión en el sector residencial se recogen en el denominado «comercial, institucional y doméstico». En la UE-28, las emisiones por combustión en este sector es la principal fuente de PM_{2,5} primaria y PM₁₀, así como las emisiones de BC (Black carbon) y BaP (Benzo[a]pireno - hidrocarburo aromático policíclico), que aportan el 56% del total de las emisiones.

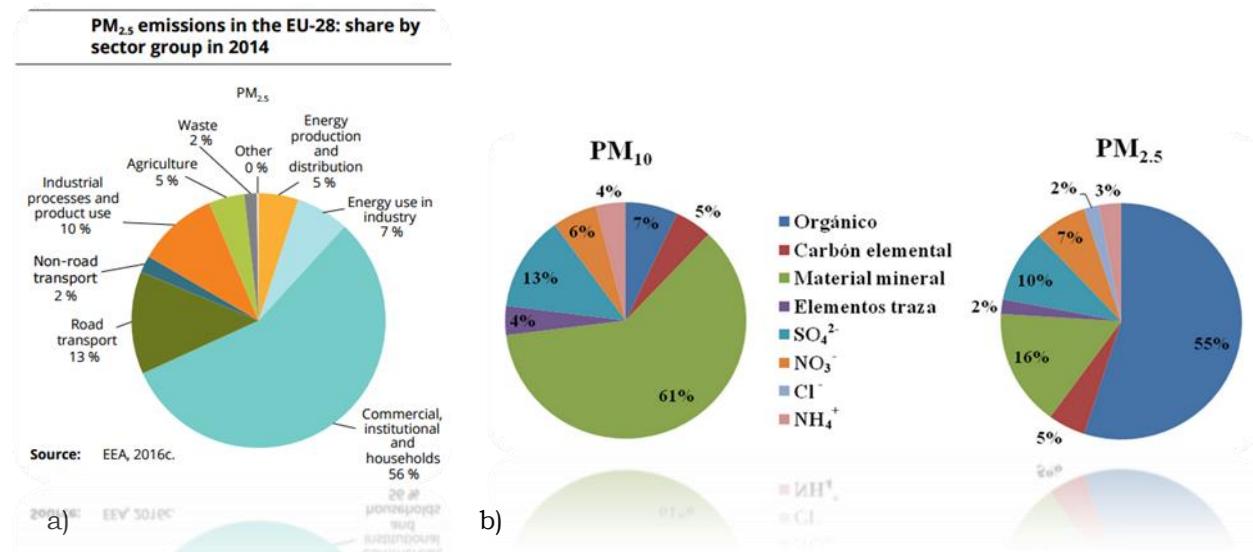


Figura 3.10. a) Emisiones PM_{2,5} en la UE-28: participación por sector en 2014. Fuente EEA 2016c, b) Caracterización química de PM₁₀ y PM_{2,5} en el Instituto Mexicano del Petróleo, Ciudad de México, durante la campaña MILAGRO, 2006. Basado en Querol *et al.* (2008).

3.3.2. Monóxido de carbono (CO)

El monóxido de carbono, también denominado óxido de carbono (II), gas carbonoso y anhídrido carbonoso (los dos últimos cada vez más en desuso), cuya fórmula química es CO, fue descubierto por el químico francés De Lassone en 1776 mientras calentaba óxido de zinc con coque. Erróneamente creyó que se trataba de hidrógeno porque generaba una llama de color azul. Más tarde, en 1800, el químico inglés William Cruikshank comprobó que dicho compuesto contenía carbono y oxígeno.

Las propiedades tóxicas del CO fueron investigadas en profundidad por el físico francés Claude Bernard en 1846. Envenenando perros con el gas detectó que su sangre se tornaba más rojiza y brillante en todos los tejidos.

Durante la Segunda Guerra Mundial, el monóxido de carbono fue usado en los motores de los vehículos ya que escaseaba la gasolina. Se introducía carbón mineral o vegetal y el monóxido de carbono generado por gasificación alimentaba al carburador. El CO también fue usado como un método de exterminio (cámaras de gas) durante el Holocausto en los campos de concentración.

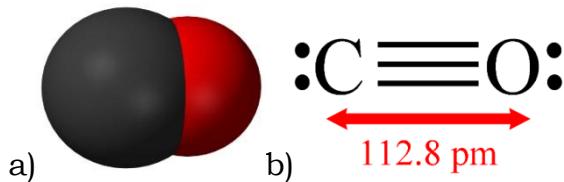


Figura 3.11.a) Representación en 3D de una molécula de monóxido de carbono. b) Estructura de la molécula de monóxido de carbono. Fuente:
https://es.wikipedia.org/wiki/Mon%C3%B3xido_de_carbono

La exposición a monóxido de carbono (CO) es uno de los principales riesgos laborales en ciudades industrializadas, siendo un problema de salud pública, que es necesario que los gobiernos atiendan con celeridad (OMS, 2002). También los países en vías de desarrollo sufren este problema; ya que la polución urbana debida principalmente al aumento del parque automovilístico, expone a la población de dichas ciudades a la inhalación de gases tóxicos, entre los que se incluye el monóxido de carbono (CO). (Téllez et al., 2006).

El CO, es el contaminante del aire más abundante en la capa inferior de la atmósfera, principalmente en el entorno de las grandes ciudades, siendo extremadamente peligroso y muy tóxico. Se origina por la combustión incompleta de hidrocarburos, componentes de la gasolina, maderas, carbón de hulla y otros. Este gas es invisible e inodoro y puede causar la muerte si es respirado en elevados niveles. Su punto de ebullición es de -192°C y presenta una densidad del 96.5 % de la del aire, siendo un gas muy ligero que no es apreciablemente soluble en agua. Es inflamable y arde con llama azul, aunque no mantiene la combustión.

El monóxido de carbono, produce en todas las combustiones incompletas, resultando especialmente peligroso los calentadores a gas en habitaciones y cuartos de baño mal ventilados, los braseros de leña, los gases de tubo de escape de los automóviles y los incendios. Los principales problemas de contaminación atmosférica por CO son debidos a la combustión incompleta de carburantes en los automóviles; así por ejemplo, la combustión de la gasolina casi nunca es completa, formando una cantidad de monóxido de carbono (CO), mayor aún en las gasolinas que no son de buena calidad [29]. También su origen puede ser debido a otras fuentes de generación de calor y energía, procesos industriales de diversas actividades productivas e incineración de desechos. El CO se produce generalmente como resultado de alguno de los siguientes procesos químicos:

- Combustión incompleta del carbono.
- Reacción a elevada temperatura entre el CO₂ y materiales que tienen carbono.
- Disociación del CO₂ a altas temperaturas.
- Oxidación atmosférica del metano (CH₄ procedente de la fermentación anaerobia (sin aire) de la materia orgánica.
- Proceso de producción y degradación de la clorofila en las plantas.

3.3.2.1. Efectos sobre la salud de las personas

La absorción del CO se origina por vía respiratoria y es transportado vía sanguínea, combinándose con la hemoglobina (Hb) dentro de los glóbulos rojos, formando un compuesto reversible denominado carboxihemoglobina; convirtiéndose en una sustancia tóxica para el organismo, por su dificultad para transportar oxígeno. La intoxicación por CO produce hipoxia en los tejidos, como resultado de la disminución en la afinidad por el transporte del oxígeno [Rojas et al., 2001].

El CO es absorbido hacia la sangre por medio de la vía alveolar sin modificar la función respiratoria; de los alvéolos pulmonares pasa a los capilares tisulares llegando a la sangre, y de esta manera se une a la hemoglobina (Hb) desplazando el O₂ de la oxihemoglobina (O₂Hb) en forma de COHb, por lo que la hemoglobina(Hb) transporta solamente una pequeña cantidad de O₂, que es difícilmente cedida a los tejidos para su utilización, esto conlleva a la Hipoxia [Alvarado & Hernández, 2008].

El monóxido de carbono (CO), se acopla a la hemoglobina con una efectividad cientos de veces superior a la del oxígeno, realizándose dicha unión de manera irreversible, formando como producto resultante de esta unión una proteína denominada carboxihemoglobina (COHb).

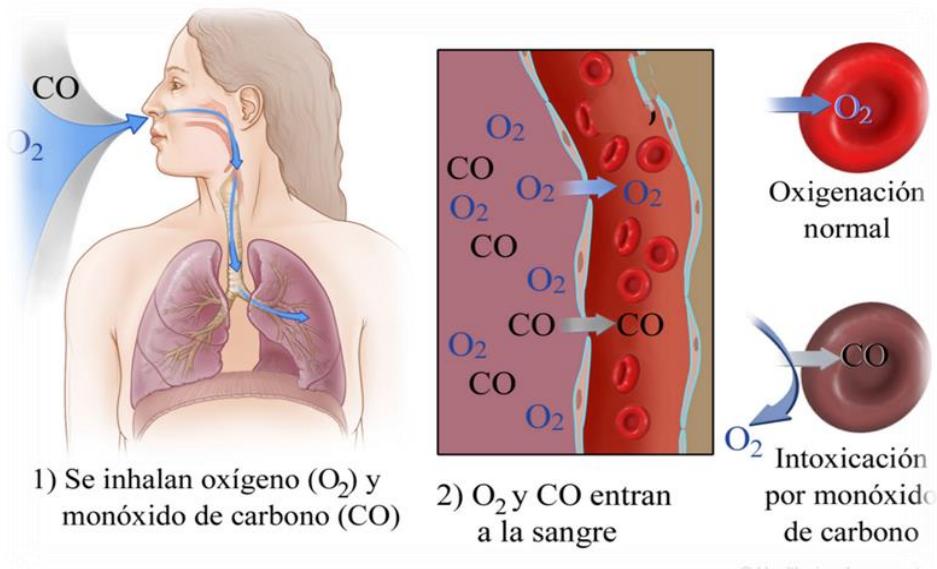


Figura 3.12. Unión de CO a los hematíes. Fuente: Álvarez, 2010

A medida que se inhala monóxido de carbono (CO), éste se va adhiriendo de forma progresiva a las moléculas de hemoglobina de la sangre. De esta forma, la cantidad de oxígeno que reciben los órganos y tejidos del cuerpo, va sumiendo a la persona que lo respira en un sopor que finalmente la conducirá a la muerte, cuando las necesidades de oxígeno del cerebro no sean adecuadamente cubiertas.

Algunos estudios de investigación experimentales realizados con animales, y epidemiológicos efectuados en humanos, en los últimos años, han evidenciado relación entre población expuesta en forma crónica a niveles medios y bajos de monóxido de carbono en aire respirable, y la aparición de efectos adversos en la salud humana, especialmente en órganos de alto consumo de oxígeno como cerebro y el corazón [30].

Por otra parte, la exposición continuada al monóxido de carbono (CO), es un factor importante a considerar y por ello es objeto de muchos estudios e investigaciones Jaimes et al., (1990); Ares et al., (2001); Crespo et al., (2001) y Rojas et al., (2001). El factor tiempo se ha señalado como responsable de la gravedad de los síntomas, incluso donde los niveles del gas no son muy elevados. También el parámetro tiempo se ha empleado para establecer diferencias entre los distintos tipos de intoxicación, aguda y crónica. Según estudios realizados, se pudo comprobar en un periodo de 4 horas continuas de permanencia, en atmósferas donde las concentraciones ambientales eran superiores a $20 \mu\text{g}/\text{m}^3$, fueron suficiente para que un individuo sano pudiera estar afectado de intoxicación aguda; mientras que, valores iguales o por encima $10 \mu\text{g}/\text{m}^3$ por periodos no consecutivos de 10 horas en ciclos de 30 días, podrían desarrollar en poco tiempo una intoxicación crónica (Ares et al., 2001).

Se han publicado varios estudios clínicos, que han proporcionado una imagen relativamente coherente, de los efectos del monóxido de carbono en el sistema cardiopulmonar. Sobre los efectos cardiovasculares, en individuos afectados de este tipo de enfermedades, se puede consultar (Allison et. al., 1996). Respecto a los relativos a la función pulmonar y tolerancia al ejercicio físico, se podría consultar los trabajos de (Ekblom and Huot, 1972; Klausen et al., 1983).

3.3.3. Óxidos de nitrógeno (NO, NO₂, NO_x)



Figura 3.13. a) Representación en 3D de una molécula de monóxido de nitrógeno
b) Estructura de la molécula de monóxido de nitrógeno.

Fuente:[https://es.wikipedia.org/wiki/%C3%93xido_de_nitr%C3%B3geno_\(II\)](https://es.wikipedia.org/wiki/%C3%93xido_de_nitr%C3%B3geno_(II))

Se denominan óxidos de nitrógeno todos aquellos compuestos químicos binarios gaseosos que están formados a partir de la mezcla de gases de nitrógeno y de oxígeno. Se forman a partir de la combustión de dichos compuestos inorgánicos, por combustión a altas temperaturas, siendo en este proceso el comburente el aire. Los óxidos de nitrógeno provienen del escape de vehículos motorizados (sobre todo diésel), de la combustión del carbón, petróleo o gas natural, y por procesos tales como la soldadura por arco, la galvanoplastia, el grabado de metales y la detonación de dinamita. También se producen por la reacción del ácido nítrico con metales o con celulosa; teniendo en cuenta que, esta reacción del ácido nítrico sobre los metales depende de la temperatura y de la concentración del ácido. Los óxidos de nitrógeno se emplean para producir lacas, tintes, combustibles para cohetes, en la manufactura de explosivos, como conservante para la carne, o para la producción de ácido nítrico, que a su vez es utilizado para crear abonos, colorantes, explosivos, fabricación del ácido sulfúrico, medicamentos y grabado de metales. El monóxido y el dióxido de nitrógeno son dos de los óxidos de nitrógeno más tóxicos que existen, no siendo inflamables ninguno de ellos.

3.3.3.1. Monóxido de nitrógeno

El óxido nítrico o monóxido de nitrógeno (NO), uno de los contaminantes de la atmósfera objeto de este estudio, es un gas que a temperatura ambiente presenta un olor dulce y penetrante, carece de color, y es difícilmente soluble en agua. Se oxida fácilmente convirtiéndose en dióxido de nitrógeno.

Como consecuencia de las transformaciones químicas que se pueden producir a partir del monóxido de nitrógeno (NO) tiene su origen el smog fotoquímico, que es una mezcla de humo y niebla muy habitual en las ciudades industrializadas. Dicho smog fotoquímico, como veremos más adelante, se forma como consecuencia de un proceso complejo, debido a la oxidación del NO (g) para formar NO₂ (g); siendo este último uno de los agentes más responsables de la lluvia ácida.

El NO, se origina en la atmósfera, por la reacción de N₂ y O₂. La formación de monóxido de nitrógeno gaseoso es un proceso que a temperatura ambiente no se produce fácilmente. Sin embargo, debido a que es una reacción de tipo endotérmica, la constante de equilibrio crece con la temperatura.

Propiedades físicas del óxido de nitrógeno (NO):

- Gas incoloro
- Densidad 1.03 kg/m³; 0,00103 g/cm³
- Masa molecular 30,01 g/mol
- Punto de fusión 109,5 K (-163,65 °C)
- Punto de ebullición 121 K (-152,15 °C)
- Temperatura crítica 180 K (°C)
- Solubilidad en agua 0,0056 g
- Momento dipolar 0,15 D

Como se ha comentado, el nitrógeno (N₂) y el oxígeno (O₂) pueden combinarse a altas temperaturas para formar el óxido nítrico (NO). Así, debido a las actividades desarrolladas por el hombre, se ha incrementado la presencia de este gas en la atmósfera. Este gas en el aire puede llegar a convertirse en ácido nítrico, produciéndose así la denominada lluvia ácida. Por otra parte, es sabido que tanto el óxido nítrico (NO), como el dióxido de nitrógeno (NO₂) participan en la degradación de la capa de ozono.

3.3.3.2. Dióxido de nitrógeno

En el caso del dióxido de nitrógeno (NO_2), en lo relativo a la contaminación, parece acentuarse el problema. Este gas, envenena a sus habitantes y preocupa cada vez más a las autoridades políticas y sanitarias. El dióxido de nitrógeno aparece en la combustión de los vehículos a motor y de las plantas eléctricas, y es el principal contaminante de los óxidos de nitrógeno. Es el causante de que el ayuntamiento de Madrid, durante el periodo navideño a finales de 2016 y comienzos del 2017, se haya visto obligado a tomar medidas extraordinarias, como reducir a 70 km/h la circulación de los vehículos por las circunvalaciones e interior de la M-30, o prohibir la circulación según el número par o impar de la matrícula [31].

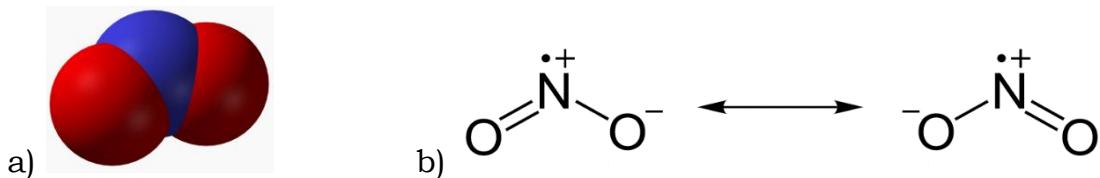
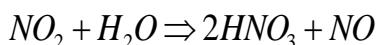


Figura 3.14. a) Representación en 3D de una molécula de dióxido de nitrógeno
b) Estructura de la molécula del dióxido de nitrógeno.

Fuente:https://es.wikipedia.org/wiki/Di%C3%B3xido_de_nitr%C3%B3geno

El dióxido de nitrógeno u óxido nítrico es un compuesto químico cuya fórmula es (NO_2), formado por un átomo de nitrógeno y dos de oxígeno, es una sustancia de color amarillento, que se forma en los procesos de combustión en los vehículos motorizados y las plantas eléctricas. A temperatura ambiente es un líquido, transformándose en torno a 20°C en un gas de color pardo rojizo. Este gas presenta un olor fuerte y muy desagradable. La reacción del dióxido de nitrógeno (NO_2) con otras sustancias químicas, originadas a partir de la luz solar, conlleva la producción de ácido nítrico (HNO_3), componente principal de la lluvia ácida, el ozono (O_3) y el smog fotoquímico [32]. Presenta por tanto buena solubilidad en agua, reaccionando y formando ácido nítrico (HNO_3) según la siguiente reacción:



Seguidamente, se dan las características principales del dióxido de nitrógeno (NO_2) [33].

- Densidad: 1,449 kg/m³.
- Masa molar: 46,1 g/mol.
- Punto de fusión: -11,20 °C.
- Punto de ebullición: 21,2 °C.
- Solubilidad en agua: Buena.
- Temperatura crítica: 158° C
- Presión de vapor: (KPa a 20° C): 96
- Densidad relativa del gas (aire = 1g/ml): 1.58

La concentración del dióxido de nitrógeno (NO_2), es muy importante puesto que interviene en diversas de las reacciones químicas, que tienen lugar en la atmósfera, dando lugar mediante las reacciones oportunas a la producción de ozono troposférico (O_3), así como a partículas en suspensión secundarias menores de 2,5 micras ($\text{PM}_{2,5}$), que son las más dañinas para la salud. Es un gas tóxico, irritante y precursor de la formación de partículas de nitrato, que conllevan la producción de ácidos, y, elevados niveles de partículas $\text{PM}_{2,5}$ en el ambiente. De modo que, a la hora de considerar los efectos del NO_2 sobre la salud, se deberían tener en cuenta no sólo los efectos directos que provoca, sino también porque el NO_2 constituye un buen indicador de la contaminación debida al tráfico rodado, y, su condición de precursor de otros contaminantes.

3.3.4. Óxidos de azufre (SO_x)

El dióxido de azufre, u óxido de azufre (IV), es un óxido cuya fórmula molecular es SO_2 . Es un gas incoloro con un característico olor asfixiante. Se trata de una sustancia reductora que, con el tiempo, el contacto con el aire y la humedad se convierte en trióxido de azufre. La velocidad de esta reacción en condiciones normales es baja. El SO_2 , tiene propiedades desinfectantes y por esto se utilizó durante siglos en la desinfección, p. ej. de las cubas de vino quemando azufre en su interior. También se utiliza en la industria alimenticia como conservante y antioxidante (E220), generalmente de zumos, frutos secos, mermeladas y vino.

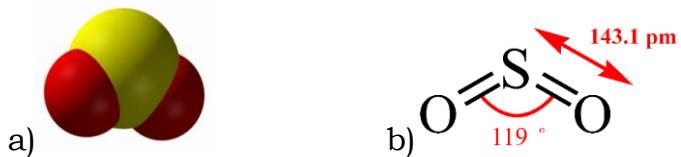
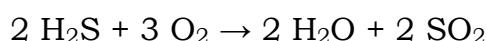
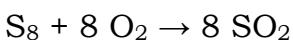
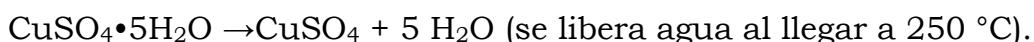
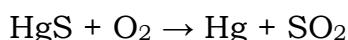
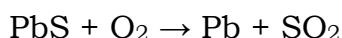
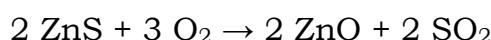
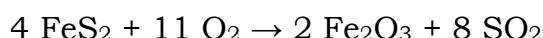


Figura 3.15. a) Representación en 3D de una molécula de dióxido de azufre. b) Estructura de la molécula del dióxido de azufre. Fuente: https://es.wikipedia.org/wiki/Di%C3%B3xido_de_azufre.

El dióxido de azufre es el principal causante de la lluvia ácida ya que en la atmósfera es transformado en ácido sulfúrico. Es liberado en muchos procesos de combustión ya que los combustibles como el carbón, el petróleo, el diésel o el gas natural contienen ciertas cantidades de compuestos que contienen azufre. Por estas razones se intenta eliminar estos compuestos antes de su combustión, p. ej. mediante la hidrodesulfuración en los derivados del petróleo o con lavados del gas natural haciéndolo más “dulce”. También los procesos metalúrgicos liberan ciertas cantidades de este gas debido a que se emplean frecuentemente los metales en forma de sulfuros. En la naturaleza, el dióxido de azufre se encuentra sobre todo en las proximidades de los volcanes, cuyas erupciones pueden liberar cantidades importantes del mismo. El dióxido de azufre se forma en el proceso de combustión del azufre y del sulfuro de hidrógeno:



También puede obtenerse por el tostado de minerales sulfurados tales como pirita (FeS), wurtzita o blenda (ZnS), galena (PbS), calcantita (CuSO₄) y el cinabrio (HgS):



Al mismo tiempo se produce dióxido de azufre (SO₂):

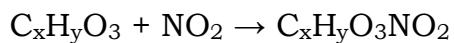


Por reacción del ácido sulfúrico con cobre elemental:



3.3.5. PAN o nitrato de peroxyacilo

El PAN o nitrato de peroxyacilo es un contaminante atmosférico que combinado con el ozono, O₃, óxidos de nitrógeno, los Compuestos Orgánicos Volátiles (COV) y la acción de la radiación solar, provoca el fenómeno del smog fotoquímico. Son los nitratos producidos en el equilibrio térmico con radicales ‘peroxi’ orgánicos por la oxidación en fase gaseosa de una variedad de compuestos orgánicos volátiles, o aldehídos y otros compuestos orgánicos volátiles oxigenados oxidantes en presencia de NO₂. Se forma por la combinación de una serie de compuestos en presencia de luz solar. La ecuación general es:



El PAN es tóxico e irritante, ya que se disuelve más fácilmente en agua que el ozono. Es una sustancia peligrosa que produce efectos lacrimógenos, causando irritación en los ojos en concentraciones de sólo unas pocas partes por mil millones; en concentraciones más altas en el ambiente, causa daños a la vegetación (Stephens, 1969) y puede producir cáncer de piel.

Es un buen marcador de la fuente de compuestos orgánicos volátiles, ya sea como biogénico o antropogénico, que es útil en el estudio de los efectos globales y locales de los contaminantes [34,35]. Puede ser empleado, puesto que es uno de sus componentes principales y por sus características, como el indicador del smog fotoquímico. Las concentraciones de PAN, mostraron fuertes correlaciones con otras partículas según un estudio realizado por investigadores de la ‘Seoul Metropolitan Government Institute of Public Health and Environment, Seoul, Republic of Korea’ [36]. Los resultados de este estudio sugieren que las concentraciones de PAN fueron afectadas sensitivamente por la estabilidad atmosférica, la deposición húmeda de NO₂, la dirección del viento, y otros factores. El PAN está formado por la reacción entre el NO₂ y radicales de peroxyacetilo (PA) (CH₃C(O)OO), originados por la oxidación fotoquímica de muchos hidrocarburos diferentes y productos de oxidación asociados, tales como aldehídos y cetonas (Seinfeld, 1986).

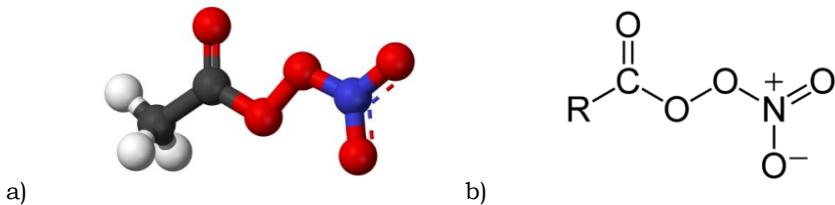
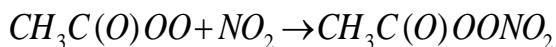
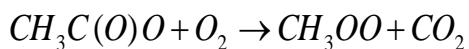
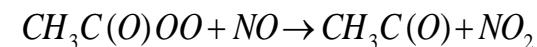
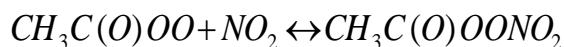


Figura 3.16. a) Representación en 3D de una molécula de PAN. b) Estructura de la molécula de PAN.
Fuente: https://en.wikipedia.org/wiki/Peroxyacetyl_nitrate

Las reacciones, que se ocasionan en la formación del peroxiacetilnitrato (PAN), se escriben de la siguiente manera:



Debido a la reacción rápida de $CH_3C(O)O$, las reacciones anteriores se pueden escribir de forma simplificada como sigue:



Las moléculas de PAN se componen de átomos de carbono, hidrógeno, oxígeno y nitrógeno. Uno de los componentes del PAN es el dióxido de nitrógeno (NO_2), el cual se forma a partir de los gases de escape provenientes de los automóviles que funcionan con combustibles fósiles, de las centrales eléctricas a carbón y de otros procesos industriales. Un segundo grupo de sustancias que pueden formar PAN, son los denominados compuestos orgánicos volátiles (COV), liberados al aire en forma de vapores, provenientes de la gasolina, pintura, solventes y pesticidas. Luego, estos compuestos se combinan con el oxígeno y el dióxido de nitrógeno para formar PAN. La energía de la luz solar facilita estas reacciones químicas a través de un proceso llamado fotodisociación. En altas concentraciones, es bastante dañino para los materiales vegetales, a los cuales “quema” produciéndose una reacción química semejante a la combustión lenta. Característicamente, los efectos del PAN sobre los vegetales aparecen como acristalamiento o bronceado de la superficie inferior de las hojas, que pueden producir colapso de tejidos, clorosis y caída de dichas hojas.

Los nitratos peroxyacílicos (PAN) son fuertes fitotoxicantes e importantes irritantes oculares en atmósferas contaminadas por los efluentes de la combustión orgánica. Los PAN son compuestos altamente inestables, en forma condensada pura, son relativamente estables en fase de vapor en la atmósfera. Las elevadas concentraciones de dichos compuestos permanecen durante toda la noche en atmósferas estancadas, mientras que las concentraciones de ozono (O_3) y otros materiales oxidantes se acercan a cero rápidamente cuando el sol se pone.

Es factible que las concentraciones de PAN puedan acumularse día a día, cuando las condiciones atmosféricas son extremadamente estables [37]. El nitrato de peroxyacilo (PAN), puede ser transportado durante largos periodos por la troposfera, debido a su estabilidad (en condiciones de baja temperatura), puesto que tiene una vida atmosférica prolongada. Como tal, el PAN puede desempeñar un papel esencial como un depósito de dióxido de nitrógeno (NO_2), y afectar así la producción de ozono troposférico (O_3) en áreas relativamente remotas (Zhang et al., 2009).

En el citado estudio sobre el PAN en atmósferas urbanas, realizado por Jun-Bok Lee, (et al). 2013, las concentraciones de peroxyacetilnitrato (PAN), mostraron fuertes correlaciones positivas con materia particulada (número total de partículas en suspensión -TSP, PM10 y PM2.5), y el carbono orgánico (OC) en partículas en todas las estaciones del año. Especialmente, se observó una correlación significativa entre el peroxyacetilnitrato PAN y el material particulado, en invierno cuando permanecían unas determinadas condiciones de estabilidad atmosférica. Las concentraciones de PAN, en una de las estaciones de medición, también fueron afectadas por las direcciones del viento (los vientos del oeste) en comparación con las mediciones de otra estación. En consecuencia, las concentraciones de PAN podrían influir en la calidad del aire urbano, al mantener una fuerte correlación con las partículas (PM) y el carbono orgánico (OC), independientemente de la estación.

3.3.6. Ozono (O_3)

El ozono (O_3) es una sustancia cuya molécula está compuesta por tres átomos de oxígeno, formada al disociarse los 2 átomos que componen el gas de oxígeno. Cada átomo de oxígeno liberado se une a otra molécula de oxígeno (O_2), formando moléculas de Ozono (O_3). En 1840, Christian Friedrich Schönbein propuso que fuera un compuesto químico distinto, nombrándolo con el verbo griego ‘ozein’, a causa del olor peculiar que se percibe durante las tormentas eléctricas eléctricas.

En 1865, Jacques-Louis Soret determinó la fórmula del ozono (O_3). Lo que fue confirmado por Schönbein en 1867. El ozono se descompone rápidamente en presencia de oxígeno a temperaturas mayores de 100 °C y en presencia de catalizadores como el dióxido de manganeso (MnO_2) a temperatura ambiente. A temperatura y presión ambientales el ozono es un gas de olor acre y generalmente incoloro, pero en grandes concentraciones puede volverse ligeramente azulado. Si se respira en grandes cantidades, puede provocar una irritación en los ojos y/o garganta, la cual suele pasar luego de respirar aire fresco por algunos minutos.

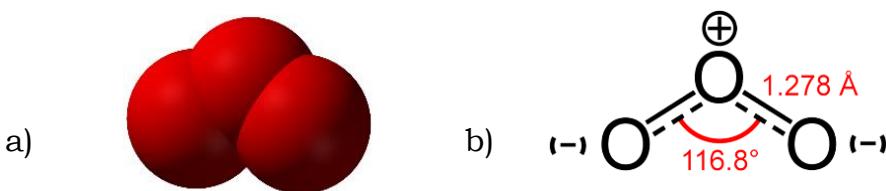


Figura 3.17. a) Representación 3D de una molécula de ozono b) Estructura de la molécula de ozono.
Fuente: <https://es.wikipedia.org/wiki/Ozono>

La mayor parte del ozono, denominado ozono estratosférico, se encuentra en la estratosfera (entre 12 – 40 km de altura), y es lo que popularmente conocemos como capa de ozono, que desempeña una función protectora para la vida en la Tierra y actúa como filtro de la radiación ultravioleta que llega a la superficie. La disminución de esta capa de ozono provoca que recibamos mayor radiación ultravioleta con los consiguientes efectos perjudiciales para la salud. En la Figura 3.18, se puede ver las reacciones del Ozono (O_3) en función de la capa terrestre.

Por otra parte, el ozono troposférico es aquel que se encuentra en la zona baja de la atmósfera, en el aire que respiramos y se trata de un contaminante que puede producir daños en la salud humana y en la vegetación dado su fuerte carácter oxidante. Este ozono troposférico es el denominado “ozono malo”, que puede tener consecuencias directas o indirectas en la salud humana, en el crecimiento y fisiología de las plantas o en el funcionamiento de los ecosistemas, potenciando el efecto invernadero.

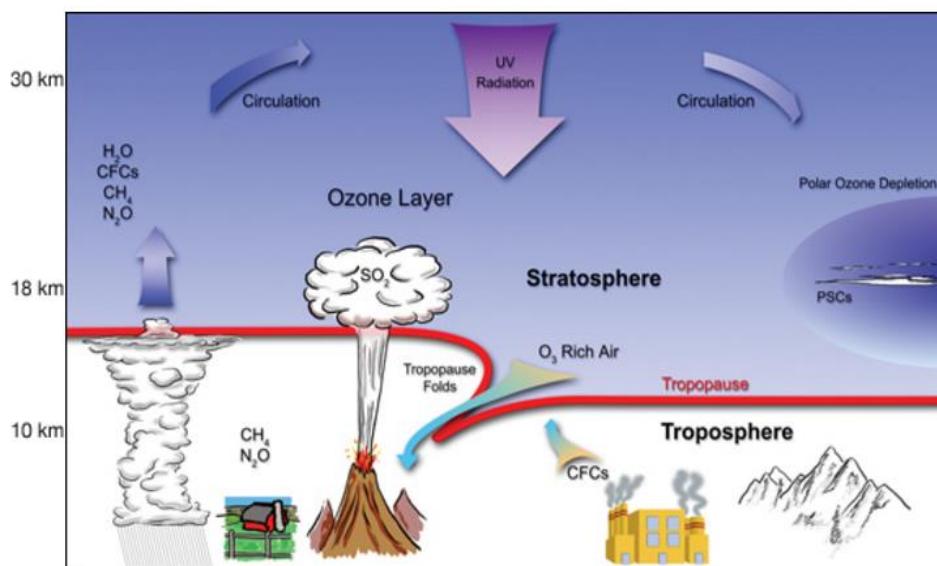
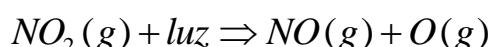


Figura 3.18. Ozono (O_3) en función de la capa terrestre. Fuente: Barbara Summey, SSAI.

3.3.6.1. El Ozono Troposférico

El ozono (O_3), es un gas muy oxidante y reactivo presente en la atmósfera en muy bajas concentraciones y con diferentes características en función del lugar donde se encuentre. Es, al mismo tiempo, un contaminante secundario de origen fotoquímico, es decir, no emitido directamente a la atmósfera, sino formado mediante reacciones fotoquímicas (en presencia de luz solar) a partir de contaminantes primarios como óxidos de nitrógeno y compuestos orgánicos volátiles (COVs). En condiciones favorables de altas temperaturas y energía solar, estos contaminantes primarios reaccionan con el oxígeno atmosférico según la secuencia de reacciones siguientes:



El aumento de la cantidad de ozono se debe principalmente a emisiones industriales, emisiones de vehículos, calefacciones, etc. El ozono, en niveles bajos, proviene de emisiones naturales de compuestos orgánicos volátiles (COVs), NO_x y CO (precursores de ozono), así como del ozono estratosférico que en ocasiones desciende a la superficie terrestre.

Las fuentes naturales de los precursores de ozono incluyen las emisiones de plantas y suelos, los incendios forestales y los rayos durante las tormentas eléctricas. En muchos lugares remotos de latitud media se observan altas concentraciones de ozono a finales del invierno y la primavera, sobre todo en alturas densamente pobladas que emiten los precursores necesarios.

Cuando se encuentra en unas determinadas concentraciones, el O₃ puede considerarse un contaminante atmosférico por sus efectos nocivos. Los precursores del ozono son fundamentalmente los óxidos de nitrógeno (NOx) y los compuestos orgánicos volátiles (COVs) los cuales se emiten de manera natural o como consecuencia de las actividades humanas. La época con mayores niveles de ozono suele coincidir con el verano debido a una mayor presencia de radiación solar.



Figura 3.19. Formación del Ozono Troposférico (O₃) en las capas bajas de la atmósfera. Fuente: Gobierno de Aragón, http://www.aragonaire.es/ozone.php?n_action=tropospheric.

Los productos químicos que destruyen el ozono están formados como hemos dicho por procesos industriales y naturales, con la excepción de la inyección volcánica y el escape de los aviones. Dichas sustancias químicas llegan a la estratosfera como consecuencia de las corrientes de aire ascendentes en los trópicos.

El metano (CH_4), los clorofluorocarbonos (CFC), el óxido nitroso (N_2O) y el agua, se inyectan en la estratosfera a través de acumulaciones de nubes. Estos compuestos se descomponen por la radiación ultravioleta en la estratosfera y originan subproductos de la descomposición de estas sustancias químicas formando “radicales”, como el dióxido de nitrógeno (NO_2) y el monóxido de cloro (ClO), que desempeñan un papel activo en la destrucción del ozono.

Table 10.1 Premature deaths attributable to $\text{PM}_{2.5}$, NO_2 and O_3 exposure in 41 European countries and the EU-28 in 2013

Country	Population	$\text{PM}_{2.5}$		NO_2		O_3	
		Annual mean (%)	Premature deaths	Annual mean (%)	Premature deaths	SOMO35 (¹)	Premature deaths
France	63 697 865	14.5	45 120	18.7	8 230	4 098	1 780
Germany	80 523 746	14.2	73 400	20.4	10 610	3 506	2 500
Greece	11 003 615	19.7	13 730	14.6	1 490	8 532	840
Hungary	9 908 798	18.2	12 890	16.8	390	4 604	460
Ireland	4 591 087	9.2	1 520	11.6	30	2 043	50
Italy	59 685 227	18.2	66 630	24.5	21 040	6 576	3 380
Netherlands	16 779 575	14.3	11 530	21.3	1 820	2 410	270
Poland	38 062 535	22.8	48 270	16.1	1 610	3 792	1 150
Portugal	9 918 548	10.0	6 070	14.0	150	5 091	420
Romania	20 020 074	18.5	25 330	17.9	1 900	2 221	430
Slovakia	5 410 836	20.1	5 620	16.0	< 5	5 116	200
Slovenia	2 058 821	17.4	1 960	17.6	150	6 540	100
Spain	44 454 505	11.0	23 940	18.0	4 280	5 895	1 760
.
.
.
Total (²)			467 000		71 000		17 000
EU-28 (²)			436 000		68 000		16 000

Tabla 3.3. Muertes prematuras derivadas de la exposición a los contaminantes indicados. Fuente: Adaptada del informe para la Agencia Europea del Medioambiente (EEA - European Environment Agency), 2016.

(¹) Las cifras Total y UE-28, se redondean hacia arriba o hacia abajo al millar más cercano. Los totales nacionales, se redondean a la decena más cercana.

(²) La media anual (en $\mu\text{g}/\text{m}^3$) y la SOMO35 (en $\mu\text{g}/\text{m}^3$ día), expresada como concentración ponderada por población, se obtiene según la metodología descrita por ETC / ACM (2016b) y no sólo desde estaciones de monitoreo.

Los aerosoles y las nubes pueden acelerar la pérdida de ozono a través de reacciones en las nubes. Así, las nubes volcánicas y las nubes estratosféricas polares pueden contribuir indirectamente a la pérdida de ozono.

España constituye uno de los principales focos de ozono “malo” o troposférico, en el contexto de los 28 países que forman la UE. En el informe de la Agencia Europea del Medio Ambiente (EEA), para el año 2016, los valores de O₃ se pueden observar en la Tabla 3.3.

3.3.6.2. Índice de concentración media anual de O₃ (SOMO35)

La legislación vigente en materia de calidad del aire establece para el ozono un valor objetivo para la protección de la salud humana de 120 µg/m³ (60 partes por billón). Este valor no deberá superarse más de 25 días por cada año civil de promedio, en un período de 3 años. No obstante, la Organización Mundial de la Salud (OMS) definió el *Índice de concentración media anual de O₃* (SOMO35) para evaluar la exposición de la población al ozono, basado en un umbral máximo de concentración –70 microgramos de ozono por m³ (35 partes por billón)–. Este índice se emplea para evaluar el daño a la salud de la exposición al ozono, de forma global y continuada en el tiempo.

La Directiva sobre la calidad del aire (UE, 2008) establece objetivos para la protección de la salud humana y la protección de la vegetación. En ella, se especifica un umbral medio máximo diario de 8 horas (120 µg/m³) para evitar riesgos en la salud de las personas. El objetivo sería que en una estación de monitoreo no se exceda el umbral límite fijado por las autoridades, más de 25 días al año (correspondiente al percentil 93.2), determinado como un promedio de 3 años, a partir del 1 de enero de 2010. El objetivo a largo plazo, para el que no se ha fijado ninguna fecha en la Directiva sobre la calidad del aire ambiente (UE, 2008), no supere el umbral en ninguna de las mediciones efectuadas.

En las AQG (en inglés- Air Quality Guidelines) de la OMS, para el O₃ se establece una concentración media diaria máxima de 8 horas, de 100 µg/m³ (OMS, 2006). Las normas europeas de calidad del aire según la Directiva de la UE sobre la calidad del aire ambiente y las directrices de la OMS sobre calidad del aire para el ozono (O₃), se dan en la Tabla 3.4 a continuación.

Air-quality standards for O₃ as defined in the EU Ambient Air Quality Directive and WHO air quality guidelines				
Averaging period	EU Air Quality Directive		WHO AQG	UNECE CLRTAP critical level
	Objective and legal nature	Concentration		
Maximum daily 8-hour mean	Human health long-term objective	120 µg/m ³	100 µg/m ³	
	Human health target value	120 µg/m ³ , not to be exceeded on more than 25 days per year averaged over 3 years		
AOT40 accumulated over May to July	Vegetation long-term objective	6 000 (µg/m ³).h		10 000 (µg/m ³).h (protection of forests)
	Vegetation target value	18 000 (µg/m ³).h averaged over 5 years		
1 hour	Information threshold	180 µg/m ³		
	Alert threshold	240 µg/m ³		

Tabla 3.4. Normas de calidad del aire para O₃ según se definen en la Directiva de la UE sobre la calidad del aire ambiente y las directrices de la OMS sobre calidad del aire.

Fuente: Agencia Europea del Medioambiente (EEA - European Environment Agency), 2016.

Existen también otros dos tipos de umbrales que se emplean, como referencia, para la protección de la salud de las personas. Son los denominados: umbrales de «información pública» y «alerta». En caso de incumplimiento del umbral de *información pública*, las autoridades de ese país están obligadas a notificarlas mediante un aviso de información pública. Cuando se supera el *umbral de alerta* durante tres horas consecutivas, el país afectado debe elaborar un plan de acción a corto plazo, de conformidad con las disposiciones específicas establecidas en la Directiva sobre la calidad del aire ambiente (UE, 2008).

3.3.6.3. Estado y tendencias en las concentraciones de ozono

Dado que la formación de O₃ requiere luz solar, las concentraciones de O₃ muestran un claro aumento a medida que se va de las partes septentrionales, en el caso de los países nórdicos, a las partes meridionales de Europa, dándose las concentraciones más altas en algunos países mediterráneos. La concentración de O₃ aumenta típicamente con altitud en los primeros kilómetros de la troposfera. Por lo expuesto, se pueden observar concentraciones más altas de O₃ en aquellas estaciones que se encuentran a mayor altitud.

Cerca del suelo, y, de las fuentes de NO_x, el O₃ se agota por deposición superficial y la reacción del NO emitido, para formar NO₂. Por lo tanto, en contraste con otros contaminantes, las concentraciones de O₃ son generalmente más altas en las zonas rurales, y más bajas en los sitios urbanos e incluso más bajas en los lugares de abundante tráfico. Las altas concentraciones de O₃ que se producen en las estaciones urbanas, especialmente en caso de superación del umbral de información (ver Tabla 2.4.), son atribuibles a su formación en grandes áreas urbanas durante episodios de alta radiación solar y altas temperaturas. Las emisiones locales y regionales de gases precursores (compuestos orgánicos volátiles (COVs), NO_x y CO), juegan un papel importante en la formación de ozono, especialmente a favor del viento de grandes áreas urbanizadas o industrializadas, como se demuestra en el caso de España por Querol et. al. (2016) [38].

3.3.6.4. Concentraciones en relación con los valores objetivo para la protección de la salud

En 2014, en 16 países [39] de la UE-28, ver Figura 3.20, se registraron concentraciones por encima del valor permitido de O₃, en más de 25 ocasiones. En total, el 11% de las estaciones (con una cobertura de datos igual o superior al 75%, para todos los países que comunicaron datos de O₃ a la EEA), tal y como se aprecia en la Figura 2.20, mostraron concentraciones por encima del valor objetivo para la protección de la salud humana en 2014, que es considerablemente menor que en 2013.

Por otra parte, sólo el 14% de estas estaciones cumplían el objetivo a largo plazo (no superar el umbral). De estas estaciones que lo cumplían, el 59% estaban en áreas de fondo, el 21% se encontraban en áreas industriales y el 20% estaban en áreas de tráfico rodado. las lecturas estaban en conformidad con el valor AQG de la OMS, fijado para la protección de la salud humana, para O₃ (media de 8 horas de 100 µg/m³), en menos del 4% de todas las estaciones y en sólo 5 de 503 estaciones de fondo rural durante 2014, Cuatro en la isla de Irlanda (Irlanda y Reino Unido) y una en Noruega.

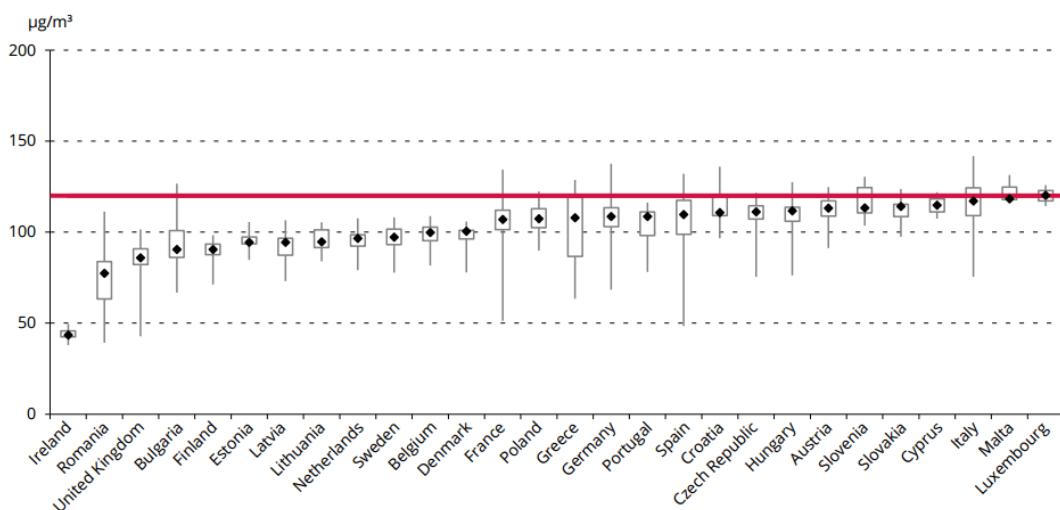


Figura 3.20. Concentraciones de O₃ en relación con el valor objetivo de 2014 en la UE-28.
Fuente: Agencia Europea del Medioambiente (EEA - European Environment Agency), 2016.

3.3.6.5. Relación entre las emisiones y concentraciones de precursores de ozono y las concentraciones ambientales de ozono

Las reducciones de las emisiones antropogénicas, debidas a actividades y procesos realizados por el hombre, de gases precursores O₃ en Europa, no han conducido a reducciones equivalentes de las concentraciones de O₃, ya que la relación de la concentración de O₃ con los precursores emitidos no es lineal. En los lugares de tráfico, la interacción entre O₃ y NO_x es el proceso dominante. El hecho de que se hayan reducido las emisiones de NO_x y, en particular, la relación entre las emisiones NO y NO_x (para vehículos diesel), conduce a que se consuma menos O₃ en la reacción de valoración con el NO.

Para los demás precursores O₃, como son los compuestos orgánicos volátiles exentos de metano - COVNM, la Directiva sobre la calidad del aire ambiente (UE, 2008) exige que los Estados miembros midan la concentración ambiental de los compuestos COV al menos en una estación, con el fin de analizar las tendencias de los precursores O₃, y comprobar la coherencia de los históricos de emisiones; así como, servir de ayuda con la finalidad de atribuir las fuentes de emisión a las concentraciones de contaminación observadas.

Existen largas series temporales para los compuestos aromáticos menos reactivos (benceno, tolueno). En el 80% de las estaciones de benceno (101 de un total de 127 estaciones de medición de benceno, de las cuales 55 son estaciones de tráfico urbanas o suburbanas), se observa una tendencia descendente significativa. Para el tolueno (35 estaciones de las que 22 son estaciones de tráfico), se dispone de un menor número de series temporales, aunque, se observa una tendencia descendente significativa en todas estas estaciones. Entre 2000 y 2014, tanto el benceno como el tolueno muestran una disminución de más del 70%, lo que refleja principalmente, teniendo en cuenta la selección de estaciones, la reducción cercana a las fuentes de tráfico.

3.3.7. Smog fotoquímico

El acrónimo ‘smog’ procede del inglés ‘smoke’ (humo), y ‘fog’ (niebla). Existen principalmente dos tipos:

el smog clásico, que se origina por la quema de grandes cantidades de carbón, produciendo una niebla rica en dióxido de azufre (SO₂); y el smog oxidante o fotoquímico, derivado de diversas reacciones, que constituye una de las principales causas de contaminación del aire en zonas urbanas.

Se denomina ‘smog’ fotoquímico (en español, “esmog fotoquímico”), a la contaminación del aire, por ozono derivado de reacciones fotoquímicas principalmente en áreas urbanas, y otros compuestos que más adelante detallaremos.



Figura 3.29. Una ciudad contaminada.
Fuente: Washington Post. Shutterstock/grynnold

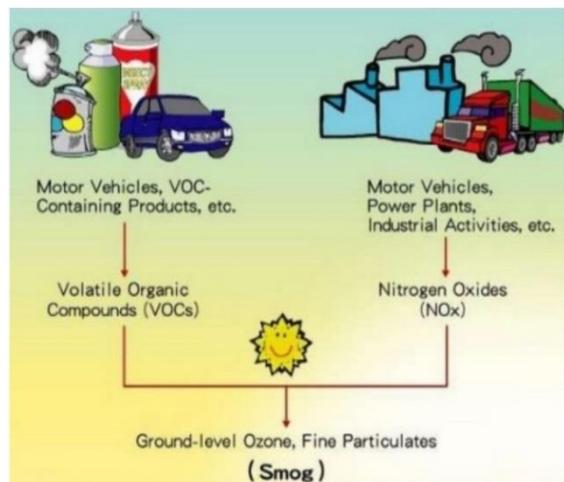


Figura 3.21. Formación de “smog” fotoquímico.
Fuente: Prof. MSc. H. Hermes

3.3.7.1. Origen del ‘smog’ fotoquímico

Según el origen del ‘smog’ fotoquímico, se distinguen dos tipos básicos en función de qué fuentes de emisión lo hayan originado; así tendremos dos tipos de misión posible: biogénicas y antropogénicas.

3.3.7.1.1. Fuentes de emisiones biogénicas

Las emisiones biogénicas son las producidas por causas naturales, y se extienden a lo largo de grandes áreas; ya que, es poco probable que coexistan varias fuentes de emisión en el mismo espacio y tiempo, reduciéndose así sus efectos.

Los incendios forestales, los rayos y los procesos microbianos que se producen en la naturaleza, generan óxidos de nitrógeno (NO_x). Los COV's se producen a partir de la evaporación de compuestos naturales, como los terpenos (compuestos orgánicos derivados del isopreno, un hidrocarburo de 5 átomos de carbono). Dichos hidrocarburos posibilitan que los aceites puedan quemarse.

3.3.7.1.2. Fuentes de emisiones antropogénicas

Las emisiones antropogénicas son producidas por la acción del hombre, y se concentran habitualmente cerca de sus fuentes de emisión; como ocurre en las grandes urbes altamente industrializadas, donde se desarrollan procesos de producción, y circulación densa de vehículos rodados que consumen combustibles fósiles. Los óxidos de nitrógeno (NO_x) se producen principalmente a partir de la combustión de los combustibles fósiles, en centrales eléctricas y vehículos de motor.

Los compuestos orgánicos volátiles (COVs) se forman a partir de la combustión incompleta de combustibles fósiles, de la evaporación de disolventes y combustibles y de la quema de materia vegetal en estufas de leña.

El smog fotoquímico es una mezcla de contaminantes, que se forman cuando los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV), emitidos a la atmósfera, reaccionan atrapados por una inversión térmica, y, son irradiados por la luz solar creando una bruma marrón característica, que se puede apreciar sobre las ciudades.

Tiende a ocurrir más a menudo en verano, porque eso es cuando tenemos más luz del sol. Se forma, cuando los fotones de la luz solar chocan con moléculas de diversos tipos de agentes contaminantes en la atmósfera, originando reacciones químicas. El smog fotoquímico contiene óxidos de nitrógeno (NO_x), componentes orgánicos volátiles (COV), ozono troposférico (O_3) y nitrato peroxyacetílico (PAN).

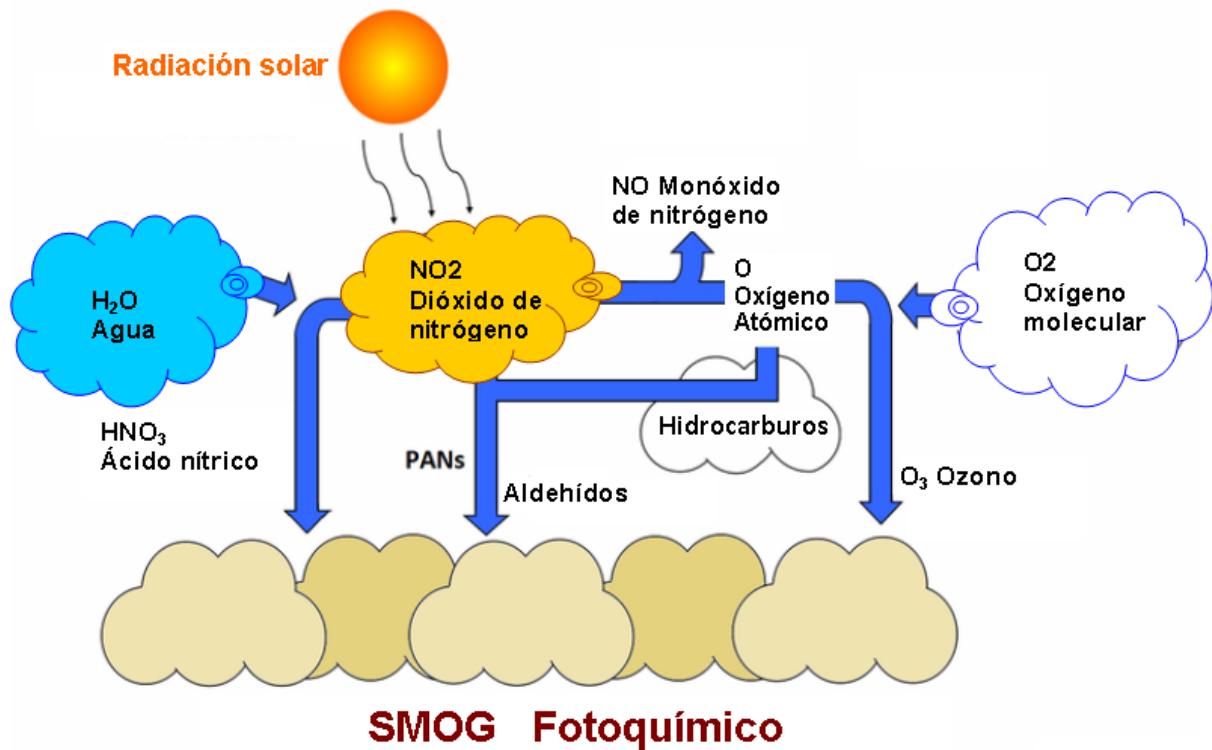


Figura 3.33. Formación de smog fotoquímico. Adaptado de vivos en el ambiente: G. Tyler Miller, Jr. y D. Hackett, “smog fotoquímico y de industria,” en vivo en un ámbito , 2^a ed. EE.UU.: Nelson, 2011, ch.20, SEC.3, pp.465-471

Los principales contaminantes secundarios que se incluyen en el smog, son los siguientes:

- Ozono (O_3)
- Ácido sulfúrico y ácido nítrico (componentes de la lluvia ácida)
- Material particulado (PM)
- Dióxido de nitrógeno (NO_2)
- Peroxiacetilnitrato (PAN)
- y otros

Estas sustancias esencialmente “cocinan” en la atmósfera, y por lo general se encuentran a favor del viento emisiones primarias debido al tiempo que se necesita para producirlos. Cuando los contaminantes primarios no pueden dispersarse debido a las capas originadas por la inversión térmica en la atmósfera, se forma el smog sobre la zona en que se produjeron, por lo que el smog es muy característico de las ciudades densas y cálidas. Los contaminantes secundarios son muy sensibles a los patrones climatológicos.

El ozono troposférico (O_3), que se encuentra en la zona más próxima a la superficie de la Tierra, se forma como consecuencia de las reacciones inducidas por la luz solar, en las que participan los óxidos de nitrógeno y los hidrocarburos presentes en el aire. El O_3 es el componente más dañino del smog fotoquímico, y, causa daños importantes en la salud cuando se encuentra en concentraciones muy altas, además de frenar el crecimiento de las plantas y de los árboles.

El smog fotoquímico tiene efectos sobre el medio ambiente, la salud de las personas e incluso sobre diversos materiales.

3.3.7.2. Efectos del smog fotoquímico sobre el hábitat y la salud

3.3.7.2.1. Sobre el medioambiente

El efecto más visible es la bruma marrón que se puede ver sobre muchas ciudades, la cual es causada por partículas líquidas y sólidas, muy pequeñas, que dispersan la luz. Productos químicos presentes en el smog fotoquímico, tales como los óxidos de nitrógeno, el ozono y el nitrato de peroxyacetilo (PAN), pueden afectar a las plantas. Dichas sustancias reducen e incluso podrían llegar a detener el crecimiento de las plantas, mediante la reducción de la fotosíntesis. Si bien pequeñas concentraciones de ozono pueden producir los efectos anteriormente descritos, el PAN es aún más tóxico para las plantas que el ozono.

3.3.7.2.2. Sobre la salud de las personas

La mayor preocupación para las personas, en relación al smog fotoquímico, es el efecto nocivo que tiene sobre nuestra salud. Las consecuencias de los principales contaminantes primarios y secundarios en el smog se dan en la Tabla 3.5, a continuación.

<u>Contaminante</u>	<u>Efecto</u>
Óxidos de nitrógeno (NO_x)	<ul style="list-style-type: none"> ▪ Afectan a los problemas de corazón y pulmones. ▪ Disminuye de la resistencia a las infecciones.
Compuestos orgánicos volátiles (COV)	<ul style="list-style-type: none"> ▪ Irritación de ojos ▪ Problemas respiratorios. ▪ Algunos compuestos podrían producir cáncer.
Ozono (O_3)	<ul style="list-style-type: none"> ▪ Tos y sibilancias. ▪ Irritación de ojos. ▪ Problemas respiratorios (especialmente para afecciones como el asma).
Nitrato de peroxiacetileno (PAN)	<ul style="list-style-type: none"> ▪ Irritación ocular. ▪ Problemas respiratorios.

Tabla 3.5. Consecuencias sobre la salud de los contaminantes contenidos en el smog fotoquímico.
Fuente EPA (Environment Protection Authority) Información 90/04.

3.3.7.3. Sobre los materiales

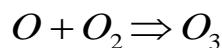
El ozono puede dañar varios de los materiales y objetos que se emplean de forma cotidiana. Puede causar el agrietamiento del caucho, la reducción de la resistencia a la tracción de los materiales textiles, el desvanecimiento de las fibras teñidas y el agrietamiento de la pintura. Debido al enorme potencial del O_3 para dañar obras de arte y libros, afectando al patrimonio cultural de las ciudades, algunos museos y bibliotecas han tomado medidas para minimizar dichos efectos, implementando medidas de protección suplementarias que hasta hace poco tiempo no se empleaban.

3.3.7.4. Formación del smog fotoquímico

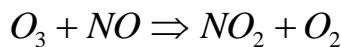
Seguidamente se detalla de forma simplificada la formación de smog fotoquímico. El dióxido de nitrógeno (NO_2) puede descomponerse, por la radiación de la luz solar, formándose óxido nítrico (NO) y un radical de oxígeno (O):



Los radicales del oxígeno reaccionan con el oxígeno atmosférico (O_2) para formar ozono (O_3):



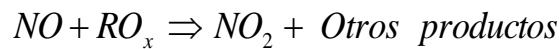
El ozono es consumido por el óxido nítrico para producir dióxido de nitrógeno y oxígeno:



Los productos nocivos, como el PAN, se originan por las reacciones de dióxido de nitrógeno con varios hidrocarburos \textcircled{R} , que son compuestos hechos de carbono, hidrógeno y otras sustancias:



La principal fuente de estos hidrocarburos son los compuestos orgánicos volátiles (COV). De forma similar, los compuestos orgánicos e inorgánicos oxigenados (RO_x) reaccionan con el óxido nítrico para producir más óxidos de nitrógeno:



La presencia de COV en las dos últimas reacciones, esencial. El ozono es consumido por el óxido nítrico, como se indica en la reacción: $\text{O}_3 + \text{NO} \Rightarrow \text{NO}_2 + \text{O}_2$. Sin embargo, cuando están presentes los compuestos orgánicos volátiles, tanto el monóxido de nitrógeno como el dióxido de nitrógeno, tal y como se indica en las dos últimas reacciones; permitiendo así, que el ozono se acumule a nivel del suelo [40].

3.3.7.5. Fenómenos atmosféricos que atrapan el smog. Inversión Térmica

El aire atmosférico se dispone en capas de acuerdo a su densidad quedando el aire más denso, con mayor peso, al nivel del suelo, mientras que el aire menos denso y con menor presión atmosférica, se coloca a mayor altura. La inversión térmica, es un fenómeno que se presenta en las noches despejadas, cuando el suelo ha perdido calor por radiación y las capas de aire cercanas a él se enfrián más rápido que las capas superiores, lo cual provoca que se genere un gradiente positivo de temperatura con la altitud. El suelo a su vez enfriá el aire en contacto con él, que se vuelve más frío y pesado que el resto de aire que se encuentra en la capa inmediatamente superior. Este fenómeno provoca que la capa de aire caliente quede atrapada entre las 2 capas de aire frío, sin poder circular; ya que, la presencia de la capa de aire frío cerca del suelo le confiere una gran estabilidad a la atmósfera, no habiendo convección térmica, ni fenómenos de transporte y difusión de gases. Esto causa una disminución de la velocidad de mezclado vertical, entre la región que hay entre las 2 capas frías de aire.

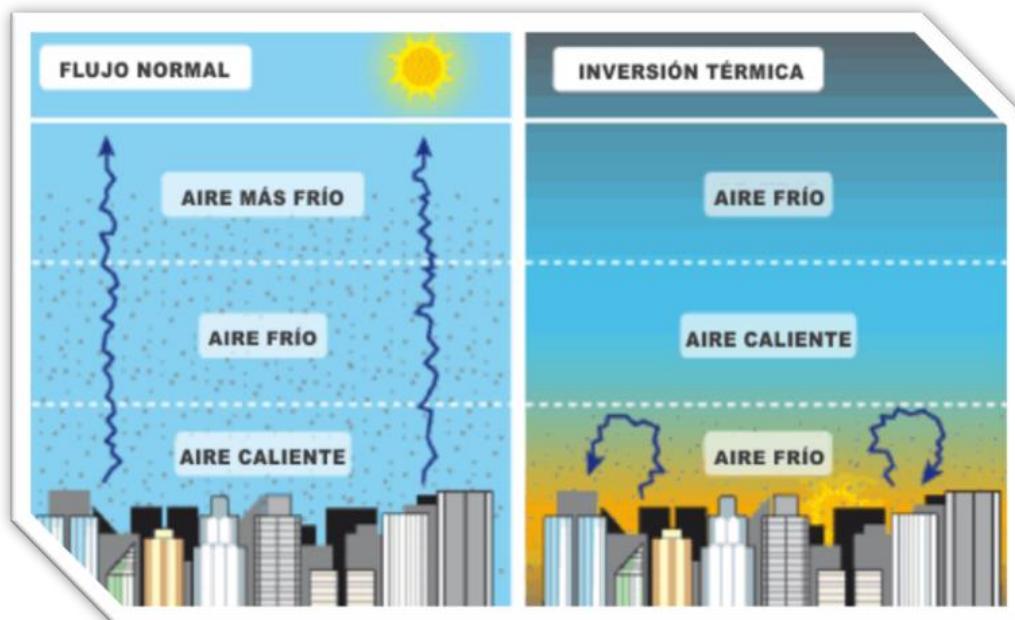


Figura 3.34. Flujo en día normal y día con inversión térmica. Fuente: http://contaminantes-atmosfericos.blogspot.com.es/2013_05_15_archive.html

Una inversión térmica ocurre cuando se da uno de los dos procesos siguientes y va usualmente acompañada por velocidades de viento bajas.

El primer proceso, se da cuando una capa superior de aire más caliente es soplada, atrapando la capa de aire frío por debajo de ella. Una variación de este proceso se da cuando una capa de aire más fría, tal como una brisa marina, es insuflada debajo de una capa más caliente, creando un efecto similar al anterior.

En *el segundo proceso*, la inversión de la radiación, ocurre generalmente durante la noche. El suelo se enfriá y, a su vez, enfriá la capa de aire más cercana a ella, dando como resultado que la capa de aire inferior sea más fría que el aire sobre ella, formando una inversión. En las noches despejadas, el suelo se enfriá rápidamente por radiación. El suelo, a su vez, enfriá el aire en contacto con él que se vuelve más frío y pesado que el que está en la capa inmediatamente superior. Al disminuir tanto, la convención térmica como la subsidencia atmosférica, disminuye la velocidad de mezclado vertical entre las dos capas de aire. Una inversión térmica puede llevar a que la contaminación aérea, como el smog, quede atrapada cerca del suelo, con efectos nocivos para la salud.

El efecto de la inversión también puede detener el fenómeno de convección, actuando como una capa aislante. La inversión térmica se rompe cuando se restablece la circulación normal en la troposfera, al calentarse el aire que está en contacto con el suelo. Esto puede suceder en cuestión de horas; pero en condiciones meteorológicas desfavorables, la inversión térmica puede prevalecer varios días.

3.3.8. Lluvia ácida

La lluvia ácida es un tipo de contaminación atmosférica secundaria y es debida a las emisiones originadas por el uso de combustibles fósiles en las operaciones industriales, combustión de desechos industriales, urbanos y agrícolas, transporte, agricultura y calefacciones domésticas; así como, por la utilización de fertilizantes y otros productos químicos para el campo.

La lluvia ácida causa daños en diversos materiales, alteraciones en la vegetación, y en los sistemas acuáticos [41]. En nuestros días la lluvia ácida presenta uno de los problemas ambientales que está padeciendo el planeta tierra, [Robertson, N. & Wunder, S. 2005].

Según diversos informes de varios países industrializados entre ellos EE.UU primer emisor de gases de efecto invernadero, se pone de manifiesto que la degradación del aire tiene un efecto significativo sobre las plantas; ya que, están viendo morir sus plantaciones por efecto de la lluvia ácida, fenómeno, que está afectando a varias zonas de cultivos y poniendo en peligro la sostenibilidad de los bosques y afectando por consiguiente a todo el ecosistema. Hay en marcha iniciativas de conservación, impulsadas globalmente en todo el mundo, que pretenden crear sistemas para los cuales los usuarios, de servicios de recursos ambientales, compensen a los regentes del recurso por la conservación de dichos servicios. Estos sistemas, denominados “pagos por servicios ambientales” (PSA), son una clase de instrumentos económicos diseñados para dar incentivos a los usuarios del suelo, de manera que continúen ofreciendo un servicio ambiental (ecológico) que beneficia a la sociedad, y podrían tener el potencial de proteger servicios ambientales. Al mismo tiempo los países en vías de desarrollo, que participen de las citadas iniciativas, pueden adquirir beneficios económicos, por mantener los ecosistemas sin perturbaciones que alteren su equilibrio [Robertson, N & Wunder, S. 2005] [42].

3.3.8.1. Efectos geoquímicos de la lluvia ácida

Se denomina capacidad de amortiguación a la facultad de los suelos de neutralizar la acidez del agua procedente de la lluvia ácida. Altas capacidades de amortiguación protegen los suelos de la acidificación. Las diferencias entre distintos suelos son una razón por la cual algunas unas áreas afectadas de lluvia ácida muestran más daño que otras. La capacidad de los suelos de los bosques para resistir la acidez depende del espesor y composición del suelo, así como del tipo de roca bajo el suelo del bosque. Los suelos de granito y las rocas volcánicas no favorecen la neutralización del ácido.

Los suelos más gruesos se comportan mejor que los delgados, debido a su mayor capacidad para el intercambio iónico; en cambio los suelos delgados, tienen menor capacidad de amortiguación porque los nutrientes contenidos en dichos suelos, son fácilmente lixiviados.

Los efectos ecológicos de la lluvia ácida también se manifiestan en medios acuáticos; donde la lluvia ácida fluye hacia las fuentes de agua, después de caer en valles, carreteras y edificaciones. Los lagos y riachuelos tienen un pH entre 6 y 8, siendo algunos lagos ácidos de forma natural, sin que sufren los efectos de la lluvia ácida. La lluvia ácida afecta principalmente a cuerpos de agua sensibles, que se encuentran en suelos con una limitada capacidad para neutralizar compuestos ácidos. Los lagos y los arroyos se acidifican, cuando el agua contenida en ellos y el suelo circundante, no tienen capacidad suficiente de amortiguación, para neutralizar la lluvia ácida que cae sobre ellos. En las zonas donde la capacidad de amortiguación es baja, la lluvia ácida también posibilita liberar el aluminio contenido en los suelos, mezclándose en lagos y arroyos. El aluminio es altamente tóxico para muchas especies de organismos acuáticos.

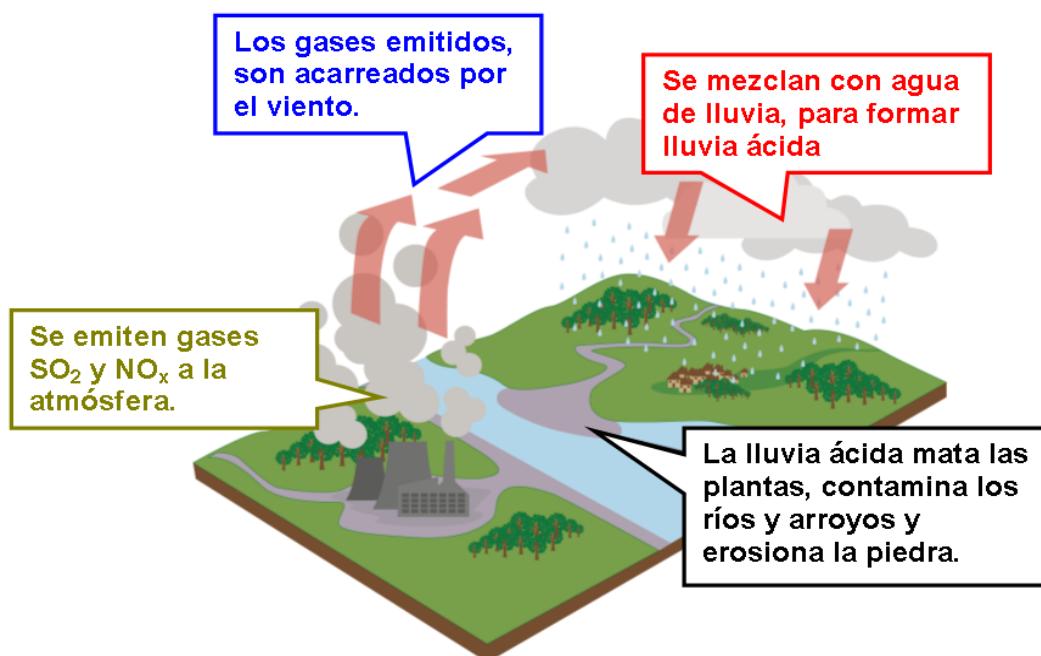


Figura 3.35. Origen y efectos de la lluvia ácida. Fuente: Adaptada de, https://www.edplace.com/worksheets_info/science/keystage3/year9/topic/94/3033/acid-rain

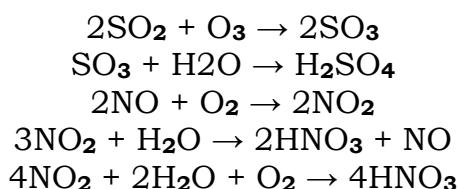
La lluvia ácida es uno de los problemas asociado a la contaminación atmosférica, las primeras apariciones de este fenómeno se observaron en Suecia (1848), Inglaterra (1877) y en Alemania (1867), donde se publicó que por alguna razón las lluvias eran más ácidas de lo normal. Lee M. & Thomas (1986) [43] señalaron que el agua de lluvia naturalmente es ácida, ya que su pH es aproximadamente 5.6.

Esto es debido a que el agua es un buen disolvente, y, cuando llueve se disuelve algo del dióxido de carbono (CO_2) presente en la atmósfera, convirtiéndose en la superficie en “ácido carbónico”. Por esta razón, los anteriores autores consideraron que ese daño ambiental no empeoraría si las emisiones ácidas no excedieran el nivel reportado en ese año. En la Figura 3.25, se muestran el origen y los efectos de la lluvia ácida.

3.3.8.2. Tipo de deposiciones de lluvia ácida

La lluvia limpia normal tiene un valor de pH comprendido entre 5.0 y 5.5, nivel levemente ácido. Llamamos lluvia ácida a toda precipitación con un $\text{pH} < 5.5$. Sin embargo, cuando la lluvia se combina con dióxido de azufre y óxidos de nitrógeno, se vuelve mucho más ácida.

La lluvia ácida típica tiene un valor de pH de entre 4.0 y 5.0. Una disminución en los valores de pH de 5.0 a 4.0 significa que la acidez es diez veces mayor. Cuando el pH tiene un valor, por debajo del que tendría la lluvia limpia, esa acidez se debe a la presencia de ácidos fuertes, como el ácido sulfúrico (H_2SO_4) y el ácido nítrico (HNO_3), que tienen como principales precursores los óxidos de azufre (SO_x) y óxidos de nitrógeno (NO_x) respectivamente, junto con la humedad de la atmósfera (H_2O). Lo anterior se explica en forma general por las siguientes reacciones [44]:



La lluvia, como hemos dicho anteriormente, se vuelve ácida debido a un descenso en su pH, estando el valor comprendido entre 4.0 a 4.2, según se puede observar en la Figura 3.26. Esto es debido a una combinación de dióxido de azufre (SO_2) y óxidos de nitrógeno (NO_x).

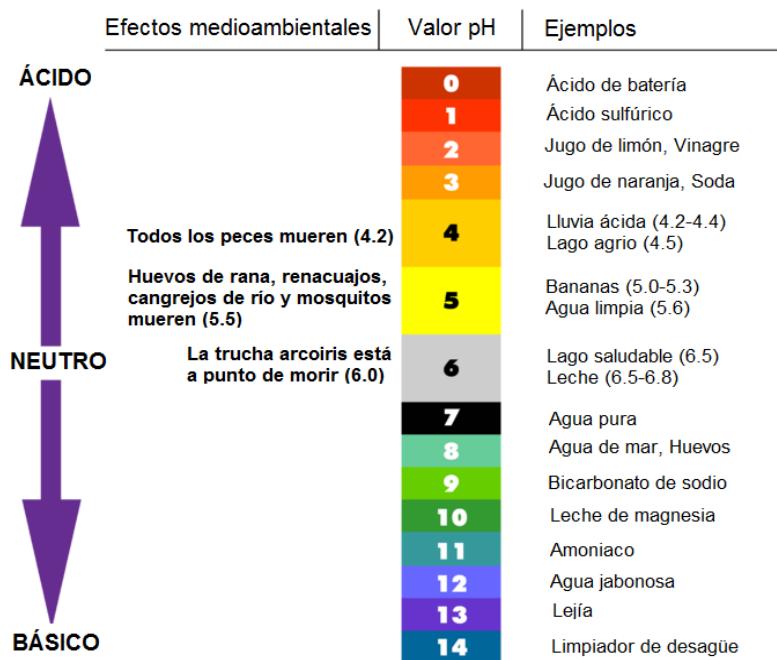


Figura 3.36. Escala del pH. Fuente: Adaptada de, Enviroment Canada <http://www.ns.ec.gc.ca/>

Dichos gases se encuentran en la atmósfera, en altas concentraciones, donde se mezclan y reaccionan con el agua, el oxígeno y otras substancias químicas, formando la lluvia ácida [Lee, J & Weber, D. 1982] [45].

La lluvia ácida puede producirse por precipitación, en el ciclo de agua, o bien por una simple deposición seca. En realidad, hay dos tipos de lluvia ácida, según su sedimentación, por lo que podrá depositarse en la superficie terrestre por sedimentación húmeda o por sedimentación seca.

3.3.8.2.1. Lluvia ácida mediante deposiciones húmedas.

La sustancia es arrastrada, disuelta o en suspensión, en forma lluvia, llovizna, nieve y rocío, volviendo al suelo, agua o superficie vegetal. La mayor parte de los SO_2 y NO_x que permanecen en el aire sufren un proceso de oxidación que da lugar a la formación de ácido sulfúrico (SO_4H_2) y ácido nítrico (NO_3H).

Estos ácidos se disuelven en las gotas de agua que forman las nubes y en las gotas de lluvia, retornando al suelo con las precipitaciones. Una parte de estos ácidos queda neutralizada por sustancias presentes en el aire tales como el amoniaco, formando iones de amonio (NH_4^+).

3.3.8.2.2. Lluvia ácida mediante deposiciones secas.

Una fracción de los óxidos vertidos a la atmósfera retornan a la superficie de la tierra en forma gaseosa o de aerosoles. Esto puede ocurrir cerca de las fuentes de emisión de los contaminantes o a grandes distancias (muchos kilómetros) de la misma, en función de las condiciones de dispersión. La deposición en seco es predominante en zonas próximas al foco emisor. En este tipo de deposición, se registran mayores niveles de acidez que en la deposición húmeda, debido a las elevadas concentraciones de sulfatos y nitratos suspendidas en el aire.

Ambos tipos de deposición pueden ser acarreados por el viento, o el agua recorriendo grandes distancias (Landsberg, H. 1984) [46].

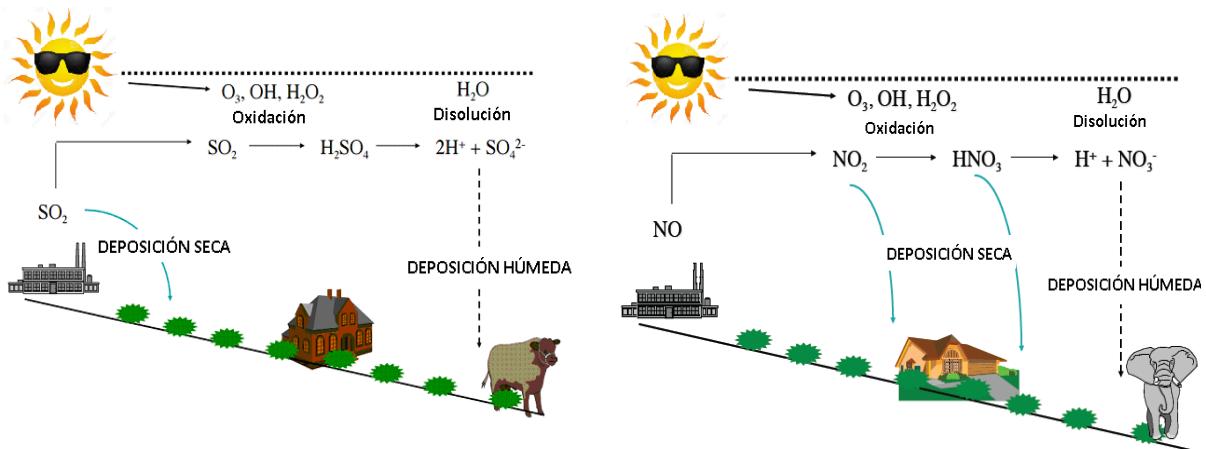


Figura 3.37. Tipos de deposiciones – Fuente: Adaptada, After Parks (1985).

<http://web.colby.edu/ch217public/files/2013/04/CH217-Acid-Rain-2016.pdf>

3.3.8.3. Origen de la lluvia ácida

La lluvia ácida es debida a las emisiones atmosféricas, típicamente originadas por el uso de combustibles fósiles en las operaciones industriales, combustión de desechos industriales, urbanos y agrícolas, transporte, agricultura y calefacciones domésticas; así como por la utilización de fertilizantes y otros productos químicos para el campo. La lluvia ácida causa daños en diversos materiales, alteraciones en la vegetación, y en los sistemas acuáticos [47].

Los contaminantes atmosféricos como el SO₂ y el NO_x se oxidan hasta convertirse en sulfatos (SO₄²⁻) y nitratos (NO₃⁻) que retornan al suelo combinados con vapor de agua, como ácido sulfúrico (H₂SO₄) y ácido nítrico (HNO₃), en forma de lluvia, llovizna, roció, nieve, niebla o granizo [48] [Manahan, S. 1994]. Las fuentes de origen de la lluvia ácida pueden ser debidas a la contaminación *biogénica* (producida por actividades naturales) y *antropogénica* (o producida por la acción del hombre), cada una con las siguientes características:

3.3.8.3.1. Por contaminación biogénica

La lluvia contaminante se puede formar de manera natural. Los volcanes, los pantanos, y las plantas en estado de descomposición desprenden dióxido de azufre, uno de los óxidos que participan en la formación de un tipo de lluvia ácida natural, que cae en cantidades inferiores a la artificial. Dicha lluvia ácida, que tiene su origen en las emisiones volcánicas, posee grandes cantidades de ácido Sulfúrico (H₂SO₄) y Ácido Nítrico (HNO₃) disueltos en gotas de agua. Ambos ácidos se forman al reaccionar el trióxido de azufre (SO₃) y el dióxido de nitrógeno (NO₂) con el agua (H₂O). De manera que la acidez de la lluvia alcanza niveles que oscilan entre 3.5 a 5.5 con respecto al pH normal del agua, que es aproximadamente 6.5. En condiciones prístinas se han encontrado precipitaciones con valores de pH inferiores a 5,6 debido a emisiones naturales de SO₂, NO_x y ácidos orgánicos, aunque los valores de pH típicos en precipitaciones ácidas, causadas por las emisiones antropogénicas, se encuentran entre 3,5 y 5,0 [Menz y Seip, 2004].

El ácido clorhídrico (HCl) es otro contribuyente significativo en la acidez de la lluvia (Herrera y Rodríguez, 2009). Las emisiones antropogénicas provienen principalmente de la producción de cloro como gas y HCl [Zhu y Liu, 2003]. Como fuente primaria de HCl, varios autores sugieren los volcanes (también la combustión de carbón) y los desechos sólidos, mientras que como fuente secundaria señalan la declorinación o eliminación del cloro, de aerosoles marinos y la oxidación in situ de los clorocarbonados de emisiones naturales.

3.3.8.3.2. Por contaminación antropogénica

La deposición de sustancias ácidas se debe, entre otras causas, a la emisión de óxidos de azufre y de nitrógeno que tienen como fuentes actividades industriales, transporte y quema de combustibles fósiles. Dichos compuestos originan la lluvia ácida, junto con otros tipos de contaminantes, que también pueden contribuir a su formación, como son los compuestos orgánicos volátiles (COVs) y los hidrocarburos, que pueden reaccionar por si mismos o combinarse [49] [Eichler, M. & Fagundes, L. 2004]. La contaminación antropogénica es debida principalmente, a las actividades que se citan a continuación:

a) Los combustibles fósiles

Los combustibles fósiles son compuestos hidrocarburos, formados por hidrógeno y carbono, pueden contener además pequeñas cantidades de otros componentes. Los combustibles fósiles empleados en procesos industriales son: el petróleo, el carbón y el gas natural. Estas materias, necesitan ser refinadas para convertirse en productos como la gasolina, el gasóleo y el queroseno. Algunos de esos hidrocarburos son transformados en plásticos, sustancias químicas, lubricantes y otros no combustibles [50] [Landsberg, H. 1984]. Un ejemplo donde se emplean los anteriores combustibles, lo constituyen las centrales termoeléctricas. Estas centrales, son instalaciones que produce energía eléctrica a partir de la combustión de carbón, fuel o gas en una caldera diseñada al efecto.

En las centrales termoeléctricas, el combustible se almacena en depósitos adyacentes, desde donde se suministra a la central, pasando a la caldera, donde tiene lugar la combustión. Debido al incremento de temperatura que se origina en la caldera se genera el vapor, a partir del agua que circula por una extensa red de tubos que cubren las paredes de la caldera. Dicho vapor, procedente del calentamiento de agua desmineralizada ($\leq 0.3 \mu\text{S}/\text{cm}$), se emplea para la producción de energía eléctrica mediante el giro de la turbina conectada al generador.

b) Fuentes de Óxidos de Azufre SO_2

El SO_2 es un gas incoloro, no explosivo, incombustible, de olor astringente. Entre los gases que contienen azufre, están: el dióxido de azufre (SO_2), trióxido de azufre(SO_3), sulfuro de hidrógeno(H_2S), anión sulfato SO_4^{2-} , donde el SO_2 es la emisión que más vamos a considerar, de los contaminantes con azufre. Las principales fuentes de emisión del SO_2 , son las centrales térmicas de carbón o combustibles líquidos, industrias donde se emplean combustibles fósiles, refinerías de petróleo y transporte.

c) Fuentes Óxidos de Nitrógeno NO_x

Los óxidos de nitrógeno están conformados por un grupo de compuestos químicos gaseosos muy reactivos, que son generados por las siguientes actividades antropogénicas: procesos de producción eléctrica e industriales, transporte por carretera empleando combustibles fósiles y actividades residenciales. Dentro de los óxidos de nitrógenos, a efectos de contaminación, se consideran fundamentalmente las siguientes especies:

Óxidos: óxido nitroso N_2O , dióxido de nitrógeno NO_2 , nitrato NO_3^- , óxido nítrico N_2O_5 .

Ácidos: ácido nitroso (HNO_2), ácido nítrico (HNO_3).

De donde los más importantes son: dióxido de nitrógeno (NO_2), monóxido de nitrógeno (NO) y óxido nitroso (N_2O) (concentraciones muy bajas).

El óxido nitroso (N_2O), es un gas que no participa en los procesos fotoquímicos troposféricos y no es tóxico. El monóxido de nitrógeno (NO), participa en los procesos fotoquímicos troposféricos es incoloro, inodoro y tóxico. Cuando el dióxido de nitrógeno (NO_2) penetra en la estratosfera, genera una de las vías de destrucción de ozono, mientras que en la baja atmósfera, y en presencia de luz solar, es uno de sus precursores [51], es pardo-rojizo y presenta olor asfixiante.

3.3.9. Contaminantes ‘Criterio’

La U.S. EPA (en inglés, United States Environmental Protection Agency), ha establecido estándares nacionales de calidad del aire ambiental (en inglés NAAQS) para seis de los contaminantes más comunes del aire. Dichos contaminantes son: monóxido de carbono (CO), plomo (Pb), ozono a nivel del suelo (O_3), partículas o aerosoles (PM), dióxido de nitrógeno (NO_2) y dióxido de azufre (SO_2). Estos contaminantes son conocidos como “Contaminantes Criterio”. Estos contaminantes en el ambiente proceden de numerosas y diversas fuentes de emisiones.

Los estándares nacionales de calidad del aire ambiental (NAAQS) primarios, se establecen para proteger la salud pública. La EPA también establece estándares nacionales de calidad del aire ambiental secundarios, para proteger el bienestar público de los efectos adversos de los contaminantes considerados como criterio, incluida la protección contra el deterioro de la visibilidad o el daño a los animales, los cultivos, la vegetación o los edificios.

Como se requiere según las actuaciones fijadas para conservar el Aire Limpio [52], la EPA realiza periódicamente revisiones exhaustivas de la literatura científica sobre los efectos en la salud y el bienestar, asociados a la exposición a los contaminantes atmosféricos [53,54]. Las evaluaciones resultantes sirven de base para tomar decisiones reglamentarias sobre si conservar o revisar los NAAQS, que especifican las concentraciones permisibles de cada uno de estos contaminantes en el aire ambiente [55].

Las normas primarias se establecen para proteger la salud pública, incluida la salud de las personas que presentan algún riesgo añadido, con un margen de seguridad adecuado. Para fijar los márgenes de seguridad, la EPA considera factores tales como las fortalezas y limitaciones de las evidencias y las incertidumbres relacionadas, la naturaleza y gravedad de los efectos sobre la salud, el tamaño de la población en riesgo y si se han identificado umbrales por debajo de los cuales no se producen efectos apreciables sobre la salud.

En la Ley de Aire Limpio no se pretende que la EPA establezca estándares nacionales de calidad del aire ambiental (NAAQS) primaria, con un nivel de riesgo cero; sino más bien, a un nivel que reduzca el riesgo de manera suficiente para proteger la salud pública, con un margen de seguridad adecuado. Las normas incluyen la consideración de proporcionar protección a una muestra representativa de personas, que comprenden poblaciones en riesgo, y, no a una sola persona; aún, siendo la más susceptible en esos grupos. Incluso en las áreas donde se cumplen las normas actuales, los miembros individuales de las poblaciones en riesgo pueden a veces experimentar efectos sobre la salud relacionados con la contaminación del aire [56,57].

Algunos de los estándares de calidad del aire han sido diseñados para proteger a la población de los efectos adversos para la salud, que pueden haber sucedido después de haber estado expuestos, durante un periodo de tiempo corto, como horas a días. Otras normas, están diseñadas para proteger a las personas de los efectos adversos para su salud, asociados con exposiciones a largo plazo (meses a años). Por ejemplo, el estándar para el ozono se basa en las concentraciones de contaminantes medidas a lo largo de un período de ocho horas. Por el contrario, la norma para el plomo considera las concentraciones medias, medidas durante un período de tres meses. Para las partículas finas ($PM_{2.5}$), se emplean tanto los estándares anuales como los de 24 horas, para proporcionar protección contra los efectos asociados con las exposiciones a largo y corto plazo respectivamente.

Los efectos sobre la salud que se han asociado con cada uno de los contaminantes de los criterios se resumen a continuación. Esta información se extrae principalmente de las evaluaciones de la EPA de la literatura científica para los criterios contaminantes.

3.3.9.1. Efectos sobre la salud de las personas de los contaminantes del aire ‘criterio’, según los informes de la US EPA

Efectos Monóxido de carbono (CO)

Los vehículos a gasolina y otras fuentes estáticas, son los causantes de estas emisiones primarias de monóxido de carbono (CO) en los Estados Unidos. La exposición al monóxido de carbono reduce la capacidad de la sangre para transportar oxígeno, disminuyendo así el suministro de oxígeno a otros tejidos y órganos como el corazón. Las personas que sufren enfermedades cardíacas ya tienen una capacidad reducida para bombear sangre oxigenada al corazón, lo que puede hacer que experimenten isquemia miocárdica (reducción del oxígeno al corazón), a menudo acompañada de dolor torácico. Para estas personas, la exposición a corto plazo al CO afecta aún más la capacidad ya comprometida de su cuerpo para responder a las mayores demandas de oxígeno, durante el ejercicio o esfuerzo.

Por lo tanto, las personas que sufren enfermedades cardíacas se identifican como el grupo de mayor riesgo a las exposiciones al CO ambiental. Otra población potencialmente en riesgo agrupa a las personas con enfermedades pulmonares obstructivas crónicas, las que sufren anemia, diabetes, y los centros de vida prenatal o de ancianos [58].

Efectos del Plomo (Pb)

La principal fuente de emisiones de plomo a la atmósfera ha sido la combustión de gasolina con plomo, en automóviles y camiones. Después de la eliminación de la gasolina con plomo en los Estados Unidos a mediados de los años noventa, en España a partir de agosto de 2001 [59], las fuentes restantes de emisiones de plomo en el aire incluyen fuentes industriales, fundición de plomo y reciclaje de baterías.

La exposición al plomo, que se acumula en los huesos, la sangre y los tejidos blandos del cuerpo, puede afectar el desarrollo del sistema nervioso central en los niños pequeños. Esto podría derivar en efectos neurológicos, tales como la disminución del coeficiente intelectual y problemas derivados del comportamiento, según se recoge en los informes de la EPA [60].

Efectos Ozono troposférico (O_3)

La exposición a corto plazo al ozono troposférico puede causar diversos efectos en la salud, como inflamación del revestimiento de los pulmones, reducción de la función pulmonar y síntomas respiratorios como tos, sibilancias, dolor torácico, ardor en el pecho y dificultades respiratorias [61, 62]. Puede disminuir la capacidad de realizar ejercicio, aumentar la susceptibilidad a la infección respiratoria. La exposición a las concentraciones ambientales de ozono se ha asociado con el agravamiento de enfermedades respiratorias como el asma, el enfisema y la bronquitis; lo que conlleva un mayor uso de medicamentos, ausencias de la escuela e incremento de las admisiones hospitalarias. La exposición a ozono a corto plazo se asocia con la mortalidad prematura. Los estudios también han descubierto que, la exposición a largo plazo al ozono puede contribuir al desarrollo del asma, especialmente entre los niños con ciertas susceptibilidades genéticas y los niños que frecuentemente se ejercitan al aire libre, pudiendo dañar permanentemente el tejido pulmonar.

Efectos Material Particulado (PM)

Aunque la evidencia científica relaciona efectos nocivos para la salud humana con exposiciones tanto a partículas finas como a partículas gruesas torácicas, es mucho más peligrosa para partículas finas que para partículas gruesas. Los efectos asociados con las exposiciones a $PM_{2.5}$ y $PM_{10-2.5}$ incluyen mortalidad prematura, agravación de la enfermedad respiratoria y cardiovascular y cambios en los indicadores subclínicos de la función respiratoria y cardíaca. Tales efectos en la salud se han asociado con la exposición a corto y/o largo plazo a PM. Las exposiciones a $PM_{2.5}$ también están asociadas con la disminución de la función pulmonar, el incremento de síntomas alérgicos y problemas respiratorios.

Los niños, los adultos mayores, las personas con cardiopatía preexistente y asma son considerados como uno de los grupos más expuestos a los efectos asociados con las exposiciones a PM. Los informes que actualmente se manejan, proporcionan evidencias de asociaciones entre la exposición a largo plazo a PM_{2,5} y los efectos en el desarrollo, como el bajo peso al nacer y la mortalidad infantil por causas respiratorias [63].

Efectos Dióxido de nitrógeno (NO₂)

El óxido nítrico (NO) y el dióxido de nitrógeno (NO₂) son emitidos por automóviles, camiones, autobuses, centrales eléctricas y motores de equipamientos no de carretera. La exposición al dióxido de nitrógeno se ha asociado con una variedad de efectos sobre la salud, incluidos los síntomas respiratorios, especialmente entre los niños asmáticos, y aumento de las consultas por urgencias respiratorias y las admisiones hospitalarias, en particular para los niños y los ancianos [64].

Efectos Dióxido de azufre (SO₂)

La quema de combustibles fósiles en las empresas de producción eléctrica y las industrias, son la principal fuente de dióxido de azufre. Las personas con asma son un grupo especialmente susceptible a los efectos del SO₂. Individuos asmáticos expuestos en periodos cortos a niveles elevados de SO₂, durante el ejercicio a un nivel moderado, ocasiona dificultades respiratorias.

Estas pueden estar acompañadas de sibilancias, opresión en el pecho o falta de aliento. Los estudios proporcionan pruebas consistentes de la relación entre la exposición a dióxido de azufre, a corto plazo, y el aumento de anomalías respiratorias en los niños, especialmente aquellos con asma o problemas respiratorios crónicos. Dichas exposiciones a corto plazo al dióxido de azufre se han asociado con el aumento de las consultas por urgencias respiratorias y las admisiones hospitalarias, en particular para los niños y los ancianos [65].

El período de desarrollo fetal, donde ha habido exposición materna a algunos de los denominados contaminantes ‘criterio’ del aire, es de particular vulnerabilidad por los posibles efectos adversos en la salud resultantes. Por ejemplo, se ha demostrado que, tanto el plomo como las PM, atraviesan la placenta y se acumulan en el tejido fetal durante la gestación (informes finales US EPA, para el plomo y PM, ya referenciados).

Además de los hallazgos señalados anteriormente con respecto a las asociaciones de exposición prenatal al PM y los resultados adversos al nacimiento (como bajo peso al nacer), estudios sobre la exposición prenatal a contaminantes atmosféricos criterio, han informado de que la exposición a PM, óxidos de nitrógeno y azufre, puede aumentar el riesgo de desarrollar asma y empeorar los resultados respiratorios entre los niños que desarrollan asma [66].

En un informe del Health Effects Institute [67], se concluye diciendo que vivir cerca de las carreteras ocupadas parece ser un factor de riesgo independiente para el inicio del asma infantil. El mismo informe también concluyó que las pruebas eran “suficientes”, para inferir una relación causal entre la exposición al tráfico y la virulencia del asma en niños. Otros estudios, han sugerido que los contaminantes relacionados con el tráfico pueden contribuir al desarrollo de una enfermedad alérgica; ya sea afectando la respuesta inmunitaria directamente, o, aumentando la concentración y la actividad biológica de los mismos alérgenos [68].

3.3.9.2. Índice de Calidad del Aire (AQI)

El Índice de Calidad del Aire (AQI, por sus siglas en inglés) es una herramienta usada por la EPA y otras agencias, para informar al público de una forma fácil de comprender sobre la calidad del aire local; indicando, si los niveles de polución son perjudiciales a la salud. El AQI es, por tanto, un índice para la notificación de la calidad del aire, a diario, del grado de pureza o contaminación atmosférica y de los efectos para la salud.

Informa a la ciudadanía si la condición del aire debe preocuparles por su salud, en base a los efectos que sobre ella podría tener haber estado expuestos a contaminantes, dentro unas horas o días después de respirar el aire contaminado.

La EPA calcula el AQI para cinco contaminantes atmosféricos principales reglamentados por la Ley del aire puro: ozono a nivel del suelo, contaminación por partículas, monóxido de carbono, dióxido de azufre y dióxido de nitrógeno.

La EPA estableció normas nacionales sobre la calidad del aire para la protección de la salud humana conforme a cada uno de estos contaminantes. El ozono al nivel del suelo y las partículas en la atmósfera son los dos contaminantes que presentan la mayor amenaza a la salud de las personas. El AQI es una regla que va de 0 a 500, cuanto más alto es el valor del AQI, mayor es el nivel de contaminación atmosférica y mayor la preocupación para la salud.

3.3.9.3. Interpretación de los colores del AQI

El objeto del AQI es ayudar a comprender el significado de la calidad del aire local para la salud. A fin de facilitar su interpretación, el AQI se divide en seis categorías.

Niveles del índice de la calidad del aire preocupantes para la salud	Valor numérico	Significado
Bueno	0 a 50	La calidad del aire se considera satisfactoria y la contaminación atmosférica presenta un riesgo escaso o nulo.
Moderado	51 a 100	La calidad del aire es aceptable pero para algunos contaminantes podría existir una preocupación moderada para la salud de un grupo muy pequeño de personas excepcionalmente sensibles a la contaminación ambiental.
Insalubre para grupos sensibles	101 a 150	Los miembros de grupos sensibles pueden padecer efectos en la salud. Probablemente no afectará a las personas en general.
Insalubres	151 a 200	Todos pueden comenzar a padecer efectos en la salud y los miembros de grupos sensibles pueden padecer efectos más graves.
Muy insalubre	201 a 300	Advertencias sanitarias de condiciones de emergencia. Son mayores las probabilidades de que toda la población esté afectada.
Peligroso	301 a 500	Alerta sanitaria: todos pueden padecer efectos sanitarios más graves.

Tabla 3.6. Niveles del Índice de Calidad de Aire (AQI) y sus consecuencias para la salud de las personas. Fuente: AirNow. https://www.airnow.gov/index.cfm?action=aqibasics.aqi_sp.

La EPA ha asignado un color específico a cada categoría del AQI para ayudar a las personas a interpretar rápidamente si la contaminación del aire está llegando a niveles insalubres en sus comunidades. Por ejemplo, el color naranja significa que las condiciones son “insalubres para grupos sensibles” y el rojo significa que las condiciones pueden ser “insalubres para todos”.

3.3.9.4. Sistema de Calidad del Aire (AQS)

Los datos presentados aquí se refieren a la concentración ambiental de los contaminantes del aire de criterio y *HAP* (*contaminantes peligrosos del aire, en inglés: Hazardous Air Pollutants*), en la zona fronteriza.

Se proporcionan resúmenes anuales sobre la concentración de contaminantes del aire de criterio, extraída mensualmente, y sobre la concentración de *HAP*, extraída trimestralmente de la base de datos de *EPA's Air Quality System (AQS)*, sistema de calidad del aire de la *EPA*. El resumen de datos incluye monitores ubicados dentro de aproximadamente 100 kilómetros (62 millas) de la frontera entre EE.UU. y México que reportaron datos en los últimos diez años.

Las agencias estatales de los EE.UU. han revisado los datos para garantizar su calidad. El AQS también contiene datos meteorológicos, información descriptiva sobre cada estación de monitoreo (incluyendo su localización geográfica y su operador), y la información de control/garantía de calidad de datos. La Oficina de Planificación y Calidad del Aire de Normas (OAQPS), y otros usuarios AQS, se basan en los datos del sistema para evaluar la calidad del aire, para evaluar los planes de aplicación del Estado de áreas de no cumplimiento, para modelización del análisis de revisión y otras funciones de gestión de la calidad del aire. La información procedente del AQS, también se utiliza para elaborar los informes del Congreso, tal y como lo establece la Ley de Aire Limpio. En el siguiente enlace, se pueden revisar todas las presentaciones AQS de la Conferencia Nacional de Monitoreo del Aire Ambiental 2016: <https://www.epa.gov/aqs/naam-conference-2016>.

Los reportes generados en el sitio Web que se pueden consultar en el enlace: https://www3.epa.gov/ttnccatc1/cica/airq_s.html#tabla (con acceso en enero 2018), suministran concentraciones de contaminantes pertinentes a cada norma de calidad del aire. Por ejemplo, un reporte de la *Total Suspended Particulate Matter (TSP)*, materia particulada suspendida total, incluye una concentración promedio durante 24 horas y una concentración promedio anual. Estos valores corresponden a las normas de calidad del aire de corto y largo plazo de México para TSP.

CONTAMINANTE	NORMA	
	Estados Unidos	México
Monóxido de carbono (CO)		
Promedio para 8 horas	9 ppm	11 ppm
Promedio para 1 hora	35 ppm	
Dióxido de nitrógeno (NO₂)		
Promedio anual	0.053 ppm	
Promedio para 1 hora		0.21 ppm
Ozono (O₃)		
Promedio para 8 horas	0.08 ppm	
Promedio para 1 hora	0.12 ppm	0.11 ppm
Dióxido de azufre (SO₂)		
Promedio anual	0.030 ppm	0.030 ppm
Promedio para 24 horas	0.14 ppm	0.13 ppm
Materia particulada de menos de 2.5 micrómetros (PM_{2.5})		
Promedio anual	15 µg/m ³	
Promedio para 24 horas	65 µg/m ³	
Materia particulada de menos de 10 micrómetros (PM₁₀)		
Promedio anual	50 µg/m ³	50 µg/m ³
Promedio para 24 horas	150 µg/m ³	150 µg/m ³
Materia particulada suspendida total (TSP)		
Promedio anual		75 µg/m ³
Promedio para 24 horas		260 µg/m ³
Plomo (Pb)		
Promedio trimestral	1.5 µg/m ³	1.5 µg/m ³

Tabla 3.7. Normas de aire ambiental basadas en la salud (concentración de los contaminantes según las normas de EE.UU.). Fuente: US EPA, https://www3.epa.gov/ttnccatc1/cica/airq_s.html#tabla Centro de información sobre contaminación de aire (CICA) para la frontera entre EE.UU-México

La Tabla 3.7 anterior enumera las normas de calidad del aire para EE.UU. y México. La tabla anterior incluye el período de tiempo a lo largo del cual se promedian las concentraciones del contaminante (es decir, el tiempo de exposición) y el valor numérico de cada norma. Los valores se proporcionan en partes por millón por volumen (ppm) y en microgramos por metro cúbico de aire ($\mu\text{g}/\text{m}^3$).

Capítulo IV

MATERIALES Y MÉTODOS

Modelos de
clasificación,
regresión y
optimización

4. Modelos de clasificación, regresión y optimización

4.1. Máquinas de vectores de Soporte (SVM)

Las máquinas de soporte vectorial o máquinas de vectores soporte (en inglés, Support Vector Machines - SVMs), son algoritmos supervisados de aprendizaje automático, para tareas de regresión y clasificación; aunque son frecuentemente empleados para la clasificación. Los modelos SVM tienen dos características particularmente deseables: robustez en presencia de datos ruidosos y aplicabilidad a una variedad de configuraciones de datos. En su núcleo, un modelo SVM es un hiperplano que separa dos clases distintas de datos (en el caso de problemas de clasificación). Para los casos que son linealmente separables, la distancia entre el hiperplano y los puntos de datos de entrenamiento más cercanos se maximiza. Los vectores que se encuentran en el margen, según podemos apreciar en la Figura 4.1, se denominan vectores soporte.

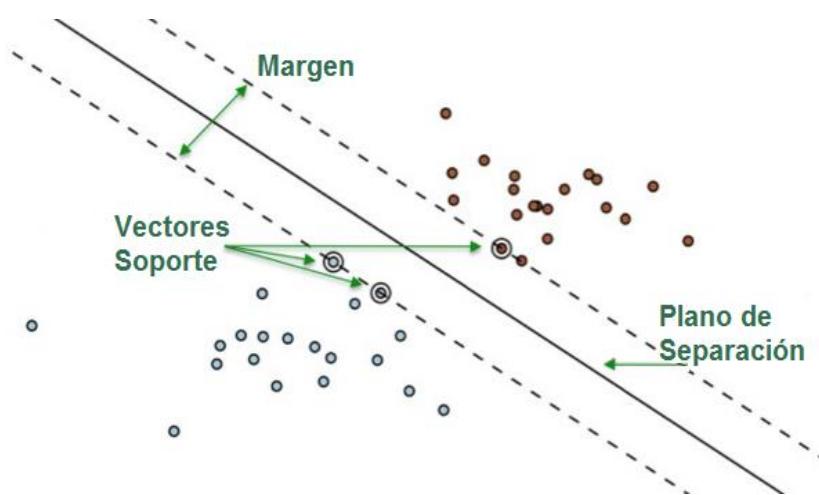


Figura 4. 1. Plano de separación en un caso linealmente separable. Fuente: Adaptada de Support Vector Machines in Apache MADlib, Xiaocheng Tang. <https://www.cloudfoundry.org/>

Una Máquina de Soporte Vectorial (SVM) “aprende” la superficie decisión de dos clases distintas de los puntos de entrada. Como clasificador de una sola clase, la descripción dada por los datos de los vectores de soporte es capaz de formar una frontera de decisión, alrededor del dominio de los datos de aprendizaje con muy poco o ningún conocimiento de los datos fuera de esta frontera.

La herramienta SVM para regresión fue desarrollada en los laboratorios de AT&T por Vaptnik [69]. Las SVM son muy utilizadas para la identificación de sistemas no lineales, esto debido principalmente dichas Máquinas de Soporte Vectorial se basan en el principio de minimización del riesgo estructural (SRM por sus siglas en inglés, "Structural Risk Minimization"), principio originado de la teoría de aprendizaje estadístico desarrollada por Vapnik, el cual ha demostrado ser superior al principio de minimización del riesgo empírico (ERM por sus siglas en inglés "Empirical Risk Minimization"), que emplean las redes neuronales convencionales. Las ventajas de la herramienta SVM son las siguientes:

- Por su capacidad de minimizar el riesgo estructurado, posee una excelente capacidad de generalización.
- Al depender solamente de los datos que contienen la mayor información, se necesitan ajustar pocos parámetros.
- La estimación de los parámetros se realiza a través de la optimización de una función de costo convexa, lo cual evita la existencia de un mínimo local.
- El modelo final puede ser escrito como una combinación de un número muy pequeño de vectores concretos, llamados vectores de soporte.

Para grandes cantidades de datos, las técnicas numéricas de optimización existentes, para resolver un problema cuadrático, no son admisibles en términos computacionales. Las SVM, resuelven un problema cuadrático allí donde el número de coeficientes es igual al número de entradas o datos de entrenamiento.

Este problema impide el uso de las SVM, para la identificación de sistemas no lineales en línea, en casos en los que las entradas son obtenidas de manera secuencial y el aprendizaje se realiza en cada paso. En la literatura existen algunas técnicas para la aplicación de SVM en línea. Estos intentos posibilitaron el diseño de algoritmos que mejoraron tanto el tiempo como el costo computacional de los algoritmos convencionales existentes, para resolver el problema cuadrático.

4.1.1. Funcionamiento de las SVMs

La teoría de las máquinas de vectores soporte (SVM) fue desarrollada por Vapnik [70]. Este método se aplica en muchas aplicaciones de aprendizaje automático, tales como clasificación de objetos, predicción de series de tiempo, análisis de regresión y reconocimiento de patrones.

En el análisis usando SVM, la idea principal es mapear los datos originales x en un espacio de características F , con mayor dimensionalidad a través de la función de correlación no lineal, que generalmente es desconocida, y luego llevar a cabo una regresión lineal. Una SVM primero mapea los puntos de entrada a un espacio de características de una dimensión mayor (i.e.: si los puntos de entrada están en \mathbb{R}^2 , entonces son mapeados por la SVM a \mathbb{R}^3) y encuentra un hiperplano que los separe y maximice el margen m entre las clases en este espacio, como se aprecia en la Figura 4.2.

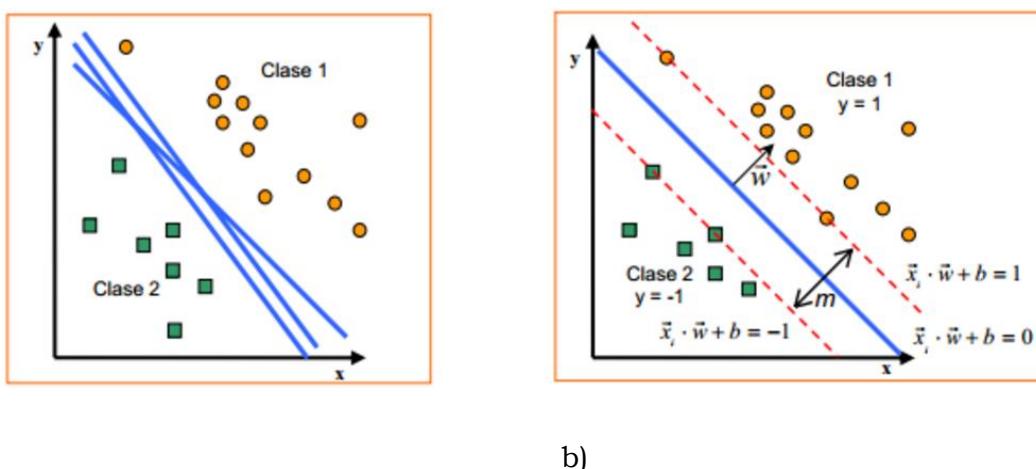


Figura 4. 2. La imagen a) muestra los posibles hiperplanos de separación. En la imagen b) se muestra el hiperplano de separación óptimo. Fuente:
<http://samuelabad1991.blogspot.com.es/2014/02/analisis-con-maquinas-de-vectores.html>

La SVM encuentra el hiperplano óptimo, utilizando el producto punto con funciones en el espacio de características que son llamadas kernels. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada, denominados vectores de soporte.

En la teoría básica de las SVM, se podrían plantear una serie de problemas de clasificación, de los que vamos a describir dos.

4.1.1.1. Conjunto de puntos linealmente separable

Supongamos que nos han dado un conjunto S de puntos, etiquetados para su entrenamiento como se aprecia en la siguiente Figura, 4.3.

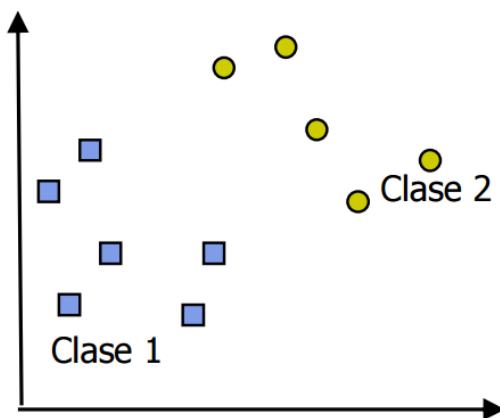


Figura 4.3. Conjunto de puntos linealmente separable. Fuente: Máquinas de soporte vectorial. Scientia et Technica Año XI, No 27, abril 2005. UTP. ISSN 0122-1701.

Cada punto de entrenamiento $x_i \in \Re^N$, pertenece a una de las dos clases y se le da una etiqueta $y \in \{-1, 1\}$ para $i = 1, \dots, l$. En la mayoría de los casos, la búsqueda de un hiperplano adecuado en un espacio de entrada es demasiado restrictivo para ser de uso práctico. Una solución a esta situación es mapear el espacio de entrada en un espacio de características de una dimensión mayor y buscar allí el hiperplano óptimo.

Sea entonces $z = \varphi(x)$, la notación del vector en el espacio de características con un mapeo φ de \Re^N a un espacio de características Z . En estas condiciones, se desea encontrar el hiperplano: $w \cdot z + b = 0$

Dicho hiperplano, viene definido por el par (w, b) , de tal forma que se podrá separar el punto x_i de acuerdo a la función:

$$f(x_i) = \text{sign}(w \cdot z_i + b) = \begin{cases} 1 & y_i = 1 \\ -1 & y_i = -1 \end{cases} \quad (4.1)$$

En donde: $w \in Z$ y $b \in \Re$. De una forma más precisa, se dice que el conjunto S es linealmente separable, si existe (w, b) , de forma que las inecuaciones siguientes sean válidas para todos los elementos del conjunto S .

$$\begin{cases} (w \cdot z_i + b) \geq 1, & y_i = 1 \\ (w \cdot z_i + b) \leq -1, & y_i = -1 \end{cases} \quad i = 1, \dots, l \quad (4.2)$$

Para el caso linealmente separable de S , podemos encontrar un único hiperplano óptimo, para el cual, el margen entre las proyecciones de los puntos de entrenamiento de dos clases distintas es maximizado.

4.1.1.2. Conjunto de puntos linealmente no separable

El problema anterior no tiene demasiado interés en la práctica, dado que en los problemas reales existe abundante ruido y no es posible establecer una separación lineal perfecta. En este tipo de casos la solución pasa por relajar el grado de separabilidad, permitiendo que alguno de los conjuntos de entrenamiento presente errores. En el caso de ejemplos no-separables se pueden dar dos casos, en el primero el punto cae dentro del margen que se asocia a la clase correcta, de acuerdo a la frontera de decisión que define el hiperplano de separación; en el otro caso, el punto cae al otro lado del hiperplano. En ambos casos se concluye que el ejemplo es no-separable, pero mientras en el primer caso la clasificación es correcta, en el segundo no lo es.

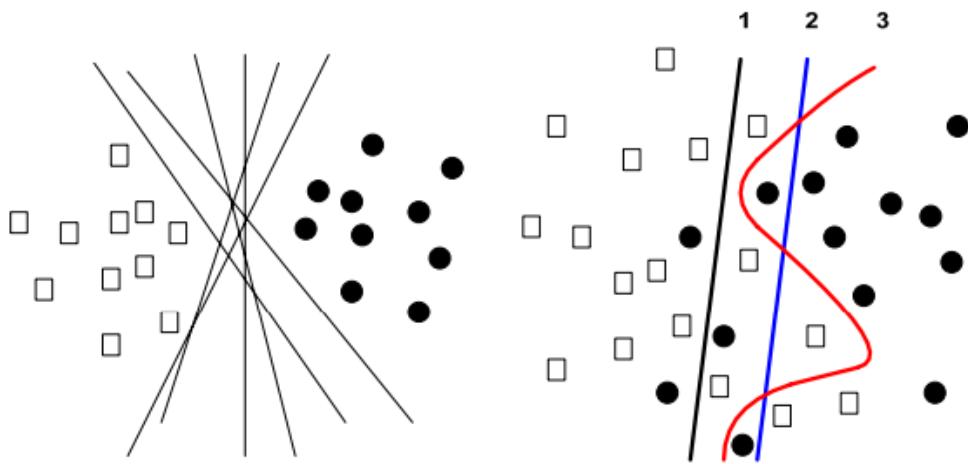


Figura 4.4. Conjunto de función de clasificación en el caso de datos linealmente separables (izquierda) y caso linealmente no separable (derecha). Fuente: Support Vector Machines with Evolutionary Feature Selection for Default Prediction, Wolfgang Karl Härdle et al. SFB 649 Discussion Paper 2012-030.

Para abordar este nuevo problema es la introducción de una serie de variables reales positivas, denominadas variables de holgura, ξ_i , $i=1, \dots, l$, que permitirán cuantificar el número de casos no-separables que estaríamos dispuestos a admitir, así:

$$y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l \quad (4.3)$$

Por tanto, dado un ejemplo (x_i, y_i) , su variable de holgura, ξ_i , representa la desviación del caso separable, medida desde el borde del margen que corresponde a la clase y_i . Según esta definición, las variables de holgura de valor '0' corresponden a ejemplos separables, las mayores de '0' corresponden a los no separables y las que son mayores de 1 corresponden a ejemplos no separables y además más clasificados. Así, la suma de todas las variables de holgura $\sum_{i=1}^l \xi_i$, permite medir el coste asociado al número de ejemplos no separables. Cuanto mayor es el valor de la suma de las variables de holgura, mayor será el número de casos no separables.

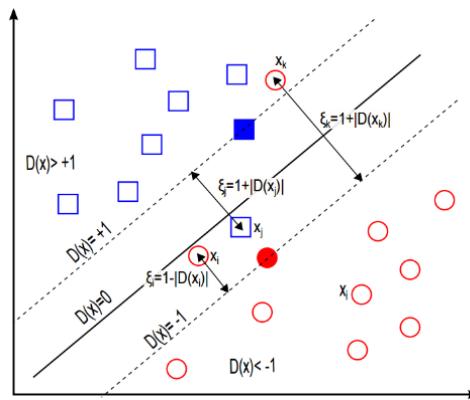


Figura 4.5. Conjunto de puntos linealmente separable. Fuente: Máquinas de soporte vectorial (SVM). Enrique J. Carmona Suárez. Dpto. de Inteligencia Artificial, ETS de Ingeniería Informática (UNED).

En la Figura 4.5, se puede ver que la distancia de cualquier punto, x_i , al hiperplano de separación óptimo viene dada por $|D(x_i)|/\|w\|$. De forma particular, si el ejemplo pertenece al conjunto de vectores soporte que están identificados por siluetas sólidas, la distancia a dicho hiperplano será siempre $1/\|w\|$. En los vectores soporte aplicados a la función de decisión, siempre se cumple que $|D(x)| = 1$.

Si disponemos de un conjunto S de puntos que no son linealmente separables, puede generalizarse el análisis previo introduciendo variables no negativas $\xi_i \neq 0$, de tal forma que la ecuación del caso anterior se modifica:

$$y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l. \quad (4.4)$$

Los $\xi_i \neq 0$ en la ecuación 4.4, son aquellos para los cuales el punto x_i no satisface la ecuación. En este sentido el término $\sum_{i=1}^l \xi_i$, puede ser tomado como una medida del error en la clasificación.

Así, el problema del hiperplano óptimo, se redefine como la solución del problema.

$$\min \left\{ \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i \right\} \quad (4.5)$$

$$s. a \quad y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \quad (4.6)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (4.7)$$

Es decir, puesto que no se imponen demasiadas restricciones, no bastaría con un único objetivo de maximizar el margen, ya que, esto se podría lograr a costa de clasificar muchos de los casos erróneamente. Por esta razón, la función a optimizar deberá incluir los errores de separación que está cometiendo el hiperplano de separación, según:

$$f(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (4.8)$$

Donde C , es una constante que puede ser definida como un parámetro de regularización. Este parámetro es el único que puede ajustarse en la formulación de la SVM. La constante C , elegida por el usuario y suficientemente grande, permite controlar en qué grado influye el término del coste de ejemplos no-separables en la minimización de la norma, es decir, permitirá regular el compromiso entre el grado de sobreajuste del clasificador final y la proporción del número de ejemplos no separables.

Por tanto, si el valor de C es muy grande permitiría valores de ξ_i muy pequeños. En el caso límite ($C \rightarrow \infty$), estamos considerando el caso de ejemplos perfectamente separables ($\xi_i \rightarrow 0$).

En caso contrario, si el valor de C es muy pequeño, permitiría valores de ξ_i muy grandes, es decir, estaríamos admitiendo un número de ejemplos mal clasificados muy elevado. En el caso límite ($C \rightarrow 0$), se estaría aprobando que todos los ejemplos estuvieran mal clasificados ($\xi_i \rightarrow \infty$).

4.1.2. Caso no lineal

Es muy común en los problemas reales, la necesidad de clasificar conjuntos donde los ejemplos no sean separables. Aunque podemos utilizar el enfoque aplicado a conjuntos de puntos linealmente no separables, es preferible emplear estrategias no lineales.

Para resolver la clasificación de este tipo de conjuntos mediante SVM, se introduce nuestro conjunto no lineal, en un espacio de dimensión mayor donde sí sean separables, y, en este nuevo espacio, se intentan encontrar el hiperplano de separación más adecuado. A este mencionado espacio, lo podríamos denominar como espacio de características.

Si $\Phi : \mathcal{X} \rightarrow \mathcal{F}$ es la función que hace corresponder a cada punto de entrada x un punto en el espacio de características \mathcal{F} , se desea encontrar el hiperplano de separación en dicho espacio \mathcal{F} . Este hiperplano, inmerso en el espacio de características, se transforma en una función no lineal que separa nuestro conjunto, en el espacio original de entradas. De esta forma, la función de decisión en el espacio de características vendría dada por: $D(x) = \langle w, \Phi(x) \rangle + b$.

El problema asociado a este caso, es similar al descrito previamente para el caso de los puntos linealmente no separables, con la excepción de que el hiperplano de separación se determina en el propio espacio de características, es decir:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\lambda} \xi_i \quad (4.9)$$

$$y_i \langle w, \Phi(x_i) \rangle + \xi - 1 \geq 0 \quad i = 1, \dots, \lambda \quad (4.10)$$

$$\xi_i \geq 0 \quad i = 1, \dots, \lambda \quad (4.11)$$

Seguidamente se presenta, como ejemplo, una posible función Φ .

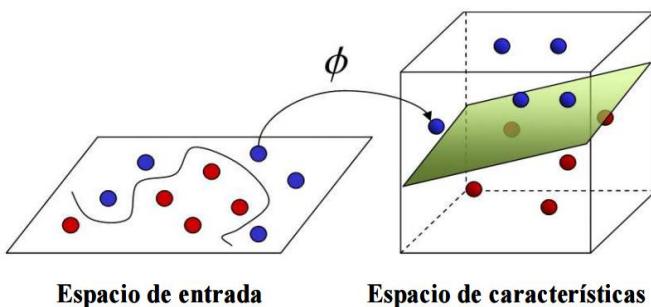


Figura 4.6. Transformación de los datos de entrada a un espacio de mayor dimensión.
Fuente: Esther Gutiérrez Alonso, Aplicación de las máquinas de soporte vectorial (ICAI) Madrid 2007.

4.1.3. La SVM de margen suave

En aquellas clases que no sean totalmente separables el algoritmo de aprendizaje no convergerá y fluctuará cada vez que encuentre un punto que no se pueda clasificar, cambiando de hipótesis (w). Para abordar este problema en Cortes y Vapnik (1995) [71], se define un margen de holgura variable $\xi > 0$ para los ejemplos (x_i, y_i) que no se puedan separar con el hiperplano (w, b) como se observa en la Figura 4.7.

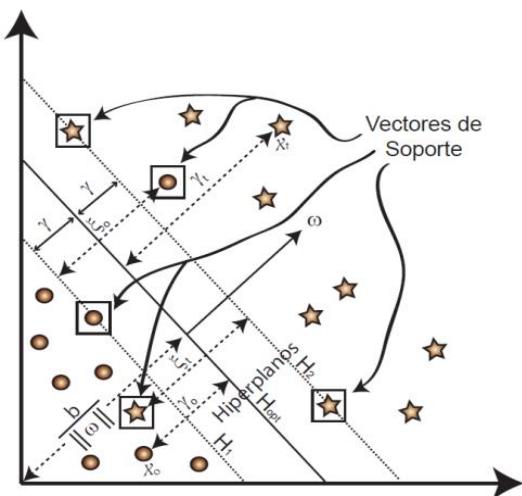


Figura 4.7. Variables de Holgura ξ .

Fuente: Cristianini y Shawe-Taylor (2004).

$$\xi_i = \max(0, y_i - (\omega \cdot x_i) - b) \quad (4.12)$$

Estas variables de holgura ξ permiten reescribir las condiciones de clasificación y encontrar un hiperplano óptimo con un margen de separación suave (soft margin), de la siguiente forma.

$$\langle w \cdot y_i \rangle + b \geq +1 - \xi_i, \text{ para } y_i = +1 \quad (4.13)$$

$$\langle w \cdot y_i \rangle + b \leq -1 + \xi_i, \text{ para } y_i = -1 \quad (4.14)$$

$$\xi_i \geq 0 \quad \forall i \quad (4.15)$$

Esto implica cambiar la función de minimización $\|w\|^2 = \mathcal{Z}$, propuesta anteriormente, $\|w\|^2 / 2 + C(\sum_i \xi_i)^k$, donde $k > 0$ y C es un factor de penalización del error asignado por el usuario, también mayor que 0. Independiente del valor de k , este sigue siendo un problema de programación cuadrática. Las dos ecuaciones anteriores se combinan en la siguiente desigualdad.

$$y_i(w \cdot x_i) + b \geq 1 - \xi_i \quad \forall i \quad (4. 16)$$

Aplicando nuevamente el método de multiplicadores de Lagrange, Para $k = 1$, se puede formular como:

$$l(w, b, \xi, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i (y_i(w \cdot x_i) + b) - 1 - \xi_i - \sum_{i=1}^l \mu_i \xi_i \quad (4. 17)$$

donde α_i y μ_i , son los multiplicadores de Lagrange.

Se puede hallar una representación dual de la ecuación anterior, usando las condiciones KKT [72], que se aplican encontrando el gradiente con respecto a w, b, ξ , e igualando a 0:

$$\frac{\partial l(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \quad (4. 18)$$

$$\frac{\partial l(w, b, \alpha)}{\partial b} = -\sum_{i=1}^l \alpha_i y_i = 0 \quad (4. 19)$$

$$\frac{\partial l(w, b, \alpha)}{\partial \xi} = C - \alpha_i - \mu_i = 0 \quad (4. 20)$$

En la última ecuación, si $\mu_i \geq 0 \Rightarrow C \geq \alpha_i \geq 0$. A esta ecuación se la conoce como “box constraints”. Si reemplazamos en la ecuación original, se obtiene la siguiente ecuación:

$$l(w, b, \xi, \alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{j=1}^l \sum_{i=1}^l \alpha_i \alpha_j y_i y_j (x_i x_j) \quad (4. 21)$$

Con la restricción $\sum_{i=0}^l \alpha_i y_i = 0$. (4. 22)

Se propone una técnica alternativa [Chen et. al, 2002], denominada v-SVM la cual maximiza el segundo término de la ecuación anterior.

$$l(w, b, \xi, \alpha) = -\frac{1}{2} \sum_{j=1}^l \sum_{i=1}^l \alpha_i \alpha_j y_i y_j (x_i x_j) \quad (4. 23)$$

Con restricciones de $\sum_{i=0}^l \alpha_i y_i = 0$, $\mu_i \geq v$, y $\frac{1}{l} \geq \alpha_i \geq 0$.

En este método el parámetro v , que reemplaza al parámetro C , es una cota superior para el margen de error ξ , y al mismo tiempo una cota inferior para los vectores soporte.

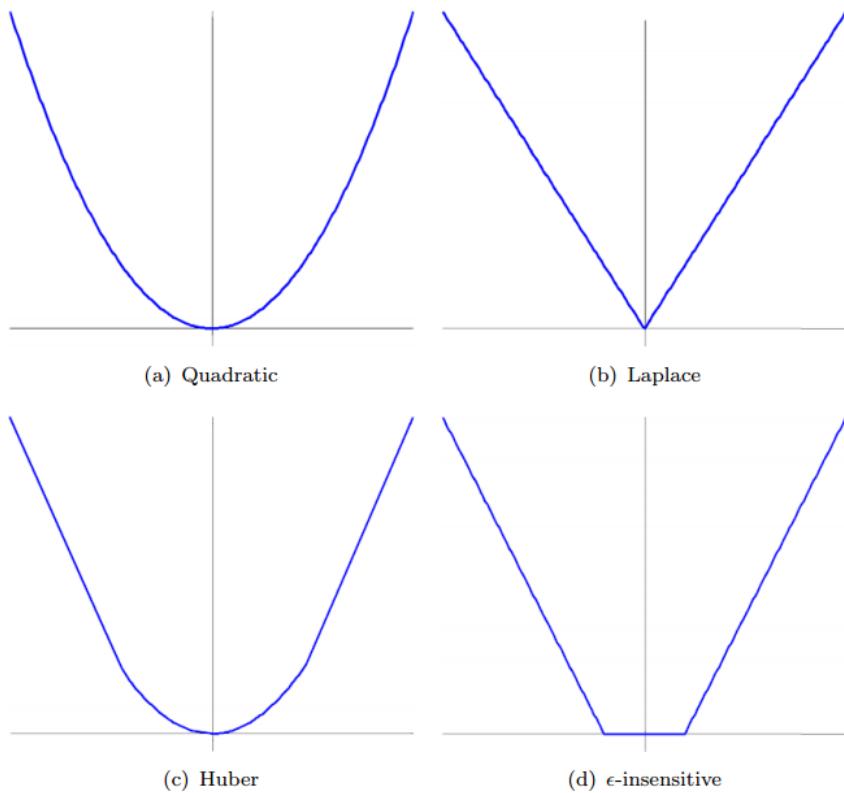
4.1.4. Máquinas de Soporte Vectorial para Regresión

La SVM es un método de aprendizaje con una raíz teórica en la teoría del aprendizaje estadístico [73]. La SVM fue originalmente desarrollada para la clasificación, y más tarde se generalizó para resolver los problemas de regresión, empleando el ‘vector de apoyo de regresión’ (SVR) [74].

El modelo producido por la clasificación de vectores soporte sólo depende de un subconjunto de los datos de entrenamiento, porque la función de costo para construir el modelo no se preocupa por los puntos de entrenamiento que están más allá del margen. Análogamente, el modelo producido por SVR sólo depende de un subconjunto de los datos de entrenamiento, ya que la función de coste para construir el modelo ignora los datos de entrenamiento que están cerca (dentro de un umbral ϵ) de la predicción del modelo.

La técnica de los vectores soporte es una herramienta universal para resolver problemas de estimación de funciones multidimensionales. En el caso de la regresión, seleccionaremos el hiperplano regresor que mejor se ajuste a nuestro conjunto de datos de entrenamiento, dado que no disponemos de clases para separar.

Figura 4.8. Funciones de pérdida. Fuente:
Support Vector
Machines for
Classification and
Regression 1998, Steve
R. Gunn.



La idea se basa en considerar una distancia margen ε , de modo que todos los ejemplos se encuentren en una *banda* o *tubo* entorno a nuestro hiperplano; es decir, que disten una cantidad menor de ε del hiperplano. A la hora de definir el hiperplano sólo se considerarán los ejemplos que disten más de ε de nuestro hiperplano, que serán considerados como vectores soporte.

Las SVM también pueden aplicarse a problemas de regresión mediante la introducción de una función de pérdida alternativa [Smola, 1996]. La función de pérdida debe modificarse para incluir una medida de distancia. La Figura 4.8 ilustra cuatro posibles funciones de pérdida.

La función de pérdida en la Figura 4.8 (a) corresponde al criterio de error de mínimos cuadrados convencional. La función de pérdida en la Figura 4.8 (b) es una función de pérdida Laplaciana, que es menos sensible a los valores atípicos que la función de pérdida cuadrática. Huber propuso la función de pérdida en la Figura 4.8 (c), como una función de pérdida robusta que tiene propiedades óptimas cuando la distribución subyacente de los datos es desconocida. Estas tres funciones de pérdida no producirán ninguna dispersión en los vectores soporte. Para abordar esta cuestión, Vapnik propuso la función de pérdida en la Figura 4.8 (d) como una aproximación a la función de pérdida de Huber que permite obtener un escaso conjunto de vectores de soporte.

En la clasificación mediante SVM, se determina la separación entre las dos clases de puntos en función del margen, a mayor margen mayor seguridad de que el hiperplano de separación es bueno. En la regresión no estamos interesados en la separación de puntos por un hiperplano, pero se desea que la función establecida esté lo más próxima posible a los puntos. Es decir, lo que para clasificación sería el margen, en la regresión sería la formación de una *banda* o *tubo*, alrededor de la función de regresión establecida.

4.1.4.1. Regresión lineal

Según Schölkopf y Smola (2004) [75] el objetivo de la regresión es encontrar una función $f(w, x_i) = \langle w \cdot x_i \rangle + b$, tal que tenga como máximo una desviación ϵ de la salida y_i para todos los datos de entrenamiento, y al mismo tiempo que este ϵ sea mínimo.

Si se considera el problema de aproximación de un conjunto de datos,

$$D = \{(x_1, y_1), \dots, (x_\ell, y_\ell)\}, \quad x \in \mathbb{R}^\ell, y \in \mathbb{R} \quad (4.24)$$

Con una función lineal,

$$f(x) = \langle w, x \rangle + b \quad (4.25)$$

La función óptima de regresión viene dada por el mínimo de la funcional,

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum (\xi_i^- + \xi_i^+), \quad (4.26)$$

Donde C es un valor pre-definido, y ξ_i^-, ξ_i^+ , representadas en la Figura 4.9 por (ξ^*, ξ) respectivamente, son variables de holgura que representan limitaciones superior e inferior en las salidas del sistema.

La SVM de Vapnik para regresión, usa una función de insensibilidad ϵ que es el valor absoluto del error $|y - f(x, w)|$, de tal forma que se genera un tubo de radio ϵ de insensibilidad como se muestra en la Figura 4.9. Este tubo de regresión está limitado por los hiperplanos $f(w, x_i) = \langle w \cdot x_i \rangle + b + \epsilon$ y $f(w, x_i) = \langle w \cdot x_i \rangle + b - \epsilon$, donde las muestras que caen dentro de éste tienen un residuo (valor absoluto para la diferencia entre y calculada y experimental) menor que ϵ considerándose el error de estas muestras como 0 y por lo tanto no se penalizan. sólo las muestras que están fuera del tubo de regresión tienen un residuo más grande que ϵ , y por tanto tiene una penalización distinta de 0 que, para la función de pérdida de insensibilidad ϵ , es proporcional a su distancia desde el borde del tubo de insensibilidad [Ivanciu, 2007] [76].

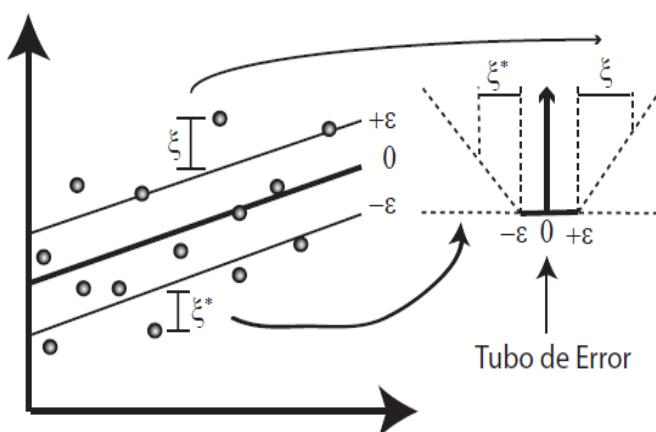


Figura 4.9. Regresión lineal con error máximo ϵ . Fuente: Schölkopf y Smola (1998).

4.1.4.1.1. ϵ -Función de pérdida insensible

Empleando una ϵ -función de pérdida insensible, de la Figura 4.8 (d),

$$L_\epsilon(y) = \begin{cases} 0 & \text{para } |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon & \text{en otro caso} \end{cases} \quad (4.27)$$

La solución viene dada por,

$$\max_{\alpha, \alpha^*} W(\alpha, \alpha^*) = \max_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) - \sum_{i=1}^l \alpha_i (y_i - \epsilon) - \alpha_i^* (y_i + \epsilon) \quad (4.28)$$

de forma alternativa,

$$(\bar{\alpha}, \bar{\alpha}^*) = \arg \min_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) - \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon \quad (4.29)$$

con limitaciones,

$$\begin{aligned} 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, l \\ \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \end{aligned} \quad (4.30)$$

Resolviendo la ecuación 4.28, teniendo en cuenta las limitaciones impuestas en la ecuación 4.30, determinamos los multiplicadores de Lagrange y la función de regresión viene dada por la ecuación 4.25 donde,

$$\begin{aligned} \bar{w} &= \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \\ \bar{b} &= -\frac{1}{2} (\bar{w}, (x_r + x_s)). \end{aligned} \quad (4.31)$$

Las condiciones del Karush-Kuhn-Tucker (KKT) satisfechas por la solución son,

$$\bar{\alpha}_i \bar{\alpha}_i^* = 0, \quad i = 1, \dots, l \quad (4.32)$$

Los vectores de soporte son puntos donde exactamente uno de los multiplicadores de Lagrange es mayor que cero. Cuando su valor es 0, obtenemos la función de pérdida L_1 y el problema de optimización se simplifica,

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j (x_i, x_j) - \sum_{i=1}^l \beta_i y_i \quad (4. 33)$$

con restricciones,

$$\begin{aligned} -C &\leq \beta_i \leq C, \quad i = 1, \dots, l \\ \sum_{i=1}^l \beta_i &= 0, \end{aligned} \quad (4. 34)$$

y la función de regresión viene dada por la ecuación 4.25, donde,

$$\begin{aligned} \bar{w} &= \sum_{i=1}^l \beta_i x_i \\ \bar{b} &= -\frac{1}{2} (\bar{w}, (x_r, x_s)). \end{aligned} \quad (4. 35)$$

4.1.4.1.2. Función de pérdida cuadrática

Empleando una función de pérdida cuadrática, Figura 4.8 (a),

$$L_{quad}(f(x) - y) = (f(x) - y)^2 \quad (4. 36)$$

La solución vendrá dada,

$$\begin{aligned} \max_{\alpha, \alpha^*} W(\alpha, \alpha^*) &= \max_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) \\ &\quad + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \frac{1}{2C} \sum_{i=1}^l (\alpha_i^2 - (\alpha_i^*)^2). \end{aligned} \quad (4. 37)$$

La optimización correspondiente se puede simplificar mediante la explotación de las condiciones KKT, ecuación 4.32, y observando que éstas implican $\beta_i^* = |\beta_i|$. Los problemas de optimización resultantes son,

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j (x_i, x_j) - \sum_{i=1}^l \beta_i y_i + \frac{1}{2C} \sum_{i=1}^l \beta_i^2 \quad (4.38)$$

con restricciones,

$$\sum_{i=1}^l \beta_i = 0. \quad (4.39)$$

y la función de regresión vendrá dada por las ecuaciones 4.25 y 4.35.

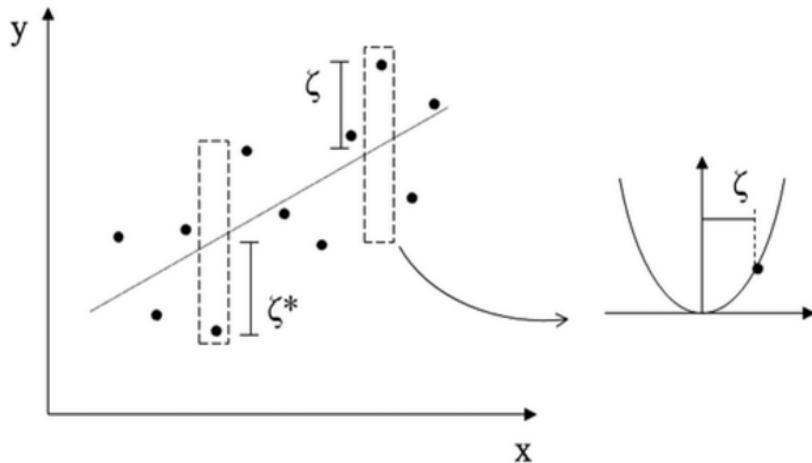


Figura 4.10. Función de pérdida cuadrática para SVR lineal. Fuente: Sez Atamturktur, M.ASCE y otros, 2015. Glenn Dept. of Civil Engineering, Clemson Univ., Clemson, SC 29634.

4.1.4.1.3. Función de pérdida de Huber

Empleando la función de pérdida de Huber, Figura 4.8 (c),

$$L_{Huber}(f(x) - y) = \begin{cases} \frac{1}{2}(f(x) - y)^2 & \text{for } |f(x) - y| < \mu \\ \mu |f(x) - y| - \frac{\mu^2}{2} & \text{en otro caso} \end{cases} \quad (4.40)$$

la solución vendrá dada por,

$$\begin{aligned} \max_{\alpha, \alpha^*} W(\alpha, \alpha^*) &= \max_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) \\ &\quad + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \frac{1}{2C} \sum_{i=1}^l (\alpha_i^2 - (\alpha_i^*)^2) \mu, \end{aligned} \quad (4. 41)$$

Los problemas de optimización resultantes son,

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j (x_i, x_j) - \sum_{i=1}^l \beta_i y_i + \frac{1}{2C} \sum_{i=1}^l \beta_i^2 \mu \quad (4. 42)$$

con restricciones,

$$\begin{aligned} -C \leq \beta_i \leq C, \quad i = 1, \dots, l \\ \sum_{i=1}^l \beta_i = 0, \end{aligned} \quad (4. 43)$$

4.1.4.1.4. Función de Riesgo

Para más simplicidad en la exposición, sólo se tendrá en cuenta el caso lineal, sabiendo que podría extenderse al caso no lineal utilizando el método kernel. Supongamos que disponemos de un conjunto $C = \{(x_1, y_1), \dots, (x_n, y_n)\}$, donde $x_i \in \mathbb{R}^d$ e $y_i \in \mathbb{R}$. Si se asume que el conjunto de datos se ha obtenido a partir de una función de probabilidad $P(x, y)$. Nuestro objetivo será encontrar una función f que minimice la función:

$$R[f] = \int c(x, y, f(x)) dP(x, y) \quad (4. 44)$$

*En la expresión anterior, $c(x, y, f(x))$ es el término que penaliza los errores de estimación. Al referirnos a esta función generada a partir del término descrito, lo haremos como función *coste*. Una posible aproximación consiste en sustituir la integración por la estimación empírica, para conseguir la denominada función de riesgo empírico:*

$$R_{emp}(f) := \frac{1}{\ell} \sum_{i=1}^{\ell} c(x_i, y_i, f(x_i)) \quad (4. 45)$$

En primera instancia, se trataría de encontrar la función $f := \arg \min_{f \in H} (R_{emp}[f])$ para una hipotética clase H . Sin embargo si la clase H es muy amplia, habrá una enorme cantidad de ejemplos, que tratándose de datos en espacios de muy alta dimensión, esto puede no ser una buena idea; ya que, existirá sobreajuste, y por tanto malas propiedades de generalización. Por lo cual, le añadimos un término de control de capacidad, que en el caso del algoritmo SV es $\|w\|^2$, lo que conduce al funcional de riesgo generalizado:

$$R_{reg}[f] := \frac{\lambda}{2} \|w\|^2 + R_{emp}[f] \quad (4.46)$$

donde $\lambda > 0$, será la constante de regularización.

4.1.4.2. Regresión no lineal

Para modelar adecuadamente los datos, algunas veces se requiere un modelo no lineal de forma similar a los problemas de clasificación. De la misma manera que el acercamiento SVC no lineal, se puede emplear un mapeo no lineal para mapear los datos en un espacio característico de alta dimensión, donde se realiza la regresión lineal. El enfoque del núcleo (kernel) se emplea para abordar el problema de la dimensionalidad. La solución SVR no lineal, utilizando una función de pérdida ϵ -insensible, Figura 4.8 (d), viene dada por,

$$\max_{\alpha, \alpha^*} W(\alpha, \alpha^*) = \max_{\alpha, \alpha^*} \sum \alpha_i^* (y_i - \epsilon) - \alpha_i (y_i + \epsilon) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) \quad (4.47)$$

con restricciones,

$$\begin{aligned} 0 \leq \alpha_i, \alpha_i^* \leq C, \quad & i = 1, \dots, l \\ \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \end{aligned} \quad (4.48)$$

Resolviendo la ecuación 4.47 con restricciones, la ecuación 4.48 determina los multiplicadores de Lagrange, α_i, α_i^* , y la función de regresión vendrá dada por,

$$f(x) = \sum_{SVs} (\bar{\alpha}_i - \bar{\alpha}_i^*) K(x_i, x) + \bar{b} \quad (4.49)$$

donde,

$$\begin{aligned}\langle \bar{w}, x \rangle &= \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) \\ \bar{b} &= -\frac{1}{2} \sum_{i=1}^l (\alpha_i - \alpha_i^*)(K(x_i, x_r) + K(x_i, x_s)).\end{aligned}\quad (4. 50)$$

Al igual que con el SVC, la restricción de igualdad puede ser eliminada si el Kernel contiene un término de sesgo, b , siendo acomodado dentro de la función Kernel, y la función de regresión viene dada por,

$$f(x) = \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*) K(x_i, x). \quad (4. 51)$$

Los criterios de optimización para otras funciones de pérdida se obtienen de manera similar reemplazando el producto de punto por una función de núcleo. La función de pérdida de ϵ -insensibilidad es atractiva porque a diferencia de las funciones de coste cuadrático y Huber, donde todos los puntos de datos serán vectores de soporte, la solución SV puede ser dispersa. La función de pérdida cuadrática produce una solución que es equivalente a la regresión de cresta, donde el parámetro de regularización $\lambda = \frac{1}{2C}$.

4.1.5. Truco del Kernel y tipos

Los problemas que no son lineales pueden ser tratados mediante el del truco del kernel. De este modo, podríamos transformar el espacio original en otro espacio en el que fuese más sencilla la separación de las muestras, sin tener que transformarlas. Las funciones del kernel (polinomial, funciones de base radial gausianas), transforman los datos en un espacio de características dimensionales superiores para permitir la realización de la separación lineal.

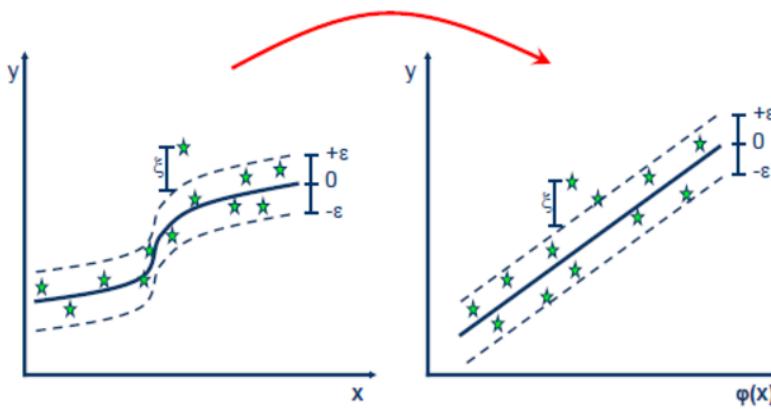


Figura 4.11. SRV-no lineal Fuente: http://www.saedsayad.com/support_vector_machine_reg.htm

Una vez hayamos comprobado que una función cumple los requisitos necesarios para ser una función kernel, existen diversas formas de trabajar con los kernel. De modo práctico, se emplean una serie de funciones kernel de las que conocemos sus propiedades y su campo de aplicación. A continuación se muestra una lista de los kernels más empleados en la práctica:

- Polinomial.
- Función de Base Radial Gaussiana.
- Función de Base Radial Exponencial.
- All-subsets Kernel.
- ANOVA Kernels.
- Perceptrón Multicapa ó Sigmoidal.
- Splines.
- Series de Fourier.

4.1.5.1. Polinomial

Definimos un kernel polinomial, derivado de K_1 como:

$$K(x, z) = p(K_1(x, z)) \quad (4.52)$$

Siendo $p(\cdot)$ un polinomio cualquiera con coeficientes positivos. Presentándose el siguiente caso especial:

$$K_d(x, x') = (\langle x, x' \rangle + R)^d \quad (4.53)$$

Definido sobre un espacio vectorial X , de dimensión n , en donde R y d son parámetros. El desarrollo del kernel polinomial K_d empleando el teorema binomial sería:

$$K_d(x, x') = \sum_{s=0}^d \binom{d}{s} R^{d-s} \langle x, z \rangle^s \quad (4. 54)$$

La dimensión del espacio de características del kernel polinomial sería: $\binom{n+d}{d}$

4.1.5.2. Función de Base Radial Gaussiana

Los kernels de base radial gaussiana son los utilizados y se han estudiado con mucho detalle en campos relacionados.

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4. 55)$$

Así por ejemplo si $K_1(x, z)$, es un kernel con función relacionada correspondiente ϕ_1 y con espacio de características F_1 , podríamos crear el kernel gaussiano en F_1 :

$$\|\phi_1(x) - \phi_1(z)\|^2 = K_1(x, x) - 2K_1(x, z) + K_1(z, z), \quad (4. 56)$$

quedando el kernel gaussiano derivado, de la forma siguiente:

$$K(x, z) = \exp(-\gamma(K_1(x, x) - 2K_1(x, z) + K_1(z, z))) \quad (4. 57)$$

Aquí el parámetro γ , controlaría el grado de flexibilidad del kernel del mismo modo que el grado d en los kernels polinomiales.

Función de Base Radial Exponencial

Estas funciones son de la forma:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|}{2\sigma}\right) \quad (4. 58)$$

4.1.5.3. “All-subsets Kernel”

Si consideramos un espacio con mapa de características ϕ_A para cada subconjunto $A \subseteq \{1, 2, \dots, b\}$ de entradas, estando incluido el conjunto vacío. Se empleará la función:

$$\phi_i(x) = x_1^{i_1}, x_2^{i_2}, \dots, x_n^{i_n}, \quad (4.59)$$

que con la restricción $i = (i_1, \dots, i_n) \in \{0, 1\}^n$, la función asociada ϕ_A vendrá dada por:

$$\phi_A(x) = \prod_{i \in A} x_i \quad (4.60)$$

Lo cual genera todos los monomios característicos para todas las combinaciones de índices hasta n , pero aquí, al contrario que en el caso polinomial, cada factor en el monomio tiene grado 1. Así, “All-subset kernel” viene determinado por la función ϕ :

$$\phi: x \mapsto (\phi_A(x))_{A \subseteq \{1, \dots, n\}} \quad (4.61)$$

Siendo el kernel $K \subseteq (x, z)$ determinado por:

$$K \subseteq (x, z) = \langle \phi(x), \phi(z) \rangle = \sum_{A \subseteq \{1, \dots, n\}} \prod_{i \in A} x_i z_i \quad (4.62)$$

4.1.5.4. Anova Kernels

Tanto el kernel polinomial, como all-subsets kernel, presentan limitaciones en cuanto qué características usar, el control y la forma de ponderar sus parámetros.

El Anova kernel K_d es un método que permite mayor libertad en la especificación del conjunto de monomios. Es como el ‘All-subsets’, excepto que se restringe a subconjuntos cardinalidad d . Emplea la notación x^i para la expresión $x_1^{i_1}, x_2^{i_2}, \dots, x_n^{i_n}$, donde $i = (i_1, \dots, i_n) \in \{0, 1\}^n$ aplicándose además la siguiente restricción:

$$\sum_{j=1}^n i_j = d \quad (4. 63)$$

Para el caso en que $d = 0$, hay una característica con valor constante 1 correspondiente al conjunto vacío. La diferencia entre el ANOVA y el polinomial es la exclusión de las coordenadas repetidas.

La función del ANOVA kernel de grado d, vendrá dada por:

$$\phi_d : x \alpha (\phi_A(x))_{|A|=d} \quad (4. 64)$$

estando cada subconjunto A, definido por:

$$\phi_A(x) = \prod_{i \in A} x_i = x^{i_A} \quad (4. 65)$$

siendo i_A , la función indicador del conjunto A.

La dimensión del resultante ϕ_A , es $\binom{n}{d}$, ya que es el número de tales subconjuntos, mientras que el producto interno resultante vendría dado por:

$$\begin{aligned} K_d(x, z) &= \langle \phi_d(x), \phi_d(z) \rangle = \sum_{|A|=d} \phi_A(x) \phi_A(z) = \sum_{1 \leq i_1 < i_2 < \dots < i_d \leq n} (x_{i1} z_{i1})(x_{i2} z_{i2}) \dots (x_{id} z_{id}) = \\ &= \sum_{1 \leq i_1 < i_2 < \dots < i_d \leq n} \prod_{j=1}^d x_{ij} z_{ij} \end{aligned} \quad (4. 66)$$

4.1.5.5. Perceptrón Multicapa ó Sigmoidal

El perceptrón más establecido, con una capa oculta, también tiene una representación válida en los kernels,

$$K(x, x') = \tanh(\tanh \langle \gamma \langle x, x' \rangle \rangle + R) \quad (4. 67)$$

Para ciertos valores escalares γ y R que aquí serían los parámetros de nuestro kernel.

4.1.5.6. Splines

Los Splines son muy empleados por su flexibilidad. Un Spline finito, de orden k , con N nodos situados en τ_s es de la forma:

$$K(x, x') = \sum_{r=0}^K x^r x'^r + \sum_{s=1}^N (x - \tau_s)^k + (x' - \tau_s)^k + d\tau \quad (4. 68)$$

Definimos un Spline infinito en el intervalo $[0,1]$:

$$K(x, x') = \sum_{r=0}^K x^r x'^r + \int_0^1 (x - \tau_s)^k + (x' - \tau_s)^k + d\tau \quad (4. 69)$$

De forma particular, si $K = 1, (S_1^\infty)$ el kernel vendrá dado por:

$$K(x, x') = 1 + \langle x, x' \rangle + \frac{1}{2} \langle x, x' \rangle \min(x, x') - \frac{1}{6} \min(x, x')^3 \quad (4. 70)$$

En donde la solución es cúbica, a trozos. Los splines son kernels univariantes, y, sabiendo que a partir del producto tensorial de kernels se obtiene un nuevo kernel, podemos construir kernels splines multivariantes.

Series de Fourier

Las series de Fourier consideran una expansión en un espacio de características $2N + 1$ dimensional. Este kernel viene definido en el intervalo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

$$K(x, x') = \frac{\operatorname{sen}\left(N + \frac{1}{2}\right)(x - x')}{\operatorname{sen}\left(\frac{1}{2}(x - x')\right)} \quad (4. 71)$$

De forma idéntica a los kernels splines, pueden emplearse de base para construir kernels multidimensionales.

4.1.6. Estimación de Parámetros del SVR, Validación Cruzada

Cuando introducimos el kernel, éste depende a su vez de unos parámetros de los cuales depende la solución. Es necesario facilitar un procedimiento de selección de dichos parámetros. Dicho procedimiento, consiste en definir una malla, sobre el espacio de los parámetros, y seleccionar aquellos valores de los parámetros que proporcionan mejores estimadores del error esperado de predicción. Para realizar dicha estimación del error, se emplea la validación cruzada, que se detalla seguidamente.

Validación Cruzada

Este método estima directamente la generalización del error cuando se aplica el método, de función de predicción $\hat{f}(x)$, a una muestra de test. La notación empleada será $Err = E[L(y, \hat{f}(x))]$, siendo $L(y, \hat{f}(x))$ la función de pérdida ajustada. En aquellos casos en los que no dispongamos de los datos suficientes, podemos emplear la validación cruzada para evaluar nuestro predictor. Si dispusiéramos de un grupo de datos lo suficientemente grande, podríamos seleccionar un conjunto de validación para evaluar nuestro modelo de predicción. Dado que la mayor parte de las veces el conjunto no es lo bastante grande como se requiere, no es viable. Cuando se aplica la validación cruzada con k -pliegues, estamos subdividiendo el conjunto de datos en k partes de aproximadamente el mismo tamaño. Cada una de las k divisiones de los datos de entrenamiento se la denomina “fold”. Utilizamos $k-1$ partes para ajustar el modelo y el resto para validar. Este mismo procedimiento se repetirá k veces, seleccionando cada vez un conjunto de validación distinto. En cada caso calculamos el error de predicción del modelo, y, por último promediamos estos errores [77]. Siendo $K : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$ una función que indica la partición, a la cual se asigna la i -ésima observación por la aleatorización. Si $j \in \{1, \dots, n\}$ denotaremos por $\hat{f}^{-K(j)}(x)$ la función ajustada con el conjunto de datos a excepción del subconjunto $K(j)$ -ésimo. La correspondiente estimación de error mediante la validación cruzada con k -pliegues es:

$$CV(\hat{f}) = \frac{1}{n} \sum_{i=1}^n L(y_i, \hat{f}^{-K(i)}(x_i)). \quad (4.72)$$

Las elecciones más comunes son $k=5$ y $k=10$. El caso $k=n$ es conocido como dejar uno fuera (en inglés, *leave-one-out (LOO)*) [78]. En este caso $-(i) = i$, y la función de predicción para el caso i -ésimo se realiza con la totalidad del conjunto excepto el i -ésimo elemento.

La validación cruzada es muy costosa, y por eso se reserva para los casos donde el conjunto de datos es muy pequeño. Hay algunos casos especiales en los que es beneficioso utilizar otros k -pliegues.

Las muestras no se dividen en subconjuntos, simplemente omiten en cada iteración un dato que será aplicado para la clasificación, y entrena con los datos restantes. El resultado con el dato que se clasifica se almacena, se pasa el siguiente dato a la fase de clasificación y se entrena con todos los demás incluyendo el dato que se clasificó la primera vez.

Una vez clasificado el último dato de la muestra, calcula con todos los datos el porcentaje de acierto [79]. Las ventajas de este algoritmo son que utiliza un mayor número de datos para entrenar, y que tiene un resultado determinista, pues el experimento siempre ofrecerá los mismos resultados. Sin embargo, poniendo un ejemplo supongamos dos clases en un conjunto de muestras, con la mitad de muestras pertenecientes a cada clase.

Cuando una muestra es extraída del conjunto de datos para la clasificación, su clase queda en minoría en el entrenamiento, la mayoría induce a predecir siempre la clase equivocada puesto que el Leave-One-Out estima un porcentaje de acierto para la mayoría de muestras, por lo que el resultado será un 0% de acierto [80].

La técnica de Leave-one-out (LOO CV)

Es una validación cruzada con $k = \text{número de datos de entrenamiento}$. Si hay N datos de entrenamiento, repetir $k=n$ veces:

- Reservar el dato número N para test
- Entrenar con los N-1 datos restantes
- Hacer el test con el dato N (el resultado sólo puede ser acierto o fallo)

El porcentaje de aciertos esperado será:

$$\frac{\text{Aciertos}}{N} \cdot 100 \quad (4.73)$$

Es preciso porque se usan casi todos los datos para entrenar, y a la vez todos los datos figuran como test en alguno de los ciclos, pero es tremadamente costoso en tiempo (hay que lanzar el algoritmo de aprendizaje N veces).

4.2. Árboles de decisión

Los árboles de decisión se utilizan desde hace siglos, y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, y otros. Un árbol de decisión es una colección de cero o más *nodos* vinculados mediante una relación de jerarquía. Si tiene cero nodos se denomina árbol *vacío*.

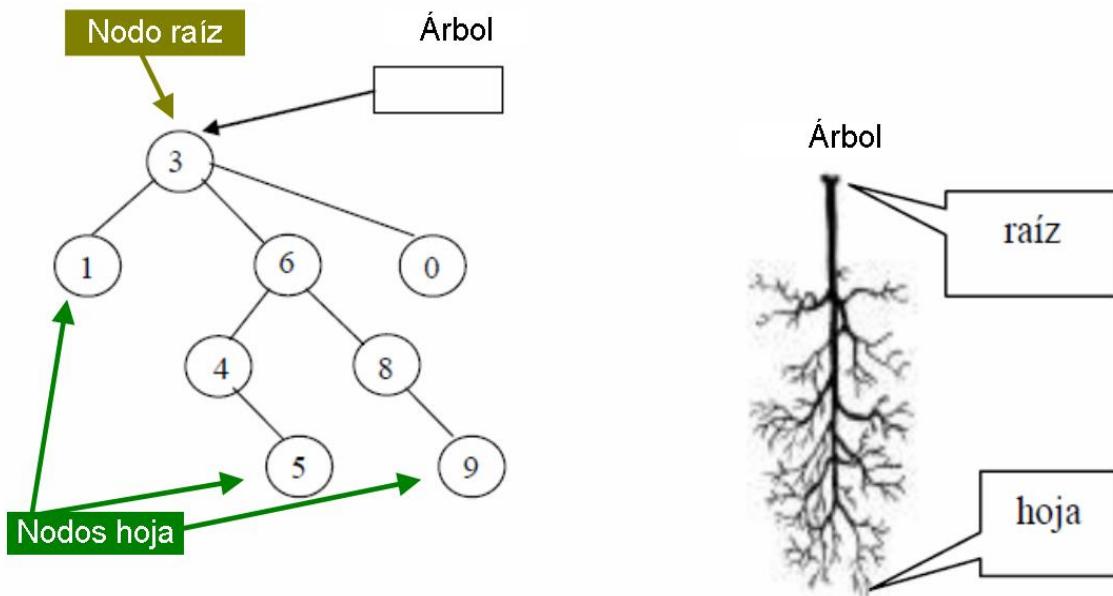


Figura 4.12. Árboles de decisión y su analogía con los árboles vegetales.
Fuente: elaborada a partir de L. Silva Bijit, 2010

En todo árbol existe un nodo especial, o punto de entrada a la estructura, denominado *raíz*, que puede tener cero o más nodos accesibles desde ella. La raíz es por tanto, el *ancestro* de sus descendientes, por lo que podemos decir que el nodo raíz no tiene ancestros. En un árbol de decisión cada nodo del árbol es un atributo o campo de los ejemplos, y cada rama representa un posible valor de ese atributo.

Un *subárbol* es un nodo con todos sus descendientes, siendo el conjunto de esos nodos *subárboles* de la raíz, y nodos *descendientes* de dicha raíz. Un nodo sin descendientes es una *hoja* que no tiene nodos *hijos*.

En la Figura 4.12, el nodo con valor 3 es la raíz. Los nodos con valores: 1, 5 y 9 son hojas. Los nodos con valores: 4, 6 y 8 son *nodos internos*. El nodo con valor 6 es hijo de 3 y padre de 8. El nodo con valor 4 es ancestro del nodo con valor 5. $\{8, 9\}$ y $\{4, 5\}$ son sub-árboles. El nodo con valor 1 es subárbol de la raíz. Los nodos con valores: 1, 6 y 0 son *hermanos*, por tener el mismo parente. El conjunto de nodos con valores: $\{3, 6, 4, 5\}$ es una trayectoria, de largo 3, siendo 2 el alto del nodo con valor 6, y 3 La profundidad del nodo con valor 5. Todos los nodos que están a igual profundidad están en el mismo *nivel*. La profundidad del árbol es la profundidad de la hoja más profunda. Se dice que un árbol es una estructura *ordenada*, ya que los hijos de un nodo se consideran ordenados de izquierda a derecha. También es una estructura *orientada*, ya que hay un camino único desde un nodo hacia sus descendientes.

4.2.1. Complejidad o altura en árboles.

4.2.1.1. Árbol completo

Se denomina árbol completo, a aquél que tiene presentes todas las hojas en el menor nivel. La raíz es de nivel cero, los hijos de la raíz están en nivel 1; y así sucesivamente. En la Figura 4.15, se representan dos árboles de nivel 1 y 3 respectivamente que tienen todas las hojas en el menor nivel.

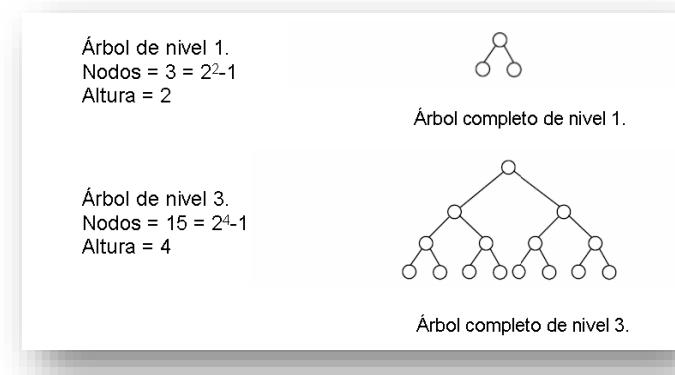


Figura 4.13. Árboles con todas las hojas en el menor nivel. Fuente: elaboración de L. Silva Bijit, 2010

De la Figura 4.13, se deduce fácilmente que para el caso de un árbol completo de ‘ m ’ niveles, el número de nodos, la altura, el número de hojas y los nodos internos se determinan de la siguiente forma:

$$\begin{aligned} &\text{Árbol de nivel } m. \\ &\text{Nodos} = n = 2^A-1 \\ &\text{Altura} = A = m+1 \\ &\text{Hojas} = 2^m \\ &\text{Nodos internos} = n - \text{Hojas} \end{aligned}$$

4.2.1.2. Árboles incompletos con un nivel de desbalance

Se ilustran los tres casos de árboles, de nivel dos, con un nivel de desbalance, para $n= 4, 5$ y 6 . Un árbol con 3 nodos es completo en caso de aceptarse sólo un nivel de desbalance. Lo mismo puede decirse de un árbol con 7 nodos y que tenga un nivel de desbalance.



Figura 4.14 . Árboles incompletos de nivel 2. Fuente: elaboración de L. Silva Bijit, 2010

De la Figura 4.14, se deduce que para el caso de un árbol incompleto de m niveles, el número de nodos en función de la altura se determina de la forma siguiente:

Árboles de m niveles.

Nodos de 2^{A-1} hasta 2^{A-2} .

Altura A .

Una de las grandes ventajas de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite analizar una situación y, llegar a una sola acción o decisión a tomar. Estos algoritmos se llaman algoritmos de *partición* o algoritmos de “divide y vencerás” [L. Silva Bijit 2010 – Universidad Técnica Federico Santa María]. El algoritmo básico de los árboles de decisión se da a continuación:

```
ALGORITMO Partición( $N$ :nodo,  $E$ :cjo. de ejemplos)
SI todos los ejemplos  $E$  son de la misma clase  $c$  ENTONCES
    Asignar la clase  $c$  al nodo  $N$ .
    SALIR; // Esta rama es pura, ya no hay que seguir partiendo.  $N$  es hoja.
SI NO:
    Particiones := generar posibles particiones.
    MejorPartición:= seleccionar la mejor partición según el criterio de partición.
    PARA CADA condición  $i$  de la partición elegida.
        Añadir un nodo hijo  $i$  a  $N$  y asignar los ejemplos consistentes a cada hijo ( $E_i$ ).
        Partición( $i$ ,  $E_i$ ). // Realizar el mismo procedimiento global con cada hijo.
    FIN PARA
FIN SI
FIN ALGORITMO
```

Para clasificar un cjo de ejemplos E , se invoca con la llamada Partición(R, E), donde R es un nodo raíz de un árbol por empezar.

Tabla 4.1. Algoritmo básico de los árboles de decisión.

Fuente: Universidad Carlos III, Madrid – Departamento de Ingeniería Telemática

El algoritmo de la Tabla 4.1, construye el árbol añadiendo las particiones y los hijos resultantes de cada partición desde la raíz. En cada partición los ejemplos se van dividiendo entre los hijos, hasta que se llega a la situación en la que todos los ejemplos que caen en los nodos inferiores son de la misma clase; por lo que, esa rama ya no seguirá creciendo. La única condición exigible es que las particiones separen los ejemplos en distintos hijos, con lo que la cardinalidad de los nodos irá disminuyendo a medida que se desciende en el árbol.

Para que el algoritmo anterior funcione bien, se deberá prestar atención a los dos puntos siguientes:

Las particiones. Se deberán elegir un conjunto de condiciones exhaustivas y excluyentes. Cuantas más particiones permitamos más explícitos y precisos serán los árboles de decisión generados. Si bien, la complejidad del algoritmo será mayor al aumentar el número de particiones. Por tanto, buscaremos un buen compromiso entre estos aspectos.

Criterio de selección de particiones. El número de particiones posibles puede dispararse. En los algoritmos clásicos de aprendizaje de decisión, una vez elegida la partición se continúa hacia abajo la construcción del árbol y no vuelven a plantearse las particiones ya construidas. Por tanto, se establecerá un criterio que permita realizar una buena elección de la partición que parezca más prometedora y que no genere demasiado esfuerzo computacional.

4.2.2. Sobreaprendizaje

4.2.2.1. El problema de sobreaprendizaje

El *problema de sobreaprendizaje* o sobre-adaptación (en inglés, “overfitting”), requiere tener en cuenta algunas consideraciones a la hora de implementar algoritmos de aprendizaje de máquinas, ya que los árboles de decisión tienden a ajustarse demasiado al conjunto de entrenamiento utilizado para la construcción del árbol. Demasiadas ramas pueden dar lugar a ruido y observaciones anómalas, originadas por una serie de valores atípicos (en inglés, ‘outliers’). Estos ‘outliers’, son en realidad valores extremos en un conjunto de datos que son muy diferentes de los otros valores. El resultado final es un árbol más complejo de lo que en realidad debería ser; como consecuencia, disminuye la precisión del clasificador en su aplicación a situaciones nuevas. Un exceso de atributos puede ocasionar problemas de sobreaprendizaje, pues incrementa la complejidad del modelo; especialmente en el supuesto de que existan pocos datos en el conjunto a estudiar.

Algunos atributos pueden resultar redundantes y ralentizar el proceso de aprendizaje. Otros, en cambio, serán irrelevantes para el caso de estudio, (p. ej., el NIF de un cliente si se quiere predecir si va a poder devolver o no un crédito). Por consiguiente, se producirá sobre-adaptación cuando el clasificador obtiene un alto porcentaje de aciertos en el entrenamiento, pero muy pequeño en el test; esto quiere decir, que el modelo no generaliza bien.

En este caso anterior, se puede decir que el clasificador está memorizando los datos en lugar de generalizarlos (p. ej., un alumno “aprende” a realizar ejercicios de los exámenes anteriores, pero no sabe resolver los del examen final).

4.2.2.2. Solución al problema del sobreaprendizaje

El problema de sobreaprendizaje o sobreajuste se resuelve aplicando técnicas de poda. Una opción sería detener el crecimiento del árbol, antes de que éste llegue a adaptarse perfectamente al conjunto de entrenamiento y clasifique perfectamente todos los ejemplos. Esta técnica se conoce como pre-poda (en inglés ‘*pre-pruning*’). Otra opción consistiría en permitir que el árbol se sobreajuste a los datos y luego efectuar sobre él una poda. Esta técnica es la post-poda (en inglés, ‘*post-pruning*’). En la práctica se usa más la técnica de ‘postpoda’, porque es difícil estimar con precisión cuando se debe detener el crecimiento del árbol. Para aquellos casos en los que no se dispone de muchos datos, se recomienda emplear la validación cruzada; es decir, aplicar técnicas de entrenamiento y testeo de los resultados.

Un árbol de decisión se poda de dos posibles formas, o bien se sustituye un subárbol por un nodo hoja que se corresponda con la clase más frecuente en el subárbol; o bien se sustituye el subárbol por otro subárbol contenido en el primero. Estos métodos, introducirán errores adicionales de clasificación en el conjunto de entrenamiento, no obstante, si la poda se realiza de manera correcta, la precisión del clasificador aumentará. La técnica propuesta por Quinlan en 1987, considera que cada nodo del árbol es un candidato para la poda, es decir, para eliminarlo del árbol.

Podar un nodo significa eliminar la rama que parte de ese nodo convirtiéndolo en un nodo hoja (que es más general). Siempre que el árbol resultante no sea peor que el original podremos realizar los pasos anteriores sobre el conjunto de test, repitiendo el proceso hasta que el árbol obtenido empeore el rendimiento.

4.2.3. Árboles de Clasificación y Regresión (ACR)

El término Árboles de Clasificación y Regresión, ACR (en inglés, *Classification and Regression Trees - CART*), es un término genérico utilizado para referirse a ambos de los procedimientos anteriores, introducido por primera vez por Breiman et al., 1984 [81]. El planteamiento de los árboles de clasificación y regresión por Leo Breiman y otros en 1984, despertó gran interés en la utilización de esta metodología por parte de la comunidad científica, debido a su fácil implementación para plantear todo tipo de problemas y su claridad a la hora de interpretar los resultados. Un gran número de investigadores han planteado variaciones del método en sus distintas etapas tras de la publicación del libro de Breiman et al., 1984, pero en numerosos casos la idea inicial del particionamiento recursivo sigue siendo la misma. Otros, por el contrario, no plantean cambios significativos, pero han optado por aplicar CART y sus variaciones a distintos campos o disciplinas como la medicina, la biología y el aprendizaje de máquinas. El Aprendizaje basado en árboles de decisión se emplea tanto en la estadística y el aprendizaje automático como en la minería de datos; así mismo, constituye una de las técnicas más eficaces para la clasificación supervisada.

En la estructura de un árbol de decisión, las hojas representan etiquetas de clase y las ramas representan los grupos de características que proporcionan esas etiquetas de clase. Son *árboles de clasificación*, aquellos árboles de decisión en donde la variable de destino puede tomar un conjunto finito de valores; en cambio, aquellos donde la variable de destino puede tomar valores continuos, que son generalmente números reales, se llaman *árboles de regresión*.

Un árbol puede ser "aprendido" mediante el fraccionamiento en subconjuntos del conjunto de datos de partida, repitiéndose el proceso en cada subconjunto mediante un particionamiento recursivo. Dicha recursividad termina, cuando el subconjunto en un determinado nodo tiene el mismo valor que el de la variable denominada objetivo, o cuando la partición efectuada ya no agrega más valor a las predicciones.

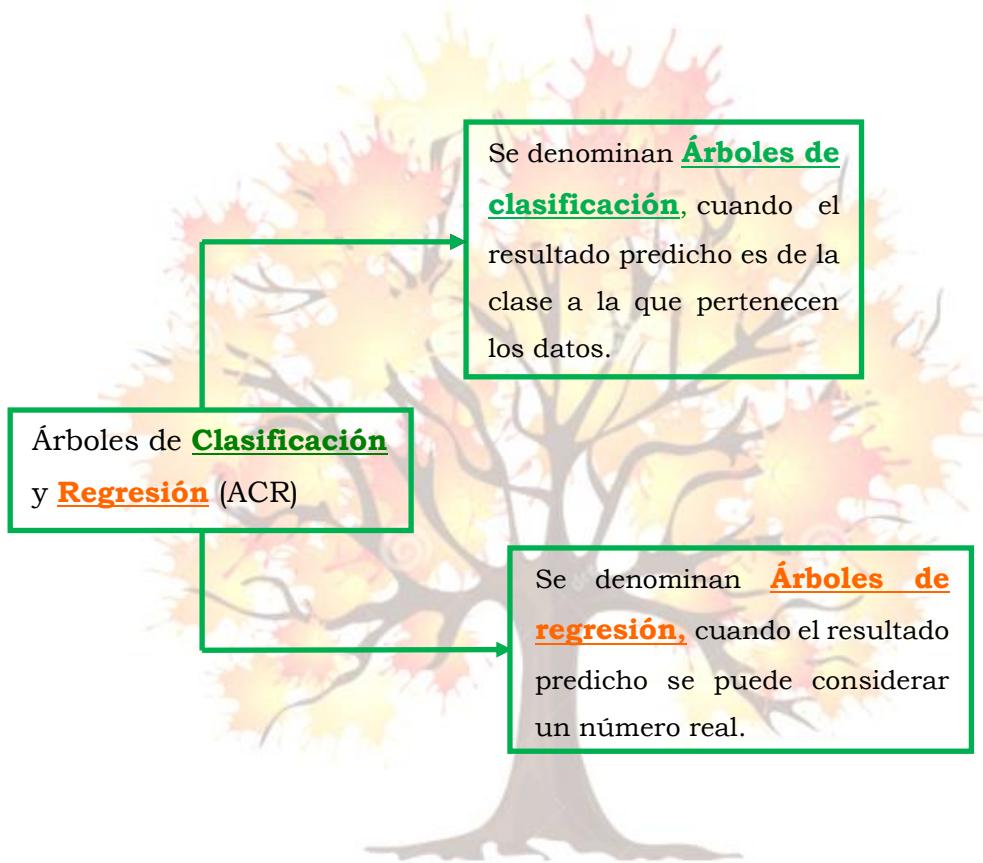


Figura 4.15. Árboles de clasificación y regresión (ACR). Elaboración propia.

Se denomina *Aprendizaje basado en árboles de decisión*, a una construcción de un árbol de decisión a partir de tuplas o listas de elementos debidamente ordenados que sirven para el adiestramiento. Cada una de estas tuplas deberá estar etiquetada con su correspondiente clase. Un árbol de decisión tiene una estructura idéntica a la de un diagrama de flujo, donde cada nodo interno (no en el caso de los nodos hoja) expresa una prueba en un atributo, cada rama representa el resultado de una prueba, y cada nodo hoja o terminal, tiene una etiqueta de clase. El nodo superior de un árbol de decisión es designado como nodo raíz.

4.2.3.1. Predicción de variables en los árboles de regresión

Un *árbol de decisión* en su forma más simple modela una decisión binaria, es decir, sí o no, o eventos mutuamente exclusivos como en la Figura 4.16.

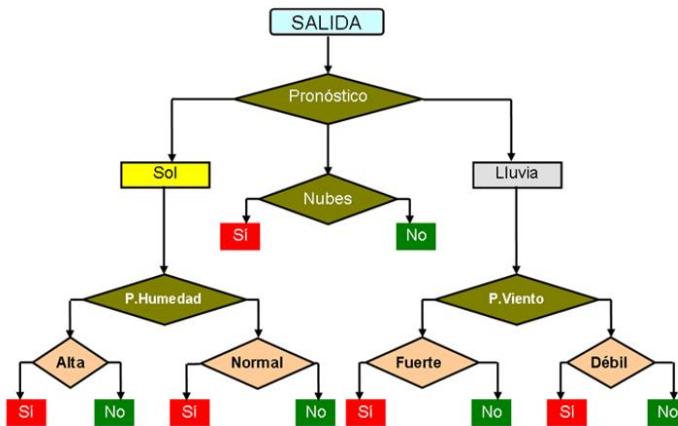


Figura 4.16. Sucesos que podrían modelarse mediante un árbol de decisión.
Fuente: elaboración propia

Un *árbol de decisión* es una forma de clasificador que decide cómo proceder cada vez modelando sus decisiones de clasificación. Según el ejemplo anterior, el clasificador en este caso tratará de predecir el pronóstico dado un conjunto de variables como el viento y la humedad.

La regresión en cambio es una forma de predecir el valor de una variable dependiente 'Y' dada la variable independiente 'X', dado que X e Y están "interconectados" por una función f . Y es una función de X o tiene una relación con X por f de manera lineal. A veces, habrá más que una función f para un intervalo dado de 'X', como en la Figura 4.17. Otra forma de verlo es que f no es lineal. Este comportamiento de la variable dependiente de Y puede ser modelado usando un *árbol de regresión* [Alket Cecaj, Universidad de Modena e Reggio Emilia].



Figura 4.17. Varias funciones en un intervalo. Fuente: <http://www.uidownload.com/free-vectors/businessman-showing-growing-charts-free-vector-436205>

Los *árboles de regresión* representan una técnica de análisis discriminante no paramétrica que permiten inferir la asignación de muestras a grupos predefinidos, en función de una serie de variables predictoras [L. C. Delgado, área de Biodiversidad y Conservación de la Universidad Rey Juan Carlos]. Es una técnica discriminante porque permite "discriminar" entre grupos. Más importante que la interpretación de la relación entre la variable dependiente y las variables predictoras, es aquí la capacidad predictiva del modelo que generemos. En los *árboles de regresión* se emplean variables cualitativas o categóricas, que son aquellas cuyos valores o características son categorías o clases excluyentes. Ejemplo: sexo, estado civil, nivel socioeconómico, color del pelo, color de los ojos, etc.

Por tanto, podrán tomar uno, de un número limitado y por lo general fijo, de posibles valores, asignando cada unidad individual o tipo de observación a un grupo en particular o categoría nominal sobre la base de alguna característica cualitativa. Las variables cualitativas o categóricas [82], son aquellas cuyos elementos de variación tienen carácter cualitativo, miden sus atributos o adjetivos y generalmente son dicotomas; es decir, que se dividen en dos.

En base a lo anterior, se puede entender que los datos categóricos pueden no tener un orden lógico y contienen un número finito de categorías o grupos distintos. Así, disponiendo de una variable de respuesta categórica, mediante árboles de regresión se pueden adoptar una serie de reglas, basadas en variables predictoras, que nos van a permitir asignar una nueva observación a un grupo u a otro. De este modo, dadas unas variables predictoras se podrá inferir correctamente la variable respuesta que, no olvidemos, deberá ser categórica.

4.2.3.2. Particionamiento recursivo

El algoritmo conocido como particionamiento recursivo [83] es el proceso paso a paso para construir un árbol de decisión y es la clave para el método estadístico no paramétrico CART [84].

Si Y es una variable respuesta y P son las variables predictoras x_1, x_2, \dots, x_p , donde las x 's son tomadas fijas, y la Y es una variable aleatoria; el problema estadístico consiste en establecer una relación entre las variables 'x' y la variable de respuesta Y , de tal manera que sea posible predecir el valor de Y basándose en los valores de las variables 'x', o variables predictoras. Matemáticamente, se quiere estimar la probabilidad condicional de la variable aleatoria Y ,

$$P[Y = y | x_1, x_2, \dots, x_p] \quad (4.74)$$

cuando la variable y es discreta, o un funcional de su probabilidad tal como la esperanza condicional,

$$E[Y | x_1, x_2, \dots, x_p] \quad (4.75)$$

Cuando la variable y es una variable continua.

Elementos que intervienen en la construcción de un árbol de decisión

Un árbol de decisión contiene varios niveles de nodos. Segundo se aprecia en la Figura 4.18 a), el primer nivel, está formado por un único nodo en forma de círculo en la cima, denominado nodo raíz. Dicho nodo raíz contiene una muestra de sujetos desde la cual se aumenta el árbol, desprendiéndose así los demás nodos. Estos sujetos constituyen lo que se llama una muestra de aprendizaje, la cual puede ser la muestra total en estudio o una parte de esta. En un segundo nivel se puede ver un nodo interno, también en forma de círculo, así como tres nodos terminales en forma de cajas que ocupan el segundo y tercer nivel respectivamente. Tanto el nodo raíz como el nodo interno se partitionan cada uno en dos nodos, en el siguiente nivel, los cuales se denominan nodos hijos, izquierdo y derecho [Zhang & Singer (2010)].

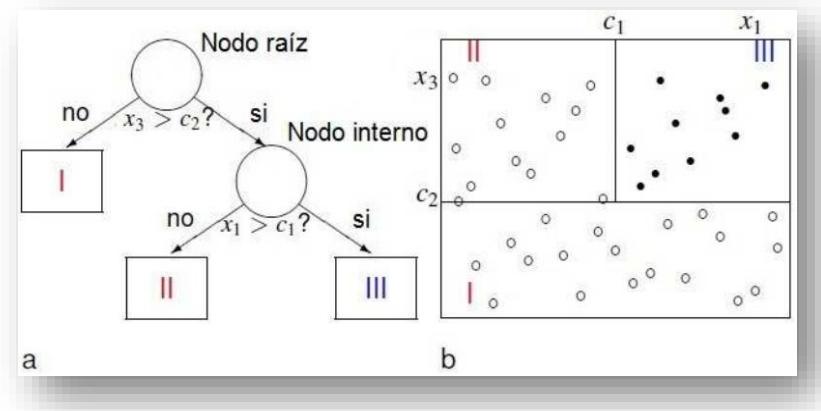


Figura 4. 18. a) Ejemplo árbol y b) Representación de las distintas clases.
Fuente: modificado de Zhang & Singer 2010.

El objetivo del particionamiento recursivo es obtener nodos terminales que sean homogéneos, en el sentido de que éstos, según el ejemplo presentado en la Figura 4.18 b), contengan sólo una de las clases presentadas, es decir, puntos o círculos. Conseguir una homogeneidad completa en los nodos terminales es difícilmente alcanzable en un análisis real de datos.

Según esto, el objetivo del particionamiento recursivo consistirá en conseguir que las variables resultantes en los nodos terminales sean tan homogéneas como sea posible. Una medida cuantitativa de la homogeneidad sería la impureza, que podemos obtener de la siguiente forma:

$$\text{Impureza del nodo} = \frac{n^o \text{ de sujetos que cumplen la característica del nodo}}{n^o \text{ total de sujetos del nodo}} \quad (4.76)$$

4.2.3.3. División de un nodo

Para dividir el nodo raíz en dos nodos homogéneos, deberemos seleccionar entre los rangos de todas las variables predictoras, el valor de la división que más lleve al límite de 0 ó 1 el cociente en la ecuación anterior, para cada nodo hijo. En la Figura 4.20 a) se seleccionó como división el valor ' c_2 ' entre el rango de la variable ' x_3 '. El proceso continúa para los dos nodos hijos, tomando en cuenta para cada nodo el rango resultante de la variable con la que se dividió el nodo padre y el rango de las demás variables involucradas.

Se busca realizar una división que de como resultado dos nodos hijos puros u homogéneos. Sin embargo, realmente los nodos hijos son por lo general parcialmente puros. Además, la bondad de una división deberá poner en una balanza la homogeneidad o la impureza de forma simultánea de los dos nodos hijos. Si se toma la covariable x_1 con el valor de corte c como alternativa para dividir un nodo, como resultado de la pregunta “¿es $x_1 > c$?” se tiene la Tabla 4.2:

	Y=0	Y=1	
Nodo izquierdo	n_{11}	n_{12}	$\mathbf{n}_{1\cdot}$
Nodo derecho	n_{21}	n_{22}	$\mathbf{n}_{2\cdot}$
	$\mathbf{n}_{\cdot 1}$	$\mathbf{n}_{\cdot 2}$	

Tabla 4. 2. Resultado de nodos tomando covariable x_1 y valor de corte c

Sea Y una variable dicotómica que toma valores 0 y 1. Se estima $P[Y=1|\tau_L]$ y $P[Y=1|\tau_R]$ por $\frac{n_{12}}{n_{1\cdot}}$ y por $\frac{n_{22}}{n_{2\cdot}}$, respectivamente y se introduce el concepto “entropía” en el nodo hijo izquierdo definida como:

$$i(\tau_L) = -\frac{n_{11}}{n_{1\cdot}} \log\left(\frac{n_{11}}{n_{1\cdot}}\right) - \frac{n_{12}}{n_{1\cdot}} \log\left(\frac{n_{12}}{n_{1\cdot}}\right) \quad (4. 77)$$

De forma análoga se define la impureza en el nodo hijo derecho como:

$$i(\tau_R) = -\frac{n_{21}}{n_{2\cdot}} \log\left(\frac{n_{21}}{n_{2\cdot}}\right) - \frac{n_{22}}{n_{2\cdot}} \log\left(\frac{n_{22}}{n_{2\cdot}}\right) \quad (4. 78)$$

Entonces, la bondad de una división, s , viene dada por:

$$\Delta I(s, \tau) = i(\tau) - P[\tau_L]i(\tau_L) - P[\tau_R]i(\tau_R), \quad (4. 79)$$

donde τ es el nodo padre de τ_L y τ_R , y $P[\tau_L]$ y $P[\tau_R]$ indican la probabilidad de que un sujeto caiga dentro de los nodos τ_L y τ_R respectivamente.

Aquí, $P[\tau_L]$ se puede considerar como $\frac{n_{1\cdot}}{n_{1\cdot} + n_{2\cdot}}$ y $P[\tau_R]$ como $\frac{n_{2\cdot}}{n_{1\cdot} + n_{2\cdot}}$.

4.2.3.3.1. Nodos terminales

El proceso de *particionamiento recursivo* continúa hasta que el árbol sea saturado, en el sentido de que los sujetos en los nodos descendientes no se pueden partir en una división adicional; es decir, cuando queda sólo un sujeto en un nodo. El número total de divisiones permitidas para un nodo disminuye cuando aumentan los niveles del árbol. Cualquier nodo que no pueda o no sea dividido se convierte en un nodo terminal.

El árbol saturado generalmente es bastante grande para utilizarse, porque los nodos terminales son tan pequeños que no permiten hacer inferencia estadística razonable debido a que los datos están sobre-ajustados; es decir, el árbol alcanza un ajuste tan fiel a la muestra de aprendizaje que cuando en la práctica se aplique el modelo obtenido a nuevos datos los resultados pueden ser muy malos, y por tanto, no es necesario esperar hasta que el árbol sea saturado. En lugar de esto, se escoge, a priori, un tamaño mínimo de nodo deteniéndose la división cuando el tamaño del nodo es menor que el mínimo elegido. La selección del tamaño mínimo dependerá del tamaño de muestra (uno por ciento), o se pueden tomar simplemente cinco sujetos (los resultados generalmente no son significativos con menos de cinco sujetos). Dependiendo del límite de parada, el particionamiento tiende a terminar muy pronto o muy tarde [Breiman et al., 1984]. En consecuencia, ellos hacen un cambio fundamental introduciendo un segundo paso llamado “poda”. La poda consiste en encontrar un subárbol del árbol saturado que sea el más “predictivo” de los resultados y menos vulnerable al ruido en los datos. Los subárboles se obtienen podando el árbol saturado desde el último nivel hacia arriba. Los pasos de particionamiento y poda se pueden ver como variantes de los procesos paso a paso (en inglés, forward y backward) en regresión lineal.

4.2.4. Árboles Modelo M5

En situaciones donde lo que se pretende predecir son valores continuos, algunas veces se requiere modificar convenientemente los algoritmos. De forma habitual, se ha venido empleando la regresión lineal para resolver este problema. El problema que se plantea es que los atributos a emplear deben ser numéricos y, a partir de esto, los modelos que se obtienen serán también numéricos. En los años 80, Breiman et al., propusieron un sistema denominado CART similar al algoritmo de inducción de árboles de clasificación ID3 [Quinlan, 1986]. El sistema CART, genera en los nodos hoja valores numéricos en vez de clases discretas. Por ello, a este tipo de distribuciones se las denominó árboles de regresión. Más adelante *J. R. Quinlan, propuso recurso novedoso basado en el sistema CART que se denominó M5, que construye modelos lineales a base de árboles* [85], por ello el autor los denominó ‘árboles de modelo’ (en inglés, ‘model trees’). El principal objetivo de este método es encontrar un árbol de regresión a partir de instancias de entrenamiento que sea capaz de predecir correctamente el valor numérico de la clase de una instancia futura [*Banzai-Inteligencia Artificial*].

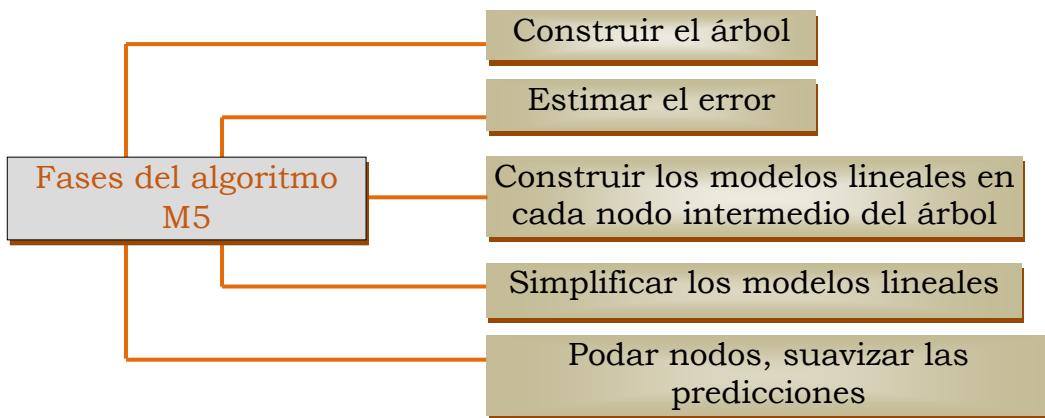
El árbol de decisión modelo M5, se emplea para tareas de regresión, con el fin de predecir los valores de la variable numérica de respuesta Y. Los árboles de predicción numérica son similares a los árboles de decisión, excepto en que la clase a predecir es continua y puede tomar valores a lo largo de un determinado intervalo. En este caso, cada nodo hoja almacena un valor de clase consistente en la media de las instancias que se clasifican con esa hoja, en cuyo caso estamos hablando de un árbol de regresión, o bien un modelo lineal que predice el valor de la clase, y se habla de árbol de modelos. En el caso del algoritmo M5, se trata de obtener un árbol de modelos, si bien se puede utilizar para obtener un árbol de regresión, por ser éste un caso específico de árbol de modelos. En un árbol de regresión, sus hojas predicen una cantidad numérica calculada como la media del valor para la variable clase de todos los ejemplos que han llegado a esa hoja durante el proceso de construcción del árbol. En cambio, las hojas de un árbol de modelos contienen una ecuación de regresión lineal local a esa partición del espacio de atributos.

M5 (Quinlan, 93) es una variación de CART que genera modelos lineales en las hojas, árbol de modelos, en lugar de valores numéricos [86]. El árbol de decisión CART es un método no paramétrico de segmentación binaria obtenido mediante la división repetida de los datos. El nodo inicial o nodo raíz es dividido en dos grupos ‘hijos’ o nodos, aplicando un procedimiento de partición a cada grupo ‘hijo’ por separado. En las divisiones se las impurezas de los hijos serán menores que la del grupo raíz y éstas vendrán definidas por el valor de una variable explicativa [Deconinck et al., 2006].

El árbol M5 selecciona el error cuadrático medio como función de impureza, pero no asigna una constante al nodo de la hoja, sino que es un modelo de regresión lineal multivariante; el árbol modelo es análogo a las funciones lineales por piezas. El árbol modelo M5, según J.R. Quinlan [87], puede ‘aprender’ eficazmente y manejar tareas con una dimensionalidad de hasta cientos de atributos. Esta capacidad sitúa al M5 al margen de los árboles de regresión en el tiempo (como en el caso de MARS), cuyo costo computacional crece muy rápidamente cuando el número de características aumenta. Otra gran ventaja del M5 sobre CART, es que los árboles modelo son habitualmente mucho más pequeños que los árboles de regresión y para determinadas tareas pueden ser más precisos [J. R. Quinlan]. El M5 permite el ajuste de forma automática y eficiente en casos de datos altamente no lineales y muy desestructurados.

4.2.4.1. Construcción del árbol M5

El árbol M5 divide el espacio de las instancias en grupos muy pequeños, lo cual genera un árbol excesivamente complejo que es preciso reducir o ‘podar’. Los datos se dividen en un conjunto T y dicho conjunto T está asociado con una hoja, o se elige algún ensayo que divida a T en subconjuntos correspondientes a los resultados de la prueba; aplicándose el mismo proceso de manera recursiva a los subconjuntos resultantes. Este proceso tal y como hemos dicho, produce a menudo estructuras excesivamente elaboradas, es decir, sobreajuste, que deben ser podadas.



La ganancia de información en el árbol M5, que es una medida de discriminación que indica el siguiente atributo a ser seleccionado para continuar con el proceso de división, se mide por el grado de reducción en la desviación estándar antes y después de la prueba. El primer paso es calcular la desviación estándar de los valores de respuesta de los casos en T. A menos que T contenga muy pocos casos o sus valores varíen sólo ligeramente, T está dividido en los resultados de una prueba. Siendo T_i el subconjunto de casos correspondientes al resultado de una prueba específica. Si tratamos la desviación $sd(T_i)$ de los valores objetivo de casos en T_i como una medida de error, la reducción esperada del error se puede escribir como sigue:

$$\Delta error = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (4. 80)$$

Entonces árbol M5 escogerá aquella que maximiza esta reducción esperada del error. Para comparar, CART elige un ensayo que proporcione la mayor reducción esperada en la varianza o la desviación absoluta [L. Breiman et al.]

4.2.4.2 Poda del árbol M5

La poda se realiza desde las hojas hasta el nodo raíz. En cada nodo interno, el árbol de regresión M5 compara el error estimado de ese nodo y el error estimado del subárbol a continuación. Seguidamente, si el rendimiento del árbol no mejora el subárbol se poda.

4.2.4.2.1. Estimación basada en errores

El árbol de modelos M5 emplea un método basado en la tasa de error de un árbol para efectuar la poda. En este método es factor clave la estimación error/precisión del modelo, en casos no vistos. M5 calcula este número haciendo primero un promedio de la diferencia absoluta entre los valores de respuesta de las observaciones y los valores predichos. Esto generalmente subestima el error en casos no vistos, por lo que M5 lo multiplica por $(n+v)/(n-v)$, donde n es el número de casos de entrenamiento y v es el número de parámetros en el modelo [J. R. Quinlan].

El efecto es aumentar el error estimado de los modelos con muchos parámetros construidos a partir de un pequeño número de casos. El error estimado de un subárbol se calcula como la suma de error estimado del árbol de la izquierda y la derecha por debajo de ese nodo, multiplicando con la proporción de muestras que baja hacia el árbol a la izquierda y la derecha.

El proceso de poda usa una estimación del error en cada nodo; siendo el error para cada nodo, la media del valor absoluto de la diferencia entre el valor predicho y el valor actual de cada ejemplo del conjunto de entrenamiento que alcanza dicho nodo. Esta media, estima por defecto el error para los ejemplos del conjunto test puesto que ha sido calculada a partir del conjunto de entrenamiento; por tanto, se multiplica por un factor de peso que depende del número de ejemplos del conjunto de entrenamiento que alcanzan ese nodo y el número de parámetros del modelo lineal obtenido en ese nodo. Una vez obtenido el modelo lineal para un nodo éste se puede simplificar eliminando términos mientras que el error estimado disminuya. Una vez simplificados los modelos lineales de cada nodo el árbol se poda desde las hojas. Un subárbol es podado siempre que su error sea mayor que el error obtenido en el nodo que debido a esta poda se convierte en hoja.

4.2.4.3. Suavizado

Cuando el valor de un caso es predicho por un árbol de modelo, el valor dado por el modelo en la hoja apropiada se ajusta para reflejar los valores predichos en los nodos a lo largo del camino desde la raíz a esa hoja. Se observó que la precisión de predicción de los modelos basados en árboles puede mejorarse mediante un proceso de suavizado [Daryl Pregibon. Private communications, 1992]. La forma y motivación del proceso de suavizado en M5, que está inspirada en Pregibon. Se describe como sigue:

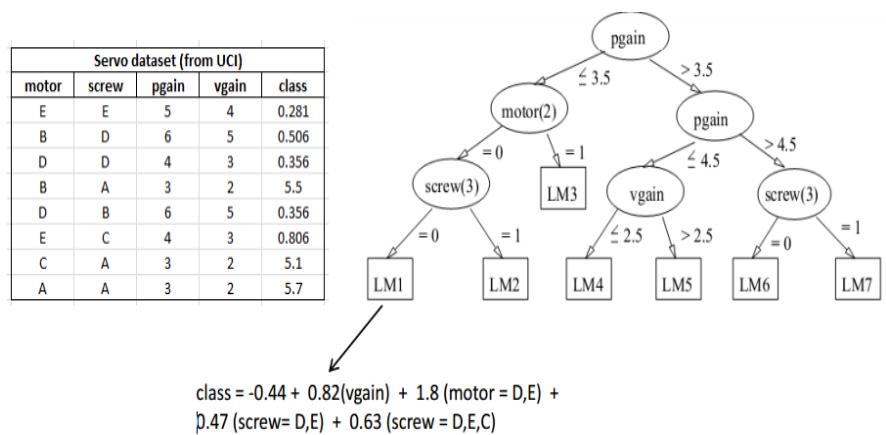
- El valor predicho en la hoja es el valor calculado por el modelo en esa hoja.
- Si el caso sigue a la rama Si del subárbol S_i , sea n_i el número de casos de entrenamiento en Si, $PV(S_i)$ el valor predicho en S_i , y $M(S)$ el valor dado por el modelo en S . El valor predicho respaldado por S es:

$$PV(S) = \frac{n_i \cdot PV(S_i) + k \cdot M(S)}{n_i + k} \quad (4.81)$$

Un proceso de suavizado resulta muy beneficioso para compensar las grandes discontinuidades que se originan entre modelos lineales adyacentes, en las hojas del árbol podado. Para ilustrar cómo funciona el árbol de modelos M5 en un problema real, se propone el ejemplo: *Servo dataset (UCI repository)*, que se muestra seguidamente. Hay 4 características que consisten en dos variables categóricas ‘motor’, ‘tornillo’ y dos variables numéricas ‘pgain’ y ‘vgain’. La variable ‘class’, toma valores reales cuyo rango va de 0.13 a 7.1. Una parte de este conjunto de datos y el árbol completo M5 para este conjunto de datos se representan en la Figura 4.19.

Figura 4.19. Ejecución del árbol de modelos M5 en *Servo dataset (from UCI)*

Servo dataset (from UCI)				
motor	screw	pgain	vgain	class
E	E	5	4	0.281
B	D	6	5	0.506
D	D	4	3	0.356
B	A	3	2	5.5
D	B	6	5	0.356
E	C	4	3	0.806
C	A	3	2	5.1
A	A	3	2	5.7



4.2.4.3.1. Modelos lineales

El proceso de suavizado necesita emplear un modelo lineal para cada nodo interior del árbol, no sólo en las hojas. Antes de podar el árbol se calcula un modelo lineal multivariado, que se ajusta a cada nodo del árbol usando la técnica de la regresión estándar. El árbol M5 no utiliza todas las características en el conjunto de datos, sino que se limita a las características que son referenciadas por modelos lineales en los subárboles por debajo de este nodo. M5 comparará la precisión de un modelo lineal con la precisión de un subárbol, lo que garantiza un sistema en el que los dos modelos usan la misma información [J. R. Quinlan et. al]. Después de construir el modelo lineal, el árbol M5 lo simplifica eliminando los coeficientes uno por uno. Esta forma podría aumentar el promedio residual, sin embargo, también reduce los factores multiplicativos de arriba, y el error estimado podría disminuir.

4.2.4.4. Procedimiento operativo seguido por el algoritmo M5

En términos de búsqueda, M5 realiza tres tipos de búsqueda:

La primera búsqueda genera un árbol de regresión a partir de los datos de entrada de forma análoga al algoritmo ID3 y calcula un modelo lineal, como en los modelos clásicos de regresión lineal, para cada nodo del árbol generado.

La segunda búsqueda intenta simplificar el árbol de regresión encontrado en el paso anterior (primera forma de post-poda), eliminando de los modelos lineales de los nodos aquellos atributos que no aumentan el error.

La tercera búsqueda tiene por objeto reducir el árbol, manteniendo el mismo poder predictivo (segunda post-poda).

4.2.4.4.1. Definición de la primera búsqueda.

Conjunto de estados: cada estado es un árbol de regresión, en el que los nodos intermedios son preguntas sobre valores de los atributos y todos los nodos también tienen un modelo asociado de regresión lineal (ecuaciones de hiperplanos con un subconjunto de los α atributos) que modelizan a los ejemplos que caen en esos nodos.

- *Conjunto de operadores*: el único operador es “introducir la pregunta por un atributo en un nodo”. Esto provoca la expansión de un nodo intermedio del árbol de regresión en construcción en varios sucesores, uno por cada posible valor del atributo. Este operador no se puede aplicar cuando en un nodo hay muy pocos ejemplos, o cuando se parecen mucho.
- *Estado inicial*: árbol de regresión vacío.
- *Meta*: árbol de regresión que separa los ejemplos de entrenamiento dependiendo de su clase, y tiene asociado en cada nodo un modelo lineal que calcula el valor de la clase dependiendo del valor de los atributos considerados.
- *Heurística* (búsqueda de la solución de un problema mediante métodos de tanteo o reglas empíricas): elegir aquel atributo que minimice la variación interna de los valores de la clase dentro de cada subconjunto. En concreto, se elige aquel atributo que maximice la reducción del error, de acuerdo a la siguiente ecuación:

El último paso de esta búsqueda consiste en crear un modelo lineal para cada nodo del árbol de regresión creado.

4.2.4.4.2. Descripción de la segunda búsqueda utilizando la técnica de escalada.

- *Conjunto de estados*: cada estado es un árbol de regresión.
- *Conjunto de operadores*: el operador es “eliminar un atributo de un modelo lineal de un nodo”. Esto provoca la simplificación de los modelos lineales de los nodos de un árbol de regresión.
- *Estado inicial*: árbol de regresión resultante de la primera búsqueda.
- *Meta*: árbol de regresión simplificado.
- *Heurística*: para decidir si quitar o no un atributo se debe calcular el error de clasificación esperado antes y después.

4.2.4.4.3. Descripción de la tercera búsqueda.

- *Conjunto de estados*: cada estado es un árbol de regresión.
- *Conjunto de operadores*: el operador es “eliminar un subárbol por debajo de un nodo”. Esto provoca la poda de un árbol de regresión.
- *Estado inicial*: árbol de regresión resultante de la segunda búsqueda.
- *Meta*: árbol de regresión que separa los ejemplos de entrenamiento dependiendo de su clase, y tiene asociado en cada nodo un modelo lineal que calcula el valor de la clase dependiendo del valor de los atributos considerados. Este árbol es el que se espera que tenga menor error de clasificación.
- *Heurística*: dado que cada nodo interno del árbol tiene un modelo simplificado lineal y un modelo descrito por su subárbol, se pueden realizar dos clasificaciones en cada nodo del árbol de regresión, y, por tanto, calcular un error de clasificación con las instancias de test: una dada por el modelo lineal y otra por la utilización del subárbol. La heurística dice que habrá que utilizar el operador de eliminar un subárbol en un nodo si el error de clasificación que se produce con el subárbol es mayor que el producido por el modelo lineal del nodo.

Para ser más eficientes, se pueden unir las dos últimas búsquedas en una sola de forma que se recorra el árbol de regresión una sola vez, simplificando los modelos lineales de cada nodo al mismo tiempo que se simplifican los nodos en sí.

Una visión general del procedimiento M5 aparece en la Tabla 4.3. Este procedimiento incluye las tres búsquedas descritas, representadas por los algoritmos *construir-árbol-regresión*, *simplificar-modelos-lineales* y *simplificar-árbol-regresión*. Llama en sucesión a estas funciones después de crear un nodo del árbol de regresión a través de la función *crea-nodo-árbol-regresión*.

Procedimiento M5 (E, A) E : conjunto de ejemplos A : conjunto de atributos con sus posibles valores $R := \text{crea-nodo-árbol-regresión}$ $R := \text{construir-árbol-regresión}(E, A, R)$ $R := \text{simplificar-modelos-lineales}(E, R)$ $R := \text{simplificar-árbol-regresión}(E, R)$ Devolver R **Tabla 4.3.** Procedimiento M5. Fuente: <http://banzai-deim.urv.net/>. Con acceso 12/2017.**4.2.5. Árbol modelo M5'**

El árbol modelo M5' fue propuesto por primera vez por Yong Wang e Ian H. Witten (1997). Dicho árbol M5', es una nueva implementación del inductor del árbol modelo basado en el árbol M5 [J. R. Quinlan, 1992], incluyendo modificaciones para reducir el tamaño del árbol, por lo que se vuelve más comprensible. M5' también introduce un modo de tratar con las variables categóricas y los valores que faltan. De acuerdo con Yong Wang e Ian H. Witten [88], M5' actúa algo mejor que el algoritmo original M5 en los conjuntos de datos estándar para los resultados publicados para clases continuas.

Para aclarar la forma de modelo lineal que no estaba clara en el árbol M5, *en primer lugar*, decir que M5' utiliza $k+1$ parámetros, donde k es el número de atributos en el conjunto de datos y 1 representa el término constante. *En segundo lugar*, durante el procedimiento de división inicial, M5' no divide un nodo si representa un número de ejemplos menor o igual que tres, o, cuando la desviación estándar de los valores de clase de los ejemplos en el nodo, es inferior al 5% de la desviación estándar de los valores de clase de todo el conjunto de datos. *En tercer lugar*, M5' decide dejar atributos, que se descartan de un modelo cuando su efecto es tan pequeño que en realidad aumenta el error estimado, en modelos de nivel superior si son útiles. *En cuarto lugar*, todos los atributos enumerados se transforman en variables binarias antes de construir un árbol de modelos.

Los valores medios de clase correspondientes a cada valor posible, se calculan a partir de los ejemplos de entrenamiento y los valores en la enumeración se ordenan de acuerdo con estos promedios. Entonces, si el atributo enumerado tiene k valores posibles, es reemplazado por $k-1$ atributos binarios sintéticos, siendo el i^{th} ‘0’ si el valor es uno de los primeros i en el orden y ‘1’ en caso contrario [Yong Wang and Ian H. Witten, 1997]. Para abordar el problema de que los atributos enumerados que tengan un gran número de valores distintos sean favorecidos de forma automática, M5' multiplica el valor de SDR por un factor que es la unidad, para una división binaria, y disminuye de forma exponencial a medida que el número de valores aumenta.

```

1 M5' ( examples )
2 {
3     SD = sd (examples)
4     for each k-valued enumerated attribute
5         convert into k-1 synthetic binary attributes
6     root = new_node
7     root.examples = examples
8     split (root)
9     prune (root)
10 }
11 split (node)
12 {
13     if sizeof (node.examples) < 4 or sd (node.examples) < 0.05 * SD
14         node.type = LEAF
15     else
16         node.type = INTERIOR
17         for each continuous and binary attribute
18             for all possible split positions
19                 calculate the attribute's SDR
20             node.attribute = attribute with max SDR
21             split (node.left)
22             split (node.right)
23     }
24     prune (node)
25     {
26         if node = INTERIOR then
27             prune (node.left_child)
28             prune (node.right_child)
29             node.model = linear_regression (node)
30             if subtree_error (node) > error (node) then
31                 node.type = LEAF
32     }

```

Tabla 4.4. Pseudocódigo del árbol modelo M5'. Hieu Chi Huynh, 2015

En términos de procesamiento de valores perdidos, M5' modifica la fórmula de SDR a:

$$SDR = \frac{m}{|T|} \times \beta(i) \times \left[std(T) - \sum_{j \in L, R} \frac{|T_j|}{|T|} \times sd(T_j) \right] \quad (4.82)$$

donde m es el número de instancias sin valores perdidos para ese atributo, (i) es el factor de corrección mencionado anteriormente, calculado para el atributo original al que corresponde este atributo sintético y T es el conjunto de ejemplos que alcanzan este nodo. TL y TR son los conjuntos que resultan de la división en este atributo. El pseudocódigo, o descripción compacta de alto nivel del principio operativo, del algoritmo del árbol modelo M5' es como se detalla en la Tabla 4.4.

4.2.6. M5' Rules

El algoritmo M5'Rules fue concebido por [Holmes et. al.] [\[89\]](#), y se ha implementado en el software Weka [Witten, Frank, Hall, 2011], utiliza las ideas básicas del algoritmo para datos nominales PART en la construcción de los conjuntos de reglas de clasificación y las aplica a M5. Se fundamenta en un modo de operar en el que se construye en cada interacción un árbol modelo donde sus nodos hojas son reglas, [Mikut y Cols, 2005]. Técnicamente genera un conjunto de reglas a partir del árbol modelo M5' siguiendo este proceso: se aplica un árbol de aprendizaje (árbol M5') al conjunto completo de datos de entrenamiento y se aprende un árbol podado. A continuación, la mejor hoja (de acuerdo con alguna heurística) se convierte en una regla y el árbol se descarta. Todas las instancias cubiertas por la regla se eliminan del conjunto de datos. El proceso se aplica de forma recursiva a las instancias restantes y termina cuando todas las instancias están cubiertas por una o más reglas. Ésta es la estrategia básica emplea el procedimiento de ‘divide y vencerás’ aplicado a las reglas de aprendizaje, evitando la generalización precipitada denominada sobre-poda [E. Frank & I. H. Witten] [\[90\]](#); no obstante, en lugar de construir una sola regla, construimos un árbol de modelos completo en cada etapa y convertimos su “mejor” hoja en una regla.

M5'Rules a diferencia del algoritmo para datos nominales PART, que emplea la misma estrategia para la predicción categórica, construye árboles completos en lugar de árboles parcialmente explorados.

La construcción de árboles parciales conduce a una mayor eficiencia computacional, y no afecta el tamaño y la precisión de las reglas resultantes. Las reglas se generan sobre la base de un modelo lineal sin arrugas. Para ilustrar este proceso de generación de reglas a partir de un árbol de modelos, la Figura 4.22 muestra un árbol elaborado por M5' y las reglas generadas por M5'Rules usando la heurística de cobertura para el conjunto de datos de pernos (Sta99). Sabemos que en cada etapa, M5'Rules elige su mejor nodo hoja y lo convierte en una regla.

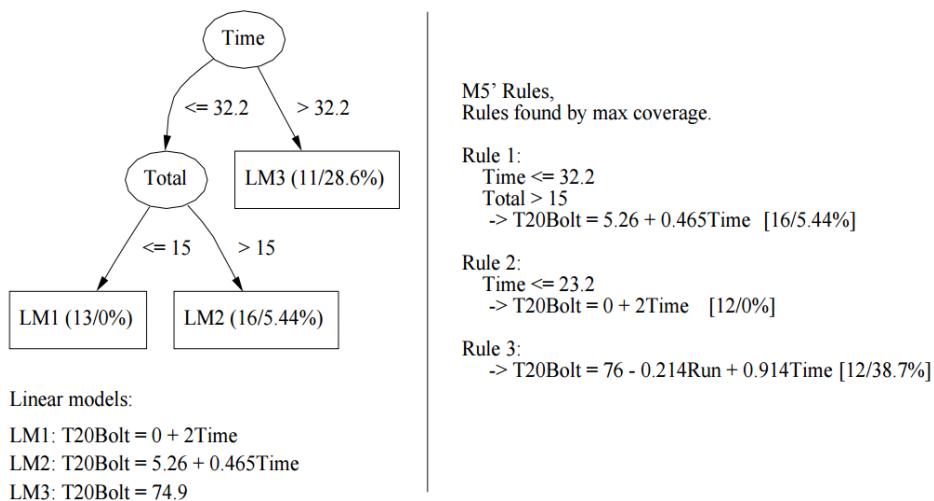


Figura 4. 20. Árbol modelo y reglas para el conjunto de datos de pernos. Fuente: Generating Rule Sets from Model Trees. Geoffrey Holmes, Mark Hall and Eibe Frank

Los valores en las hojas del árbol y por consiguiente de las reglas son la cobertura y el porcentaje de error cuadrático medio, respectivamente, para las instancias que alcanzan esas hojas y satisfacen las reglas. Se deberá tener en cuenta que la primera regla siempre se asignará directamente a una rama del árbol, pero las reglas posteriores a menudo no. En la Figura 4.22, la Regla 1 y LM2 son idénticas al igual que la Regla 2 y LM1, sin embargo, la Regla 3 y LM3 son muy diferentes.

Se han probado varias heurísticas, diseñadas para identificar reglas precisas y para compensar la precisión (número de veces que la regla es correcta) contra la cobertura (frecuencia de ocurrencia de la regla en los datos).

Una regla puede ser muy precisa, pero ocurrir muy pocas veces y ser poco relevante. La cobertura representa el porcentaje de veces que ocurre la regla, lo útil que es; es decir, el porcentaje de casos clasificados correctamente. Estas medidas son similares a las utilizadas en el procedimiento 'divide y vencerás' al evaluar la especialización de una regla de otra [Torgo 95] [91]. La pregunta es ¿cuáles son los criterios para comparar la "calidad" de los nodos y luego elegir la más informativa? En [Holmes et. al.], se han propuesto 4 heurísticas.

La primera y más obvia es *elegir la hoja que cubre la mayoría de los ejemplos*. En realidad, la Figura 4.22 utiliza esta heurística para extraer un conjunto de reglas.

La segunda calcula *el porcentaje de la raíz del error cuadrático medio*, como se muestra a continuación:

$$\% RMS = \frac{\sqrt{\sum_{i=1}^{N_r} (Y_i - y_i)^2 / N_r}}{\sqrt{\sum_{i=1}^{N_r} (Y_i - \bar{Y})^2 / N}} \quad (4. 83)$$

donde Y_i es el valor de clase real para el ejemplo i , y_i es el valor de clase predicho por el modelo lineal en una hoja, N_r es el número de ejemplos cubiertos por la hoja, \bar{Y} es la media de los valores de clase y N es el número total de ejemplos.

En este caso, los valores pequeños (menores de 1) de %RMS indican que el modelo en una hoja está haciendo mejor que la simple predicción de la media de los valores de la clase. Un problema potencial con el error cuadrático promedio de la raíz es que puede favorecer la precisión a expensas de la cobertura.

La tercera y cuarta, respectivamente, muestran dos medidas diseñadas para el tratamiento de la precisión contra la cobertura. El tercero, simplemente normaliza el error absoluto medio en una hoja, usando el número de ejemplos que cubre; la cuarta, multiplica la correlación entre los valores de clase pronosticados y reales, para instancias en una hoja, por el número de instancias que alcanzan la hoja.

$$MAE / Cover = \frac{\sum_{i=1}^{N_r} |Y_i - y_i|}{2N_r} \quad (4.84)$$

$$CC.Cover = \frac{\sum_{i=1}^{N_r} Y_i y_i}{N_r \sigma_Y \sigma_y} N_r \quad (4.85)$$

En [Holmes et al., 1999], también se implementa un experimento para poner a prueba el rendimiento de las 4 heurísticas enumeradas. Llegando a varias conclusiones, tales como que:

- M5'Rules, usando cobertura heurística, es mejor que el M5' en términos de precisión.
- M5'Rules no produce grandes conjuntos de reglas y produce conjuntos más pequeños en todos los conjuntos de datos analizados.
- En comparación con el sistema comercial CUBIST [Cubist, 2000], algoritmo de aprendizaje capaz de concluir categorías continuas, M5'Rules lo supera en tamaño y es comparable en precisión.
- Cuando se basa en el número de hojas, es más de tres veces más probable que produzca un número significativamente menor de reglas.
- El uso del algoritmo M5'Rules se considera una alternativa eficaz al análisis de datos, en cuanto a que utiliza sólidos conceptos de estadística y técnicas de aprendizaje automático bien conocidas.

4.2.6.1. Regresión lineal multivariante en el árbol modelo M5

Una de las propiedades más importantes de un árbol modelo M5 es que utiliza una regresión lineal multivariante en cada hoja, para predecir valores numéricos. Además, el modelo de regresión lineal también se construye en nodos internos en el proceso de poda. Por lo tanto, el rendimiento de un árbol modelo M5 se basa principalmente en lo buenos que sean esos modelos lineales. En esta sección de nuestro trabajo, nos gustaría presentar lo que es el modelo de regresión lineal, conocidos enfoques utilizados para resolver este problema y el planteamiento de nuestro enfoque en la aplicación del modelo de árbol M5. La mayoría de los conocimientos presentados en las siguientes secciones están elaboradas a partir de [William H. et al.] [92].

4.2.6.1.1. Modelo de regresión lineal

Para encontrar un *modelo de regresión lineal multivariante* para un conjunto de datos que tengan M instancias que contengan valores de un grupo de N atributos y valores de una clase, debemos resolver un conjunto de ecuaciones algebraicas lineales como ésta:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1N}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2N}x_N &= b_2 \\ a_{31}x_1 + a_{32}x_2 + \dots + a_{3N}x_N &= b_3 \\ &\dots \\ a_{M1}x_1 + a_{M2}x_2 + \dots + a_{MN}x_N &= b_M \end{aligned} \tag{4. 86}$$

Aquí las N incógnitas, x_j , $j = 1, 2, \dots, N$ están relacionadas por M ecuaciones. Los coeficientes a_{ij} con $i = 1, 2, \dots, M$ y $j = 1, 2, \dots, N$ son números conocidos recogidos de valores de atributos del conjunto de datos. Las cantidades de lado derecho b_i con $i = 1, 2, 3, \dots, M$ son de valores de clase. Si $N=M$ entonces hay tantas ecuaciones como se desconoce y hay una buena probabilidad de resolver para un conjunto de solución único de las ' x_j '.

Sin embargo, si una o más de las ecuaciones M es una combinación lineal de las otras (degeneración de fila), o si todas las ecuaciones contienen sólo ciertas variables exactamente en la misma combinación lineal (degeneración de columna), puede fallar en encontrar una única solución. Un conjunto de ecuaciones degenerado se llama singular [William H. et al.][33]. Este es un concepto importante en ecuaciones algebraicas lineales y se dispone de métodos de separación para tratar con esos casos. El conjunto de las ecuaciones lineales indicadas puede escribirse en forma de matriz como:

$$A \cdot x = b \tag{4. 87}$$

Donde A es la matriz de coeficientes y b es el lado derecho escrito como un vector de columna.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & & & \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_M \end{bmatrix} \quad (4. 88)$$

Para resolver el conjunto de ecuaciones lineales, una de las estrategias más simples es utilizar la estimación lineal por mínimos cuadrados. El conjunto de ecuaciones $M \times N$ puede escribirse como el conjunto $N \times N$.

$$(A^T \cdot A) \cdot x = (A^T \cdot b) \quad (4. 89)$$

El método de estimación de mínimos cuadrados lineales minimiza la suma de los residuos al cuadrado y conduce a una expresión de forma cerrada para el valor estimado de x desconocido. El problema ahora se convierte en la solución de un conjunto de ecuaciones $N \times N$ y esto representa una ventaja en lugar de $N \neq M$, ya que hay una buena probabilidad de resolver una solución de incógnitas única. Hay muchos métodos para resolver este tipo de problema. Aquí presentamos los tres más populares: eliminación Gauss-Jordan, eliminación Gausiana con eliminación hacia atrás (del inglés Back-substitution) y factorización o descomposición LU (del inglés *Lower-Upper*).

4.2.6.1.2. Eliminación Gauss-Jordan

Este método puede usarse para encontrar una inversión matricial y también para resolver conjuntos de ecuaciones lineales. Para una mejor ilustración del método, escribiremos ecuaciones para sólo el caso de cuatro ecuaciones y cuatro incógnitas y un vector del lado derecho de cuatro elementos.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad (4. 90)$$

La rutina de eliminación de Gauss-Jordan se procesa de la siguiente manera [William H. et al.] [33]. La primera fila se divide por el elemento a_{11} y ésta es una combinación lineal trivial de la primera fila con cualquier otra fila-coeficiente cero para la otra fila. Luego a cada otra fila se le resta una cantidad de primera fila multiplicando sus primeros elementos. Como resultado, la primera columna de A ahora está en concordancia con la matriz de identidad. Nos movemos a la segunda columna y dividimos la segunda fila por a_{22} , luego restamos la primera, tercera y cuarta filas de una cantidad de segunda fila multiplicando su segundo elemento, para hacer sus entradas en la segunda columna cero. Y así sucesivamente para las filas restantes y A se transforma en una matriz de identidad. Al hacer las operaciones en A, también hacemos las operaciones correspondientes de las b 's. En esta rutina, el elemento que dividimos por se llama pivote. Podemos intercambiar hileras (pivotamiento parcial) o hileras y columnas (pivotamiento completo) para colocar un elemento particularmente deseable en la posición diagonal a partir de la cual se va a seleccionar el pivote. Hay un problema obvio cuando uno de los elementos en diagonal, o pivote, es cero y no podemos dividir la fila por ella. Este es un caso de resolución de una matriz singular o no invertible.

En términos de encontrar la matriz inversa, sólo tenemos que aplicar la rutina anterior para resolver 4 ecuaciones

$$A \cdot x_1 = b_1 \quad A \cdot x_2 = b_2 \quad A \cdot x_3 = b_3 \quad A \cdot x_4 = b_4 \quad (4.91)$$

donde b_i , $i = 1, 2, 3, 4$ son columnas de matriz identidad. La matriz resultante está construida por x_i , $i = 1, 2, 3, 4$ como columnas.

4.2.6.1.3. Eliminación Gaussiana con Sustitución hacia atrás

Emplea una rutina similar a la eliminación de Gauss-Jordan, la única diferencia es que la matriz A no se reduce todo el camino a la matriz de identidad, sólo a medio camino.

En cada etapa, restamos las filas alejadas sólo por debajo del elemento de pivote actual. Cuando a_{22} es el elemento pivote, por ejemplo, dividimos la segunda fila por su valor (como antes), pero ahora usamos la fila pivote para cero sólo a_{32} y a_{42} , no a_{12} [William H. et al.] [33]. La ecuación después de la rutina de reducción será:

$$\begin{bmatrix} a'_{11} & a'_{12} & a'_{13} & a'_{14} \\ 0 & a'_{22} & a'_{23} & a'_{24} \\ 0 & 0 & a'_{33} & a'_{34} \\ 0 & 0 & 0 & a'_{44} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b'_1 \\ b'_2 \\ b'_3 \\ b'_4 \end{bmatrix} \quad (4.92)$$

Observe que aquí también se modifica la matriz original b , de acuerdo con las operaciones realizadas para convertir la matriz A en A' , en la ecuación anterior. La resolución de x ahora se hace bastante fácil siguiendo esta fórmula

$$x_4 = b'_4 / a'_{44} \quad (4.93)$$

$$x_3 = \frac{1}{a'_{33}} [b'_3 - x_4 a'_{34}] \quad (4.94)$$

a continuación, se procede x antes x_3 con,

$$x_i = \frac{1}{a'_{ii}} \left[b'_i - \sum_{j=i+1}^N a'_{ij} x_j \right] \quad (4.95)$$

El procedimiento definido se denomina sustitución hacia atrás. La combinación entre la *eliminación gaussiana* y la *sustitución hacia atrás* proporciona una solución para resolver el conjunto de ecuaciones lineales. La ventaja de este método sobre la eliminación Gauss-Jordan es la rapidez en el recuento de operaciones en bruto o crudas [William H. et al.][33].

4.2.6.1.4. Descomposición LU

Supongamos que la matriz A puede descomponerse en dos matrices, una matriz triangular inferior L y una matriz triangular superior U

$$L \cdot U = A \quad (4.96)$$

En el caso de la matriz 4×4 , la ecuación anterior tiene este aspecto:

$$\begin{bmatrix} \alpha_{11} & 0 & 0 & 0 \\ \alpha_{21} & \alpha_{22} & 0 & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & 0 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix} \cdot \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & \beta_{22} & \beta_{23} & \beta_{24} \\ 0 & 0 & \beta_{33} & \beta_{34} \\ 0 & 0 & 0 & \beta_{44} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad (4. 97)$$

Para resolver el problema lineal de las ecuaciones, se puede escribir como,

$$A \cdot x = (L \cdot U) \cdot x = L \cdot (U \cdot x) = b \quad (4. 98)$$

Primero resolvemos la ecuación,

$$L \cdot y = b \quad (4. 99)$$

Y después resolvemos,

$$U \cdot x = y \quad (4. 100)$$

Para resolver las ecuaciones, podemos usar el procedimiento descrito en la sección de eliminación gaussiana y sustitución posterior. El problema restante es resolver la descomposición y encontrar L y U.

$$i < j: \quad \alpha_{i1} \beta_{1j} + \alpha_{i2} \beta_{2j} + \dots + \alpha_{ii} \beta_{ij} = a_{ij} \quad (4. 101)$$

$$i = j: \quad \alpha_{i1} \beta_{1j} + \alpha_{i2} \beta_{2j} + \dots + \alpha_{ii} \beta_{ij} = a_{ij} \quad (4. 102)$$

$$i > j: \quad \alpha_{i1} \beta_{1j} + \alpha_{i2} \beta_{2j} + \dots + \alpha_{ii} \beta_{ij} = a_{ij} \quad (4. 103)$$

Debido a que existen ecuaciones $N^2 + N$ desconocidas, pero sólo N^2 en las ecuaciones de (11), (12) y (13), podemos asumir que

$$\alpha_{ii} = 1 \quad i = 1, 2, \dots, N$$

Entonces, usamos el algoritmo de Crout, mediante el cual se obtiene una matriz U cuyos elementos de la diagonal son todos 1, para resolver el conjunto de $N^2 + N$ ecuaciones, para todos los α 's y β 's sólo por organizar las ecuaciones en un cierto orden de la siguiente manera.

Establecemos $\alpha_{ii} = 1, \quad i = 1, 2, \dots, N$

Para cada $j=1, 2, 3, \dots, N$ hacemos dos procedimientos: Primero, para $i=1, 2, \dots, j$ resolvemos β_{ij}

$$\beta_{ij} = a_{ij} - \sum_{k=1}^{i-1} \alpha_{ik} \beta_{kj} \quad (4. 104)$$

Segundo, para $i=j+1, j+2, \dots, N$ resolvemos α_{ii}

$$\alpha_{ij} = \frac{1}{\beta_{ij}} (a_{ij} - \sum_{k=1}^{i-1} \alpha_{ik} \beta_{kj}) \quad (4. 105)$$

4.2.6.1.5. Resolver una matriz singular

Hay un método bien conocido para asumir el control de la matriz singular, es la descomposición en valores singulares: la matriz $A_{M \times N}$ cuyo número de filas M es mayor o igual que su número de columnas N, se puede escribir como el producto de una $M \times N$ matriz de columnas ortogonal U, y una matriz $N \times N$ diagonal producto de $M \times N$ y una matriz ortogonal V.

$$[A] = [U] \cdot \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \dots & \\ & & & \dots \\ & & & & w_N \end{bmatrix} \cdot [V^T]$$

Entonces

$$A^{-1} = V \cdot \left[\text{diag}(1/w_j) \right] \cdot U^T \quad (4. 106)$$

donde reemplazamos $1/w_j$ por 0, si w_j es casi 0. Así podemos resolver x como sigue:

$$x = V \cdot \left[\text{diag}(1/w_j) \right] \cdot (U^T \cdot B) \quad (4. 107)$$

Sin embargo, en la práctica, SVD no es un método óptimo porque requiere muchos recursos. En nuestra implementación, usamos el método de descomposición LU en una combinación con una matriz modificada A para resolver conjunto de ecuaciones lineales.

Cada vez que A se prueba como una matriz singular, una constante predeterminada, es decir, se añade una cresta a cada elemento en la diagonal de la matriz A para formar una nueva matriz que ya no es singular. La cresta es lo suficientemente pequeña para asegurar que el cambio efectuado en la matriz A es realmente trivial. Seguidamente se puede ver el ejemplo de pseudocódigo de matriz modificada para resolver ecuaciones lineales [Hieu Chi Huynh, 2015]

```

1 int success = 0;
2 double ridge = 10^(-8);
3 int i ,j ;
4 while (success == 0){
5     info = LAPACKE_dgesv( LAPACK_ROW_MAJOR,n, nrhs , &A[ 0 ] ,
6     lda , &ipiv [ 0 ] , &B[ 0 ] , ldb );
7     /* Check for the exact singularity */
8     if( info > 0 ) {
9         for ( i = 0; i < n; i++){
10             A [ i*n + i ] = A[ i*n + i ] + ridge ;
11         }
12         ridge = ridge * 10.0;
13     }
14     else
15         success = 1;
16 }
```

4.3. Técnicas de Optimización

En este epígrafe se pretende ilustrar que el desarrollo de un algoritmo que se construya en base a la solución de un problema puede mejorarse muchísimo, analizando aspectos como el gasto de tiempo y los recursos de ejecución. Este trabajo de tesis propone el uso de una técnica de optimización por nube o enjambre de partículas, que hace referencia a una metaheurística que evoca el comportamiento de las partículas en la naturaleza. Dicha técnica permite optimizar un problema a partir de una población de soluciones candidatas, denominadas "partículas", moviendo éstas por todo el espacio de búsqueda según reglas matemáticas que tienen en cuenta la posición y la velocidad de las partículas.

El movimiento de cada partícula se ve influido por su mejor posición local hallada hasta el momento, así como por las mejores posiciones globales encontradas por otras partículas a medida que recorren el espacio de búsqueda.

4.3.1. Clasificación de las técnicas de optimización

Los problemas de optimización han tenido gran importancia a lo largo de la historia de la Informática, por lo que se han desarrollado múltiples métodos para tratar de resolverlos. En la Figura 4.21 se muestra una clasificación de las técnicas empleadas para abordar estos problemas de optimización. Las técnicas empleadas se pueden clasificar en dos grupos: *exactas* o enumerativas y *aproximadas*. Las *técnicas exactas* garantizan encontrar la solución óptima para cualquier instancia de cualquier problema en un tiempo acotado. El inconveniente de estos métodos es que el tiempo necesario para llevarlos a cabo, aunque está limitado crece exponencialmente con el tamaño del problema. Esto provoca en muchos casos que el tiempo necesario para la resolución del problema sea inabordable. Los algoritmos basados en *técnicas de optimización aproximadas* sacrifican la garantía de encontrar la solución ‘óptima’ a cambio de encontrar una ‘buena’ solución en un tiempo ‘razonable’. Estos últimos, vienen recibiendo una atención especial en los últimos años.

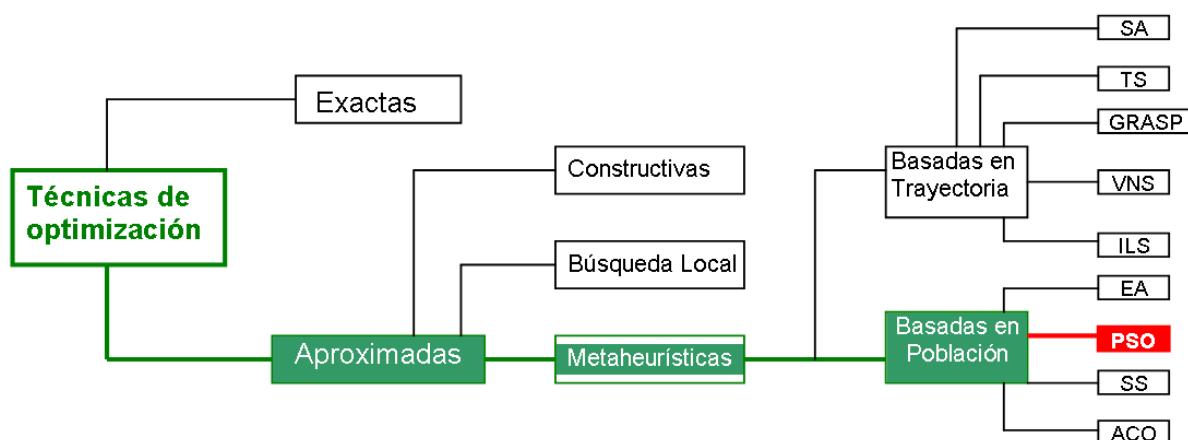


Figura 4.21. Clasificación de las técnicas de optimización. Elaboración propia.

Entre los mencionados algoritmos basados en técnicas de optimización no exactas, se encuentran los tres tipos que se describen a continuación:

- Heurísticos constructivos o voraces
- Métodos de búsqueda local
- Los metaheurísticos

4.3.1.1. Heurísticos constructivos o voraces

Los algoritmos heurísticos, generan una solución partiendo de una propuesta vacía a la que se le van añadiendo componentes, hasta tener una solución completa que sería la deducción del algoritmo. Estos suelen ser los métodos más rápidos, pero en cambio las soluciones ofrecidas son de muy baja calidad, resultando difícil encontrar métodos que produzcan buenas soluciones dado que es necesario tener un conocimiento muy extenso del problema.

En muchos supuestos es casi imposible disponer de la información suficiente, p. ej., en aquellos casos con muchas restricciones, donde la mayoría de las soluciones parciales sólo conducen a soluciones no factibles.

4.3.1.2. Métodos de búsqueda local o de seguimiento del gradiente

Parten de la solución completa teniendo presente el concepto de ‘vecindario’, que notamos como $N(s)$, y que es el conjunto de soluciones que se pueden construir a partir de (s) aplicando un operador específico de modificación; y recorren una parte del espacio de búsqueda hasta encontrar un ‘óptimo local’, que es una solución mejor o igual a cualquier otra solución de su vecindario. Se parte de una solución inicial y se examina el ‘vecindario’ eligiendo el ‘mejor vecino’, continuando el proceso hasta que encuentran un ‘óptimo local’. En muchos casos, la exploración completa del vecindario es inabordable, dando lugar a diferentes variaciones del esquema genérico.

4.3.1.3. Los metaheurísticos

En los años setenta surgió una nueva clase de algoritmos denominados metaheurísticos, no exactos, que combinan a un nivel más alto distintos métodos heurísticos, consiguiendo explorar el espacio de búsqueda de una manera más eficaz. El término metaheurístico fue introducido por primera vez por Glover [93]. Estos métodos eran denominados anteriormente como heurísticos modernos [94], cuando todavía no habían sido completamente aceptados en el ámbito científico. Se definen a continuación las propiedades fundamentales que caracterizan el método basado en la metahurística:

- Su planteamiento básico es general, no dependiendo del problema a resolver.
- Son algoritmos no exactos y generalmente no deterministas, que podrían incorporar elementos que eviten áreas no óptimas del espacio de búsqueda.
- Son estrategias o plantillas generales que ‘guían’ el proceso, explorando del espacio de búsqueda para encontrar soluciones (casi) óptimas.
- Emplean funciones de bondad, como las funciones de ‘fitness’, para cuantificar el grado de adecuación de una determinada solución. La función de aptitud (fitness) es la que permite valorar la aptitud de los individuos y debe tomar siempre valores positivos.

En general, una metaheurística es una estrategia de alto nivel que emplea diferentes métodos para explorar el espacio de búsqueda. Se trata por tanto de una plantilla general no determinista, que nos permite abordar problemas con espacios de búsqueda de gran tamaño, que debe ser rellenada con datos específicos del problema.

Según T. Cranic y M. Toulouse, existen distintas maneras de clasificar y describir las técnicas metaheurísticas [T. Crainic and M. Toulouse, 2003.] [95]. Así, según las características seleccionadas, se podrían obtener taxonomías basadas en la naturaleza o no basadas en la naturaleza, basadas en memoria o sin memoria, con función objetivo estática o dinámica, etc.

Según el esquema de la Figura 4.21, tendremos metaheurísticas:

Basadas en la trayectoria, si en cada paso manipulan un único punto del espacio de búsqueda.

Basadas en la población, si en cada paso manipulan un único punto del espacio de búsqueda o trabajan sobre un conjunto, o ‘población’, de ellos.

<p>El Enfriamiento Simulado o Simulated Annealing (SA). S. Kirkpatrick, C. Gelatt, and M. Vecchi. <i>Optimization by Simulated Annealing. Science</i>, 220(4598):671-680, 1983.</p>	<p>Permite elegir una solución peor que la solución actual, para evitar quedar atrapado en un óptimo local. Se calcula siguiendo la distribución de Boltzmann</p>
<p>La Búsqueda Tabú o Tabu Search (TS). F. Glover. <i>Future Paths for Integer Programming and Links to Artificial Intelligence. Computers & Operations Research</i>, 13:533-549, 1986.</p>	<p>Emplea explícitamente el historial de la búsqueda (memoria de corto plazo), tanto para escapar de los óptimos locales como para implementar su estrategia de exploración y evitar buscar varias veces en la misma región. En cada iteración se elige la mejor solución entre las permitidas y la solución es añadida a la lista tabú.</p>
<p>El Procedimiento de Búsqueda Miope Aleatorizado y Adaptativo o The Greedy Randomized Adaptive Search Procedure (GRASP). T. Feo and M. Resende. <i>Greedy Randomized Adaptive Search Procedures. Journal of Global Optimization</i>, 6:109-133, 1999.</p>	<p>Es un procedimiento iterativo compuesto de dos fases: primero una construcción de una solución y después un proceso de mejora, que combina heurísticos constructivos con búsqueda local. La solución mejorada es el resultado del proceso de búsqueda.</p>
<p>La Búsqueda en Vecindario Variable o Variable Neighborhood Search (VNS). N. Mladenovic and P. Hansen. <i>Variable Neighborhood Search. Computers Oper. Res.</i>, 24:10971100, 1997.</p>	<p>Aplica explícitamente una estrategia para cambiar entre diferentes estructuras de vecindario de entre un conjunto de ellas definidas al inicio del algoritmo.</p>
<p>La Búsqueda Local Iterada o Iterated Local Search (ILS). T. Stützle. <i>Local Search Algorithms for Combinatorial Problems Analysis, Algorithms and New Applications. Technical report, DISKI Dissertationen zur Künstlichen Intelligenz. Sankt Augustin, Germany</i>, 1999.</p>	<p>En cada iteración, la solución actual es perturbada y a esta nueva solución se le aplica un método de búsqueda local para mejorarlala. El nuevo óptimo local obtenido por el método de mejora es aceptado como nueva solución actual si pasa un test de aceptación.</p>

Tabla 4.5. Características de las metaheurísticas basadas en trayectoria. Elaboración propia.

<p>Los Algoritmos Evolutivos o Evolutionary Algorithms (EA). <i>T. Bäck, D. Fogel, and Z. Michalewicz. Handbook of Evolutionary Computation. IOP Publishing and Oxford University Press, New York and Bristol (UK), Feb 1997.</i></p>	<p>Siguen un proceso iterativo y estocástico (aleatorio) que opera sobre una población de individuos, los cuales de forma individual, representan una solución potencial al problema que se está resolviendo. Cada individuo en la población tiene asignado, por medio de una función de aptitud denominada ‘fitness’, una medida de su bondad con respecto al problema considerado. Dicho valor es la información cuantitativa que el algoritmo usa para guiar su búsqueda.</p>
<p>La Búsqueda Dispersa o Scatter Search (SS). <i>F. Glover and G. Kochenberger. Handbook of Metaheuristics. Kluwer Academic Publishers, Norwell, MA, 2002.</i></p>	<p>Se basa en mantener un conjunto de soluciones tentativas relativamente pequeño denominado ‘conjunto de referencia’, que se caracteriza tanto por contener buenas soluciones como soluciones diversas. Este conjunto se divide en subconjuntos de soluciones a las cuales se les aplica una operación de recombinación y mejora.</p>
<p>Los sistemas basados en Colonias de Hormigas o Ant Colony Optimization (ACO). <i>M. Dorigo. Optimization, Learning and Natural Algorithms. PhD thesis, Dipartamento di Elettronica, Politecnico di Milano, 1992.</i></p>	<p>Se basan en que las hormigas exploran el área cercana a su nido de forma aleatoria. Cuando la hormiga encuentra comida la lleva al nido, y mientras recorre este camino va depositando una sustancia química denominada feromonas. Dicha sustancia ayudará al resto de las hormigas a encontrar la comida. La comunicación indirecta, mediante el rastro de feromonas, entre las hormigas las faculta para encontrar el camino más corto entre el nido y la comida. En el método para resolver problemas de optimización, el equivalente al rastro de feromonas es simulado mediante un modelo probabilístico.</p>
<p>Los Algoritmos Basados en Cúmulos de Partículas o Particle Swarm Optimization (PSO). <i>J. Kennedy, R. Eberhart, and Y. Shi. Swarm Intelligence. San Francisco: Morgan Kaufmann Publishers, 2001.</i></p>	<p>Se fundamentan en los factores que influyen en la toma de decisión de un agente que forma parte de un conjunto de agentes similares. La toma de decisión por parte de cada agente se realiza conforme a una componente social y una componente individual, mediante las que se determina el movimiento o dirección de este agente, para alcanzar una nueva posición en el espacio de soluciones.</p>

Tabla 4.6. Características de las metaheurísticas basadas en población. Elaboración propia.

METAHEURÍSTICAS BASADAS EN TRAYECTORIA

Estos algoritmos manan de extensiones de los métodos de búsqueda local simples a los que se les va añadiendo alguna característica para evitar los mínimos locales, lo cual implica que exista una condición de parada más compleja que encontrar un mínimo local.

Los procedimientos seguidos, parten de un punto y van actualizando la situación actual mediante la exploración del vecindario, formando así una trayectoria. La búsqueda finaliza tras alcanzar un número máximo predefinido de iteraciones, o encontrar una solución aceptable, o se detecta que el proceso se paraliza. En la Tabla 4.5 se presentan las características de las metaheurísticas basadas en trayectoria.

METAHEURÍSTICAS BASADAS EN POBLACIÓN

Los métodos metaheurísticos basados en población, a diferencia de los basados en trayectoria, se caracterizan por trabajar con un conjunto de soluciones, o población, en cada iteración. Su resultado final depende principalmente de la forma en que se manipula la población. En la Tabla 4.6 se presentan las características de las metaheurísticas basadas en población.

4.3.2. Optimización por Enjambre de Partículas (PSO)

La técnica de Optimización por Enjambre de Partículas (en inglés, Particle Swarm Optimization, (PSO)) se utiliza en espacios de búsqueda con muchas dimensiones, con el fin de facilitar la visualización. Este método definido por James Kennedy y Russell C. Eberhart [96], está inspirado en el comportamiento de los enjambres de insectos, tales como p. ej., un enjambre de abejas que para encontrar el polen buscan un área en el que exista más densidad de flores, porque es más probable encontrar polen en dicho espacio. Según los autores, "el algoritmo de enjambre de partículas imita el comportamiento social, humano o el de los insectos, donde los individuos interactúan unos con otros mientras aprenden de su propia experiencia, y, poco a poco los miembros de la población se mueven en mejores regiones del espacio del problema". Esta idea ha sido trasladada al campo de la computación en forma de algoritmo y se emplea actualmente para la mejora de los distintos tipos de sistemas. La inteligencia del enjambre es una técnica de inteligencia artificial basada en el estudio del comportamiento colectivo en sistemas descentralizados y auto-organizados. Estos sistemas se componen típicamente de una población de agentes simples que interactúan localmente entre sí y con su entorno. Aunque normalmente no existe una estructura de control centralizada que dicte cómo deben comportarse los agentes individuales, las interacciones locales entre estos agentes a menudo conducen a la aparición de un comportamiento global (de Wikipedia).

El algoritmo PSO se fundamenta en el enfoque denominado ‘metáfora social’ [J. Kennedy et. al] [97], que describe el algoritmo y se reduce a que los individuos que conviven en una sociedad tienen una opinión que es parte de un ‘conjunto de creencias’ y constituye el espacio de búsqueda compartido por todos los individuos, pudiendo éstos modificar su propia opinión basándose en tres factores:

- Conocimiento del entorno (valor de ‘fitness’).
- Conocimiento basado en experiencias anteriores (memoria).
- Conocimiento de experiencias anteriores de los individuos del vecindario.



Figura 4.22. Comportamiento para el aprendizaje de los elementos o partículas que constituyen el enjambre. Fuente: Rommy Meyer. Swarm Intelligence Group, Peking University.

Los elementos que constituyen el ‘enjambre’ se denominan ‘partículas’, y no ‘puntos’, porque las velocidades y las aceleraciones se aplican con más propiedad a las ‘partículas’ [Kennedy y Eberhart, 1995].

4.3.2.1. El paisaje de fitness

Sea $f(x, y)$, una función desconocida, cuya expresión se desconoce, que podemos evaluar en los espacios que queramos a modo de caja negra. El objetivo consistirá en encontrar unos valores de (x) e (y) , para los cuales la función $f(x, y)$ sea máxima o mínima, o verifique una determinada relación respecto a alguna otra función. A la función $f(x, y)$ se la denomina ‘paisaje de fitness’, asemejándola a un paisaje de cerros y valles formados por los valores que toma la función. Dicha función deberá establecer lo buena que es la posición actual para cada partícula. Situando las partículas al azar en el espacio de búsqueda, y dándoles la posibilidad de que se muevan a través de dicho espacio, según unas reglas que tengan en cuenta el conocimiento personal de cada partícula y el conocimiento global del enjambre, se puede comprobar que proporcionándoles capacidad de movimiento a través del citado ‘paisaje de fitness’ y comunicación entre ellas, pueden llegar a descubrir valores particularmente altos para $f(x, y)$ empleando relativamente pocos recursos de tiempo, cálculo y memoria.

El paisaje de ‘fitness’, se obtiene asignando aleatoriamente valores a cada uno de los puntos de la malla, para posteriormente aplicar un proceso de difusión que suaviza los valores obtenidos, lo que proporciona un espacio con numerosos mínimos locales o ‘valles’ y máximos locales o ‘colinas’, por lo que su optimización por medios más clásicos, como el ascenso de la colina, es difícil.

4.3.2.2. Enfoque basado en PSO. Modo de operación

Cada partícula ocupa una posición, p (que en 2 dimensiones vendrá determinado por un vector de la forma (x, y)), en el espacio de búsqueda y una velocidad, v (que en 2 dimensiones vendrá determinado por un vector de la forma (v_x, v_y)), con la que se mueve a través del espacio. Además, como el resto de partículas físicas, tienen una cantidad de inercia, que los mantiene en la misma dirección en la que se movían, así como una aceleración (cambio de velocidad), que depende principalmente de dos características:

1. Cada partícula es atraída hacia la mejor localización que ella, personalmente, ha encontrado en su historia ‘mejor personal’.
2. Cada partícula es atraída hacia la mejor localización que ha sido encontrada por el conjunto de partículas en el espacio de búsqueda ‘mejor global’.

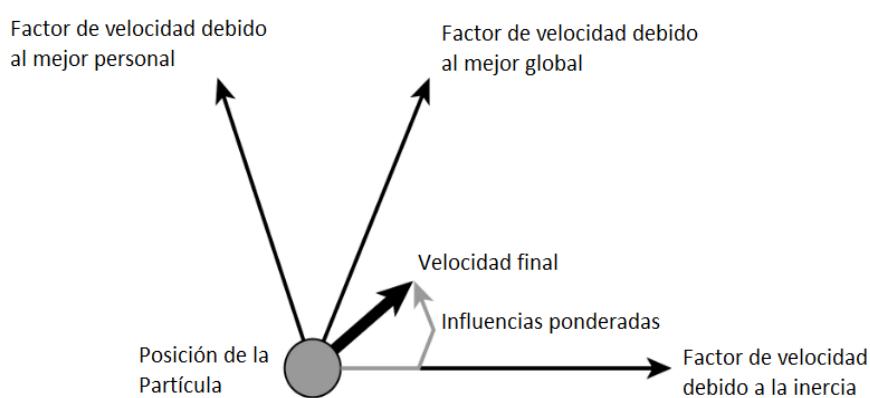


Figura 4.23. Dirección y fuerza que empuja a la partícula. Fuente: F. Sancho - Dpto. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla

La fuerza con que las partículas son empujadas en cada una de estas direcciones depende de dos parámetros que pueden ajustarse (*atracción-al-mejor-personal* y *atracción-al-mejor-global*), de modo que a medida que las partículas se alejan de estas localizaciones mejores, la fuerza de atracción es mayor. También se suele incluir un factor aleatorio que influye en cómo las partículas son empujadas hacia estas localizaciones.

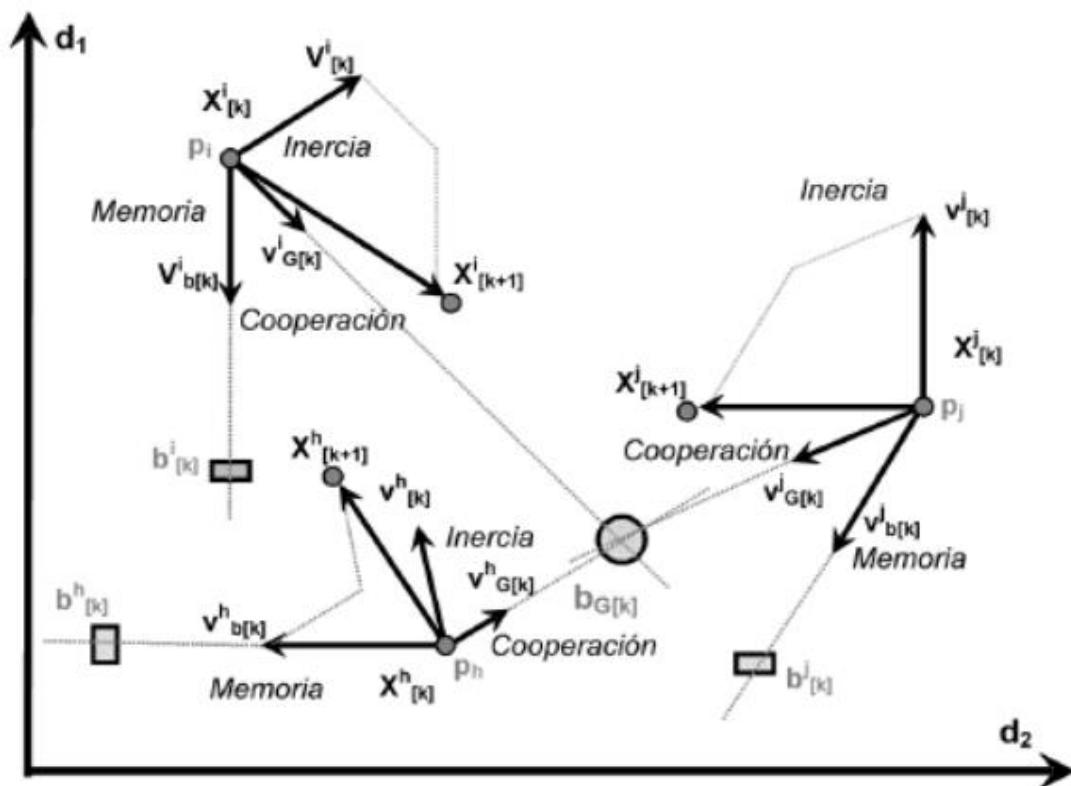


Figura 4.24. Direcciones y fuerzas de empuje de varias partículas. Fuente: F. Sancho - Dpto. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla

4.3.2.3. 'Posición actual' y 'Posición siguiente'

Un ave voladora cuando va en busca de alimento tiene una posición y una velocidad determinada que, en cualquier momento, el pájaro cambia de posición ajustando dicha velocidad a las necesidades puntuales. La velocidad cambia en base a su experiencia pasada y también a la retroalimentación recibida de sus vecinos. Este proceso de búsqueda puede ser artificialmente simulado para resolver problemas de optimización no lineal.

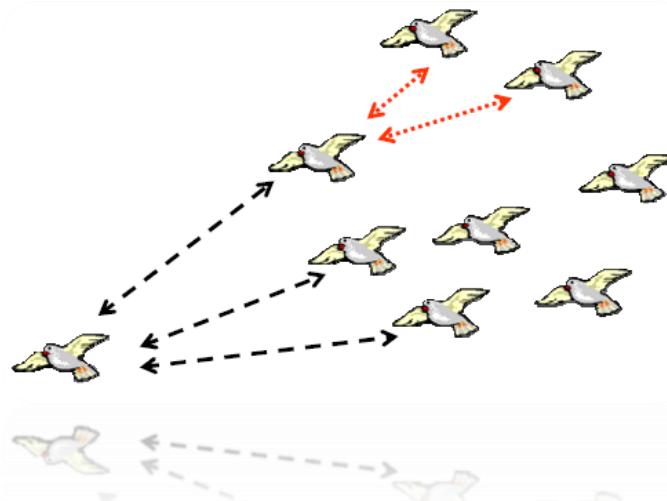


Figura 4.25. Cambio de velocidad y posición de la partícula debido a la realimentación del vecindario. Fuente: Xiaodong Li School of Computer Science and IT, RMIT University, Melbourne SEAL'06.

A continuación, se representa un código simplificado del algoritmo de estudio. En este modelo, cada partícula realiza sólo operaciones vectoriales básicas, por lo cual para actualizar las velocidades y posiciones bastaría realizar algunas sumas de vectores, y como la experiencia muestra que hacen falta pocas partículas y el número de iteraciones es bajo, la complejidad del proceso se centra sobre todo en la evaluación del ‘fitness’ de cada partícula en cada iteración, algo que no puede evitarse y que depende exclusivamente del problema específico que se quiera optimizar.

Asignar posiciones y velocidades aleatorias iniciales a las partículas

Repetir

Cada partícula:

Actualizar su velocidad considerando:

Inercia de la partícula (la hace seguir con la misma velocidad)

Atracción al mejor personal

Atracción al mejor global

Actualizar la posición de la partícula

Calcular el valor de fitness en la nueva posición

Actualizar su mejor personal

Actualizar el mejor global del sistema

Devolver el mejor global

Tabla 4.7. Pseudocódigo a grandes rasgos del algoritmo PSO. Fuente: F. Sancho – Dpto. de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla.

Este método de optimización es especialmente adecuado cuando el coste de la función de ‘fitness’ es muy elevado, pues requiere pocas evaluaciones para encontrar valores cercanos al óptimo, lo que resulta una ventaja. Como inconveniente, no podemos estar seguros de alcanzar el óptimo global, como pasa en todo este tipo de metaheurísticas, siendo el principal problema que comparten todos los métodos de optimización que trabajan con cajas negras.

El algoritmo PSO resulta adecuado para trabajar con datos dinámicos, es decir, debido a la forma en que funciona el algoritmo, si la función que se desea optimizar cambia en el tiempo, y esta dinámica verifica algunas propiedades de continuidad tales como que el cambio efectuado en dicho tiempo no sea muy brusco, entonces el mismo algoritmo hace que en cada iteración las partículas vayan evaluando el valor presente de la función a optimizar, por lo que pueden ir adaptando sus trayectorias a la situación real de la función. Algunos algoritmos que se emplean para optimización considerarían la función en un instante determinado, intentarían optimizarla para ese instante, y tras ejecutar el algoritmo la función ya habría cambiado. Así, dependiendo del coste en tiempo del algoritmo de optimización usado, podemos encontrarnos que el valor devuelto como óptimo poco tiene que ver con el valor actual de la función; porque en el transcurso de su ejecución ha ido cambiando. En este sentido, el PSO funciona bien desde el punto de vista de optimización de funciones dinámicas.

En relación con lo anterior, no se puede afirmar que el PSO permita predecir los valores óptimos futuros con carácter general. Para realizar predicciones quizás se debería mezclar, con algún tipo de jerarquización, PSO con algoritmos de aprendizaje, o en general de dinámica de sistemas, para aprovechar simultáneamente la capacidad optimizadora de PSO con la capacidad predictora de algún otro método [F. Sancho – Dpto. Inteligencia Artificial, Universidad de Sevilla].

4.3.2.4. Parámetros del modelo PSO

Se realiza un conjunto ordenado de operaciones sistemáticas, que constituye un proceso aleatorio e iterativo, que opera sobre el enjambre de partículas; representando la posición de cada partícula una solución potencial al problema. Cada partícula del enjambre, p_i , se compone de tres vectores y dos valores de ‘fitness’:

- El vector $x_i = \langle x_{i1}, x_{i2}, \dots, x_{in} \rangle$ almacena la *posición actual* o localización de la partícula en el espacio de búsqueda.
- El vector $pBest_i = \langle p_{i1}, p_{i2}, \dots, p_{in} \rangle$ almacena la posición de la *mejor solución encontrada* por la partícula hasta el momento.
- El vector de velocidad $v_i = \langle v_{i1}, v_{i2}, \dots, v_{in} \rangle$ almacena la dirección o gradiente, según el cual se moverá la partícula.
- El valor de ‘fitness’ $fitness_x_i$ almacena el valor de adecuación de la solución actual (vector x_i).
- El valor de ‘fitness’ $fitness_pBest_i$ almacena el valor de adecuación de la mejor solución local encontrada hasta el momento (vector $pBest_i$).

El enjambre, en el que las partículas son agentes simples que se mueven por el espacio de búsqueda guardando y comunicando la mejor solución que han encontrado, es un sistema multi-agente (SMA) que se inicializa generando las posiciones y las velocidades iniciales de las partículas. Las posiciones se pueden generar aleatoriamente en el espacio de búsqueda mediante una heurística de construcción, de forma regular o con una combinación de ambas. Cada partícula tiene un ‘fitness’, una ‘posición’ y un ‘vector velocidad’ que dirige su corriente, estando el movimiento de las partículas guiado por el espacio mediante las partículas óptimas en el momento actual. Cada partícula se desplaza, mediante un proceso iterativo, desde una posición del espacio de búsqueda hasta otra, añadiendo el vector velocidad v_i al vector de posición x_i para obtener así un nuevo vector de posición.

Una vez generadas las posiciones, se calcula el ‘fitness’ de cada una y se actualizan los valores de $fitness_x_i$ y $fitness_pBest_i$. Una vez calculada la nueva posición de la partícula, se evalúa actualizando el $fitness_x_i$ y si el nuevo ‘fitness’ es el mejor encontrado hasta el momento, se actualizan los valores de mejor posición $pBest_i$ y $fitness_pBest_i$. El vector velocidad de cada partícula es modificado en cada iteración utilizando la velocidad anterior, una componente cognitiva y una componente social. El modelo matemático resultante del algoritmo PSO viene expresado por las siguientes ecuaciones:

$$v_i^{k+1} = w \cdot v_i^k + \varphi_1 \cdot rand_1 \cdot (pBest_i - x_i^k) + \varphi_2 \cdot rand_2 \cdot (g_i - x_i^k) \quad (4. 108)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (4. 109)$$

v_i^{k+1} : velocidad de la partícula i en la iteración $k+1$,

v_i^k : velocidad de la partícula i en la iteración k ,

w : factor de inercia,

φ_1 , φ_2 : pesos o ratios de aprendizaje que definen las componentes *cognitiva* y *social*,

$rand_1$ $rand_2$: son números aleatorios comprendidos entre 0 y 1,

x_i^k : posición de la partícula i en la iteración k ,

$pBest_i$: mejor posición de la partícula i encontrada hasta el momento,

g_i : posición de la partícula con el mejor $fitness_pBest_i$ del entorno de p_i ‘mejor local’ o de todo el enjambre ‘mejor global’.

La ecuación (4.108) expresa el reajuste del vector velocidad de cada partícula i en cada iteración k . La *componente cognitiva* viene dada por el factor $\varphi_1 \cdot rand_1 \cdot (pBest_i - x_i^k)$, que representa la distancia entre las posiciones ‘actual’ y la que tomará, influenciada por su propia experiencia, por ser la ‘mejorconocida’ por esa partícula. La *componente social* se modela mediante $\varphi_2 \cdot rand_2 \cdot (g_i - x_i^k)$, y define la distancia entre la ‘posición actual’ y la decisión que tomará la partícula según la influencia que el resto del enjambre ejerce sobre ella, adoptando así la ‘mejor posición del vecindario’.

En la expresión siguiente (4.110), la velocidad v_{t+1} , expresa el valor del cambio, que es una función de la diferencia entre las individuales ‘mejor personal’ p_i , y su posición actual y la diferencia entre ‘lo mejor del vecindario’ p_g y su posición actual [Kennedy & Eberhart, 2001]. Obsérvese que el símbolo ‘ \otimes ’ expresa un producto vectorial.

$$v_{t+1} = v_t + R_1 \left[0, \frac{\varphi}{2} \right] \otimes (p_i - x_t) + R_2 \left[0, \frac{\varphi}{2} \right] \otimes (p_g - x_t) \quad (4.110)$$

Donde los factores de la expresión expresan el concepto que se indica:



Puesto que $\varphi_1 = R_1 \left[0, \frac{\varphi}{2} \right]$ y $\varphi_2 = R_2 \left[0, \frac{\varphi}{2} \right]$ dan vectores de números aleatorios dentro de un rango especificado, la partícula recorrerá desigualmente alrededor de un punto definido como el promedio ponderado de p_i y p_g :

$$\frac{\varphi_1 \otimes p_i + \varphi_2 \otimes p_g}{\varphi_1 + \varphi_2}$$

Seguidamente se escribe el pseudocódigo de un PSO básico:

```

Generar aleatoriamente una población inicial
reiterar
    para i = 1 hasta generar el tamaño de población
        si  $f(x_i) < f(p_i)$  entonces  $p_i = x_i$ ;
         $p_g = \min(p_{vecino})$ ;
    para d=1 hasta generar dimensiones
        actualizar_velocidad ();
        actualizar_posición ();
    fin
fin
hasta que se cumpla el criterio de terminación.
```

4.3.2.4.1. Peso de inercia PSO

Para controlar aún más la búsqueda, Shi y Eberhart (1998) propusieron utilizar un parámetro de ‘peso de inercia’. En este trabajo elaborado por los investigadores, se introduce un nuevo parámetro, denominado peso inercial, en el optimizador del enjambre de partículas original. Para ello, realizaron simulaciones para ilustrar el impacto significativo y efectivo de este nuevo parámetro sobre el optimizador de enjambre de partículas [Shi y Eberhart, 1998] [98]. Se concluye que el PSO con el peso de inercia en el rango [0.9, 1.2] en promedio tendrá un mejor rendimiento; es decir, tiene una mayor oportunidad de encontrar el óptimo global dentro de un número razonable de iteraciones. Además, se introduce un peso de inercia decreciente en el tiempo que trae una mejora significativa en el rendimiento de PSO.

$$v_{t+1} = W v_t + R_1 \left[0, \frac{\varphi}{2} \right] \otimes (p_i - x_t) + R_2 \left[0, \frac{\varphi}{2} \right] \otimes (p_g - x_t) \quad (4.111)$$

En la anterior ecuación, el factor ‘W’ es el peso de inercia.

$$x_{t+1} = x_t + v_{t+1}$$

Eberhart y Shi sugirieron también que utilizar un peso de inercia que disminuye con el tiempo, típicamente de 0.9 a 0.4, con $\frac{\varphi}{2} = 2.0$, tiene el efecto de reducir la búsqueda, cambiando gradualmente de un modo exploratorio a un modo explotador.

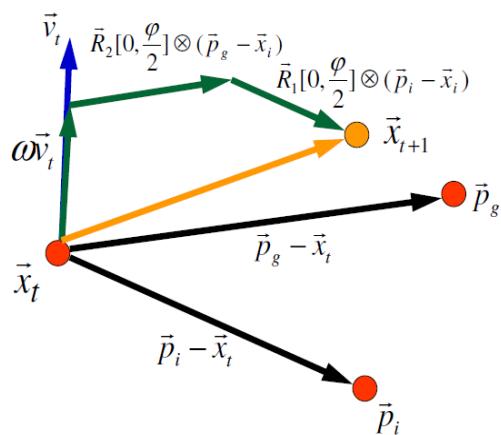


Figura 4.26. Visualización del efecto del peso de inercia en el PSO. Fuente: Xiaodong Li, School of Computer Science and IT, RMIT University, Melbourne, Australia

Clerc y Kennedy sugirieron un PSO más generalizado, donde se aplica un coeficiente de restricción (coeficiente de Tipo 1") a ambos términos de la fórmula de velocidad. Clerc muestra que la restricción PSO puede converger sin usar V_{max} :

$$v_{t+1} = \chi(v_t + R_1 \left[0, \frac{\varphi}{2} \right] \otimes (p_i - x_t) + R_2 \left[0, \frac{\varphi}{2} \right] \otimes (p_g - x_t)) \quad (4.112)$$

$$x_{t+1} = x_t + v_{t+1} \quad (4.113)$$

Donde $\frac{\varphi}{2}$ es un número positivo, fijado a menudo a 2.05; y χ el factor de restricción establecido 0,7289 [Clerc y Kennedy 2002] [99].

Utilizando el coeficiente de restricción, la amplitud de la oscilación de la partícula disminuye, asegurando su convergencia en el tiempo.

4.3.3. PSO completamente informado

El PSO Completamente Informado (en inglés, Fully Informed Particle Swarm, FIPSO), es una variación del modelo de factor de restricción, que hace uso de todas las variables para calcular la velocidad. Dado que los dos términos en la PSO de restricción son de la misma forma, se pueden comprender en la siguiente expresión [Mendes & Kennedy, 2004]:

$$v_{t+1} = \chi(v_t + \varphi_m \otimes (p_m - x_t)) \quad (4.114)$$

$$x_{t+1} = x_t + v_{t+1} \quad (4.115)$$

Donde $\varphi_m = \varphi_1 + \varphi_2$ y $p_m = (\varphi_1 \otimes p_i + \varphi_2 \otimes p_g) / (\varphi_1 + \varphi_2)$.

Esto demuestra que una partícula tiende a converger hacia un punto determinado por p_m , el cual es el promedio ponderado de su mejor p_i anterior y el mejor del vecindario p_g . p_m puede generalizarse a cualquier número de términos:

$$v_{t+1} = \chi(v_t + \sum_{k \in N} R \left[0, \frac{\varphi}{|N|} \right] \otimes (p_k - x_t)) \quad (4.116)$$

$|N|$ indica la *vecindad*, y p_k la ‘*mejor posición anterior*’ encontrada por la k -ésima partícula en N . Si $|N|$ es igual a 2, entonces lo anterior es una generalización del PSO canónico.

Kennedy (2006) describe PSO en la siguiente forma:

Nueva Posición = Posición Actual + Persistencia + Influencia social.

Si sustituimos $v_t = x_t + v_{t-1}$ en FIPSO, entonces tenemos:

$$x_{t+1} = x_t + \chi((x_t + v_{t-1}) + \sum_{k \in N} R \left[0, \frac{\varphi}{|N|} \right] \otimes (p_k - x_t)) \quad (4. 117)$$

$(x_t + v_{t-1}) \longrightarrow \text{‘Persistencia’}$

$\sum_{k \in N} R \left[0, \frac{\varphi}{|N|} \right] \otimes (p_k - x_t) \longrightarrow \text{‘Influencia social’}$

Entendiendo por ‘*persistencia*’ la tendencia de una partícula a persistir en moverse en la misma dirección que se estaba moviendo previamente.

El término de ‘*influencia social*’ puede concretarse aún más:

Nueva Posición =

Posición Actual + Persistencia + Tendencia Social Central + Dispersión Social

Se puede estimar la ‘*tendencia social central*’, por ejemplo, tomando la media de los mejores anteriores en relación con la ‘*posición actual*’ de la partícula (preguntas aún abiertas). La ‘*dispersión social*’ puede estimarse tomando la distancia entre la anterior mejor de una partícula y la ‘*anterior mejor*’ de cualquier vecino; o promediando distancias en pares entre la partícula y algunos vecinos. [Kennedy (2006)] utilizó algunas distribuciones como Gaussiana, doble exponencial y Cauchy.

4.4. Método de optimización de SVM mediante PSO

Las máquinas de vectores soporte, presentan el inconveniente de la existencia de parámetros libres o hiperparámetros y parámetros de kernel, que deben ser definidos por el usuario, y que limitan su uso en las plataformas académicas e industriales. Puesto que la aptitud de los modelos de regresión de SVM depende de una configuración adecuada de los anteriores parámetros, el principal problema para quienes tratan de aplicar SVMs, radica en la manera de establecer los correspondientes valores; con el fin de garantizar un buen rendimiento de generalización, para un determinado conjunto de datos de entrenamiento [Xiaoyong Liu &Hui Fu, 2014]. El método de SVM basado en PSO, optimiza dos hiperparámetros importantes, como son C y ε . El hiperparámetro C , determina el equilibrio entre la complejidad del modelo y el grado en que se toleran las desviaciones mayores que ε .

En este trabajo se propone la utilización de un modelo híbrido basado en la combinación de un algoritmo puro de clasificación, junto con un algoritmo de optimización. Concretamente, se aplicará una técnica de clasificación conocida como Support Vector Machine (SVM), potenciada con una metaheurística de optimización por enjambre de partículas (en inglés, Particle Swarm Optimization (PSO)). La función que deberá desempeñar el algoritmo de optimización (PSO), será mejorar el funcionamiento de la máquina de soporte vectorial (SVM), a través de la búsqueda de parámetros que una vez construido el modelo conformen un clasificador optimizado por PSO, que tenga la capacidad de responder bien frente a cualquier dato, ofreciendo mayores ventajas de uso y aumentando la efectividad del conjunto.

4.4.1. Enfoque basado en el PSO-SVM

El objetivo fundamental de optimizar el sistema de clasificación basado en el enfoque de máquinas de soporte vectorial (SVM), mediante la técnica de optimización por enjambre de partículas (PSO), es el de mejorar la precisión del clasificador SVM. Para que el algoritmo implementado, basado en PSO-SVM actúe eficientemente, deberemos detectar un subconjunto que contenga las mejores características informativas y la estimación de los mejores valores, para la regularización de los parámetros del núcleo en el modelo SVM. Por tanto, el algoritmo de optimización por enjambre de partículas (PSO), sirve para orientar la selección de aquellos potenciales subconjuntos, que nos puedan proporcionar de la forma más precisa la mejor predicción. Es decir, se emplea para encontrar los subconjuntos de características óptimas a partir de los conjuntos de datos procesados, con el fin de hallar las mejores combinaciones de características, a medida que progresan dentro del espacio del problema.

El proceso de ejecución del algoritmo de optimización por enjambre de partículas (PSO), se ejecuta de forma iterativa iniciándose con ' n ' partículas aleatoriamente seleccionadas en busca de una partícula óptima; representando cada partícula un vector ' m -dimensional' que constituye una posible solución. Mediante una técnica de *validación cruzada* se evalúan los resultados del análisis estadístico, procurando que sean independientes de la partición entre datos de entrenamiento y de prueba. En cada posible solución se va implementando el clasificador SVM y calculando la media aritmética obtenida en las medidas de evaluación efectuadas sobre las diferentes particiones. Cada subsiguiente población de partículas candidatas, en promedio, encajará mejor que su predecesora; continuando el mismo procedimiento hasta que el rendimiento del SVM converge [Radoslav Goldman, et.al, 2007]. En conclusión, el método de optimización por enjambre de partículas (PSO), se emplea para encontrar los subconjuntos de características óptimas a partir de los conjuntos de datos procesados, con el fin de hallar las mejores combinaciones de características, a medida que progresan dentro del espacio del problema.

Seguidamente se describe, en forma de diagrama de flujo, la operativa del método PSO-SVM propuesto.

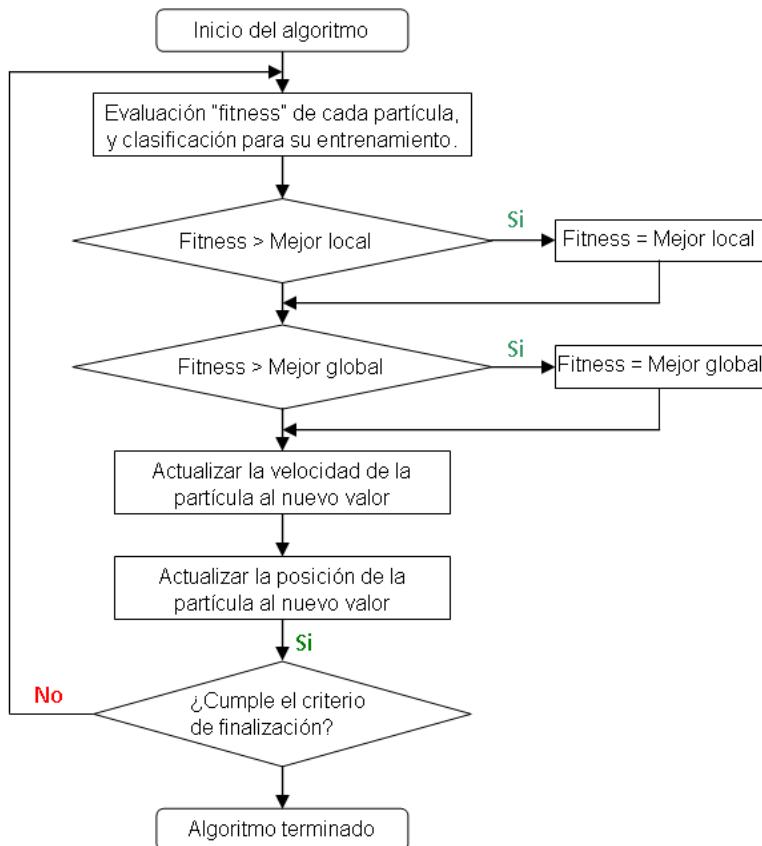


Figura 4.27. Flujograma del método de optimización basado en PSO-SVM.

Elaboración propia a partir de Intelligent Information Management, [Sheng Ding, Li Chen 2010].

El escenario de partida en el flujograma anterior es el siguiente: el tamaño de la población, el peso de inercia y las generaciones no presentan mejora cuando se inicia el algoritmo PSO. En primera instancia, se evalúa la aptitud de cada partícula. Tras la evaluación, se comparan los valores de aptitud y se determinan la partícula '*mejor local*' y la partícula '*mejor global*'. El siguiente paso, consiste en actualizar la *velocidad* y la *posición* de cada partícula hasta que el valor de la función de aptitud converja. Después de converger, la *mejor partícula global* en el enjambre es alimentada al clasificador SVM para su entrenamiento. Como paso último, se formará el clasificador de la SVM.

Capítulo V

ESTADO DEL ARTE

Hallazgos
documentales y
revisión

5. Hallazgos documentales y revisión

5.1. Evolución de la técnica

Los primeros teóricos que concibieron los fundamentos de la computación neuronal fueron el neurofísico Warren McCulloch, y el matemático Walter Pitts, quienes formularon una teoría acerca de la forma de trabajar de las neuronas en 1943.

En 1949 Donald Hebb escribió el libro: La organización del comportamiento, en el que se establece una conexión entre psicología y fisiología.

En 1957 Frank Rosenblatt comenzó el desarrollo del Perceptrón, que es la red neuronal más antigua.

En 1960 Bernard Widrow y Marcial Hoff, desarrollaron el modelo ADALINE (en inglés, ADaptive LINear Elements), que fue la primera red neuronal aplicada a un problema real de filtros adaptativos para eliminar los ecos en las líneas telefónicas.

1969 - Marvin Minsky/Seymour Papert. En este año surgieron críticas que frenaron, hasta 1982, el crecimiento que estaban experimentando las investigaciones sobre redes neuronales.

1974 - Paul Werbos. Desarrolló la idea básica del algoritmo de aprendizaje de propagación hacia atrás (backpropagation); cuyo significado quedó definitivamente aclarado en 1985.

1985 - John Hopfield. Provocó el renacimiento de las redes neuronales con su libro: «Computación neuronal de decisiones en problemas de optimización».

1986 - David Rumelhart/G. Hinton. Redescubrieron el algoritmo de aprendizaje de propagación hacia atrás (backpropagation).

Son numerosos los trabajos que se realizan y publican cada año. Dado que las técnicas que se emplean en esta tesis están relacionadas con las máquinas de vectores soporte, como consecuencia del estudio efectuado sobre el Estado del Arte, se presentan seguidamente los siguientes hallazgos que reflejan lo más puntero de la investigación actual en este sentido.

5.2. Situación actual del arte y revisión documental

Las máquinas de vectores soporte (SVM) pueden emplearse para solucionar u optimizar distintos tipos problemas de detección y clasificación, en la práctica se plantean optimización de algoritmos realizando híbridos de las SVM con otras metodologías como Wavelet, redes neuronales y cadenas de Markov. En la página web de la IEEE Xplore Digital Library, <http://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=SVM>, se presentan los últimos trabajos sobre máquinas de vectores soporte desarrollados en los últimos 2 años para conferencias, magazines y artículos, a los que se puede acceder fácilmente.

Los usos de las máquinas de vectores soporte son muy variados, se pueden encontrar aplicaciones que van desde el campo de la medicina hasta la seguridad en redes. No obstante, las SVM son ampliamente utilizadas en la clasificación de muestras y reconocimiento. Por ejemplo, en el caso de la visión artificial, se usan para el reconocimiento de objetos, rostros, caracteres y matrículas de automóviles, mediante interfaz hombre máquina. En el campo de la seguridad informática, se puede analizar el tipo de correo entrante y verificar si es o no correo spam; también se puede usar para la simulación y predicción de la evolución de la economía.

Las SVM se pueden aplicar a muchísimas áreas como hemos visto en el primer párrafo de este punto; no obstante, después de la revisión documental efectuada respecto al estado del arte en relación a la temática de esta investigación, son de nuestro interés sólo una pequeña parte de estas aplicaciones. Seguidamente se presentan algunas de las investigaciones revisadas, que tienen un vínculo importante con la temática, el contenido y el desarrollo de esta tesis doctoral:

5.2.1. Hallazgos documentales de interés

Máquinas de vectores de soporte y redes de perceptrón multicapa para evaluar la presencia de cianotoxinas a partir de concentraciones experimentales de cianobacterias en el embalse de Trasona (norte de España). J. A. Vilán Vilán et al., Oviedo, 2013.

El objetivo de este estudio es construir un modelo de diagnóstico de cianotoxina utilizando máquinas de vectores de soporte y redes de perceptrón multicapa a partir de concentraciones de cianobacterias determinadas experimentalmente en el embalse de Trasona (embalse recreativo utilizado como centro de entrenamiento de alto rendimiento de canotaje en el norte de España). Los resultados del presente estudio son dobles. En primer lugar, el significado de cada variable biológica y físico-química sobre la presencia de cianotoxinas en el reservorio se presenta a través del modelo. En segundo lugar, se obtiene un modelo predictivo capaz de pronosticar la posible presencia de cianotoxinas.

Modelado del desgaste de la herramienta de fresado mediante el uso de un modelo evolutivo basado en SVM con datos experimentales. P. J. García Nieto, et al., Oviedo, 2015.

El modelo se basó en Particle Swarm Optimization (PSO) en combinación con máquinas de vectores de soporte (SVM). Este mecanismo de optimización implicó la configuración de los parámetros del kernel en el procedimiento de entrenamiento SVM, lo que influye significativamente en la precisión de la regresión. Teniendo esto en cuenta, un modelo basado en PSO-SVM, que se basa en la teoría del aprendizaje estadístico, se utilizó con éxito aquí para predecir el desgaste del flanco de la herramienta de fresado (variable de salida) en función de las siguientes variables de entrada: la duración del tiempo experimento, profundidad de corte, alimentación, tipo de material, etc. Este mecanismo de optimización implicó la configuración de los parámetros del kernel en el procedimiento de entrenamiento SVM, lo que influye significativamente en la precisión de la regresión.

Sistema distribuido basado en Hadoop para la predicción en línea de la contaminación del aire basado en máquina de vectores soporte. Z. Ghaemi et al., Teherán, Irán, 2015.

El propósito de esta investigación es presentar un enfoque de pronóstico ‘on line’ basado en la Máquina de Vectores de Soporte (SVM), para predecir la calidad del aire con un día de anticipación. Con el fin de superar los requisitos computacionales para el análisis de datos a gran escala, se ha empleado la computación distribuida basada en la plataforma Hadoop para aprovechar la potencia de procesamiento de múltiples unidades de procesamiento. En este estudio, se adopta el modelo de programación ‘MapReduce’ para el procesamiento paralelo masivo. Hadoop es un marco de código abierto, para almacenar datos y ejecutar aplicaciones en grupos de equipos básicos, que proporciona un almacenamiento masivo para cualquier tipo de datos, un enorme poder de procesamiento y la capacidad de manejar tareas o trabajos con muchos datos.

Predicción de eventos críticos de calidad del aire utilizando máquinas de vectores soporte y optimización por enjambre de partículas. J. C. Zapata-Hernandez, et al., Santander, Colombia, 2016.

Se propone una metodología para predecir eventos críticos de calidad del aire (CAQE) en el Valle de Aburrá basados en máquinas de vectores de soporte (SVM) optimizadas con optimización de enjambre de partículas (PSO) y un esquema de caracterización para evaluar las tendencias actuales y pasadas de contaminantes y clima comportamiento, analizar el comportamiento estadístico en diferentes intervalos de tiempo.

Una distribución de predicción de contaminantes atmosféricos utilizando máquinas de vectores de soporte, análisis discriminante y herramientas cartográficas (caso de estudio: Túnez). Bedoui, S. et al., Túnez, 2016.

La máquina de vectores de soporte (SVM), como método de análisis de aprendizaje supervisado, se considera una herramienta estadística efectiva para la predicción y el análisis de la calidad del aire.

El trabajo presentado aquí examina la viabilidad de aplicar el SVM para predecir las concentraciones de ozono y partículas en dos ciudades tunecinas, a saber, Túnez y Sfax. Utilizamos el SVM con el kernel lineal, SVM con el kernel polinomial y SVM con el kernel RBF para predecir las concentraciones de ozono (O_3) y partículas suspendidas en Túnez durante un año. El kernel RBF produjo buenos resultados para los dos contaminantes con una tasa de error del 0%. Los núcleos polinomiales y lineales produjeron errores suficientemente bajos para los contaminantes, en 9.09% y 18.18%, respectivamente. Los resultados del análisis discriminante muestran que la caracterización espacial permite la discriminación exitosa entre las dos ciudades con una tasa de error de 4.35% para el análisis lineal y 0% en el caso del análisis cuadrático.

Estimación de los efectos causales de la contaminación del aire local en las muertes diarias: efecto de los niveles bajos. Joel Schwartz, et al. Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA , 2016.

En esta investigación se emplea una versión de SVM para *regresión* [Cortes y Vapnik 1995], con un kernel radial para estimar la variación en $PM_{2.5}$ (o en 'Black Carbon' BC o NO_2) y sus productos, incluyendo posibles dependencias no lineales en los predictores. Este enfoque combina varios instrumentos en un único instrumento calibrado, y permite comparar los cambios de rango intercuartil en los instrumentos de contaminación local calculados utilizando cada uno de los contaminantes ($PM_{2.5}$, BC, o NO_2) como indicador.

La investigación concluye que existe una asociación causal entre la contaminación del aire local con las muertes diarias, en concentraciones por debajo de los estándares de la EPA. El riesgo atribuible estimado en Boston superó las 1,800 muertes durante el período de estudio, lo que indica que se pueden obtener importantes beneficios de salud pública como resultado de aplicar mecanismos adicionales de control.

Aplicación de la máquina de vectores de soporte y la programación de expresión génica (GEP) para el pronóstico de ozono troposférico en Teherán Metropolitano. Vahid Mehdipour, Mahsa Memarianfard, Teherán, Irán, 2017.

En este documento, la máquina de vectores de soporte (SVM) y la programación de expresión génica (GEP) son dos métodos potentes con resultados fiables en estudios previos, utilizados para predecir el ozono troposférico en Teherán metropolitano utilizando precursores fotoquímicos y parámetros meteorológicos como predictores. Para el SVM, se utilizan varios kernel seleccionando el mejor en función del resultado.

Clasificación de calidad del aire ambiente mediante algoritmo de optimización Grey Wolf Optimizer basado en máquinas de vectores soporte (SVM). Akash Saxena and Shalini Shekhawat. Jaipur, India, 2017.

En esta investigación, se elabora un marco matemático para formular un Índice acumulativo (CI) sobre la base de la concentración individual de cuatro contaminantes principales (SO_2 , NO_2 $\text{PM}_{2.5}$ y PM_{10}). Además, se propone un clasificador basado en un algoritmo de aprendizaje supervisado empleando máquinas de vectores soporte (SVM) para clasificar la calidad del aire en dos tipos, es decir, bueno o nocivo. Las posibles entradas para este clasificador son los valores calculados de los CI. La eficacia del clasificador se prueba con los datos reales de tres ubicaciones: Kolkata, Delhi y Bhopal. Se observa que el clasificador se comporta bien para clasificar la calidad del aire.

Los objetivos de investigación para este estudio son los siguientes:

- Presentar un marco matemático en la formulación del CI empleando los valores numéricos de los índices de calidad del aire de (SO_2 , NO_2 , $\text{PM}_{2.5}$ y PM_{10}).
- Presentar un análisis comparativo del IC propuesto con los índices de calidad del aire existentes.

- Establecer una rutina de optimización y proponer una solución basada en el algoritmo de Grey Wolf Optimization (GWO) estimando los parámetros del núcleo y el sesgo para SVM.
- Derivar una función lineal discriminante del SVM para clasificar los datos en función de la calidad del aire.

5.3. Origen de la investigación

El siguiente estudio de investigación se toma como documento de referencia y origen, para la elaboración de la presente tesis doctoral. El siguiente texto es un resumen de la publicación, en el que se han eliminado las referencias bibliográficas con el fin de simplificar y resumir el trabajo. Las particularidades muy resumidas de esta investigación se detallan a continuación:

A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study.

Un modelo de regresión basado en SVM para estudiar de la calidad del aire a escala local en el área urbana de Oviedo (norte de España): Un caso de estudio.

P. J. García Nieto, et al., Oviedo, 2013.

Resumen

En este trabajo, basado en los datos de NO, NO₂, CO, SO₂, O₃ y materia particulada (PM₁₀), registrados para los años 2006, 2007 y 2008, se utiliza la regresión de vectores soporte (en inglés, Support Vector Regression SVR), para construir un modelo dinámico no lineal de la calidad del aire, en la ciudad de Oviedo, en el norte de España.

5. Hallazgos documentales y revisión

Capítulo VI

Aplicación al caso de estudio

Modelado y análisis de la calidad del aire en la ciudad de Oviedo (norte de España), mediante el enfoque PSO-SVM, red neuronal MLP y árbol de regresión M5

6. Modelado y análisis de la calidad del aire en la ciudad de Oviedo (norte de España), mediante los enfoques PSO-SVM, red neuronal MLP y árbol de regresión M5. Caso de estudio

6.1. Planteamiento del caso de estudio

El objetivo principal de este estudio será construir varios modelos de regresión de la calidad del aire utilizando técnicas basadas en el aprendizaje estadístico, en el área metropolitana de Oviedo, que está situada en el norte de España. En este trabajo de investigación, se implementará un modelo híbrido basado en una regresión de vectores soporte y un algoritmo evolutivo de optimización de una nube de partículas a partir del conjunto de datos experimentales, para predecir la calidad del aire (específicamente óxidos de nitrógeno, monóxido de carbono, dióxido de azufre, ozono y materia particulada) recogidos desde 2013 a 2015 en el área urbana de Oviedo. Al mismo tiempo, también se ajustarán al conjunto de datos experimentales un modelo de árbol de regresión M5 y una red de perceptrón multicapa (MLP). Con ello, se quiere demostrar que los resultados obtenidos mediante el modelo híbrido propuesto, son más robustos que los obtenidos mediante los métodos de predicción MLP y árbol de regresión M5, en términos de estimadores estadísticos y de las pruebas de rendimiento.

Palabras clave. Máquinas de Vectores Soporte (SVMs); Optimización por enjambre de partículas (PSO); Redes neuronales artificiales (RNAs); Árbol de regresión modelo M5; Modelización y control de la calidad del aire.

6.2. Introducción

La contaminación del aire puede definirse como la introducción en la atmósfera de sustancias químicas, partículas o elementos biológicos que pueden causar malestar, enfermedades e incluso la muerte de seres humanos, animales o plantas. También puede deteriorar el entorno natural o construido [100-102]. La contaminación del aire procede de fuentes diversas y puede tener origen natural o humano: (a) fuentes de origen natural, tales como erupciones volcánicas y partículas arrastradas por el viento; (b) fuentes de origen humano, fuentes estáticas originadas por el hombre, tales como fábricas, centrales eléctricas o algunas operaciones de limpieza en seco y desengrasado; y fuentes móviles originadas por el hombre, tales como vehículos motorizados, aviones y trenes, contribuyen a la contaminación del aire.

En el control de la calidad del aire, la primera respuesta a una amenaza, conocida o potencial que afecte a la norma o directriz de la calidad del aire establecida, es reducirla. Los Planes Estatales de Implementación (por sus siglas en inglés, SIP) formalizan dichas respuestas en España [100-102]. La contaminación atmosférica es un problema medioambiental importante en las áreas metropolitanas [100-104] como Oviedo (Principado de Asturias, España). Puede causar problemas de salud que conducen a dificultad para respirar, tos y el empeoramiento de problemas cardíacos y respiratorios existentes [102-104]. Por ejemplo, las emisiones diesel (en inglés, *Emissions from diesel-DE*) constituyen una de las principales fuentes de emisión de material particulado, que se originan durante la combustión. Las emisiones diesel (DE), se han relacionado con el aumento de la creación de trombos y la disfunción vascular aguda en varios estudios de salud humana. Esto explicaría el vínculo anteriormente descrito entre el aumento de la mortalidad y la morbilidad cardiovascular y la contaminación debida al material particulado del aire [100-102, 105]. Oviedo es el centro administrativo del Principado de Asturias, en el norte de España. Tiene una población de 221.202 habitantes y una superficie de 186.65 km², con una altitud de 232 m sobre el nivel del mar y una densidad de población de 1185,12 habitantes por kilómetro cuadrado.

El clima de Oviedo, al igual que en el resto del noroeste de España, es más diverso que en otras partes de España. Los veranos son generalmente húmedos y cálidos, con sol, pero también algo de lluvia, mientras que los inviernos son fríos y muy lluviosos. La nieve suele estar presente en las montañas que rodean la ciudad de Oviedo, de octubre a mayo. Tanto la lluvia como la nieve ocasional son características meteorológicas regulares de los inviernos de Oviedo. Por otra parte, la central de carbón de Soto de Ribera, planta que suministra la mayor parte de la energía eléctrica consumida en la ciudad de Oviedo, está situada a siete kilómetros al sur de dicha ciudad, a una altitud de 126,50 m sobre el nivel del mar, en el distrito de Ribera de Arriba (Figura 6.1). Las localizaciones geográficas de las tres estaciones meteorológicas empleadas para captar los datos de estudio y de la central de carbón de Soto de Ribera, anteriormente citada, se muestran en la Figura 6.1.



Figura 6.1. Ubicación geográfica de las estaciones meteorológicas en la zona urbana de Oviedo (norte de España) y de la central eléctrica de Soto de Ribera (central eléctrica a carbón cerca de la ciudad de Oviedo). Elaboración propia.

La vigilancia de la contaminación meteorológica, la medición de componentes como el CO, SO₂, NO, NO₂, O₃ y partículas (PM₁₀) es cada vez más importante debido a sus efectos adversos sobre la salud humana [100-102, 106,110]. Por lo tanto, la UE y muchos organismos nacionales de medio ambiente han establecido normas y directrices de calidad del aire para los niveles permitidos de estos contaminantes en el aire [104,110, 111]. El principal objetivo de este trabajo es construir un modelo para la contaminación media diaria, que pueda ser de utilidad para las autoridades responsables de la regulación de la contaminación atmosférica en la región correspondiente. Los datos empleados en este estudio han sido recogidos a lo largo de tres años: concretamente de 2013 a 2015. Los experimentos numéricos efectuados aplicando la técnica basada en PSO-SVM, han permitido obtener un buen grado de precisión de modelado diario para todos los contaminantes considerados. Estos resultados, se presentan y se discuten en este trabajo de investigación.

Para fijar las ideas, el objetivo de este estudio es evaluar el enfoque basado en la aplicación de máquinas de soporte vectorial (SVMs) [112-119] en combinación con las técnicas de optimización evolutiva, consistente en la optimización por nube de partículas u optimización por enjambre de partículas (conocida por sus siglas en inglés: PSO, de «*particle swarm optimization*») [119-123]; así como la red neuronal artificial del Perceptrón Multicapa (MLP) [124-130] y árboles de regresión M5 [131-133], para modelizar la calidad del aire en el área urbana de Oviedo (norte de España) a escala local, comparando los resultados obtenidos. El soporte teórico para los algoritmos de aprendizaje de SVMs está basado en la teoría del aprendizaje estadístico y la minimización del riesgo estructural. En concreto, se han creado cinco modelos basados en PSO-SVM para el dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y partículas de aerosol menores de 10 μm (PM₁₀), como una función que utilizó los otros contaminantes relevantes medidos en la calidad del aire como variables independientes: óxido nítrico (NO), monóxido de carbono (CO) y ozono (O₃); con el propósito de obtener estimaciones de concentración precisas para los contaminantes NO₂, SO₂ y PM₁₀ [134-136].

Los modelos basados en SVM pueden ser utilizados como una alternativa a los clásicos de enfoques regresión y son una nueva familia de modelos que se pueden utilizar para la estimación de valores en muy diferentes áreas [112-119]. Se encontró que los cinco modelos basados en PSO-SVM mejoran la precisión para el caso de problemas de regresión no lineal, tales como los estudiados en este trabajo relacionados con la calidad del aire.

La técnica de optimización de enjambre de partículas (PSO) se usó aquí con éxito para optimizar la puesta a punto de los hiperparámetros óptimos del núcleo (kernel) en la fase de entrenamiento del SVM. La optimización de enjambre de partículas (PSO) fue introducida por Kennedy y Eberhart en 1995 [120] y es un algoritmo bio-inspirado en la inteligencia del enjambre (en inglés, Swarm Intelligence, SI). El PSO se basa en la simulación del comportamiento de la emigración de los pájaros o ‘flocado’ [120-123] y es similar a otros algoritmos de cálculo evolutivo basados en SI. También explora el modelo de intercambio social de información [137-138]. Los modelos PSO hibridados con SVM (PSO-SVM) [137-138] se emplearon como herramienta de aprendizaje, y entrenamiento para estimar la calidad del aire de otros contaminantes atmosféricos en el área urbana de Oviedo, a escala local.

Se empleó un modelo híbrido PSO optimizado mediante SVM (PSO-SVM), así como un modelo MLP y el árbol modelo M5 [124-133], como herramientas de aprendizaje automatizado, entrenándolos para predecir la calidad del aire en el área urbana de Oviedo, a partir de los datos de entrada físico-químicos medidos experimentalmente.

Este innovador trabajo de investigación se organiza de la siguiente manera. En primer lugar, se describen los materiales y métodos necesarios para llevar a cabo el estudio, en segundo lugar, se muestran y discuten los resultados obtenidos y finalmente, se exponen las principales conclusiones extraídas de los resultados.

6.3. Marco teórico y metodología del experimento

6.3.1. Fuentes y tipos de contaminación del aire, en el área de localización del problema.

Un contaminante del aire es una sustancia contenida en el aire atmosférico, que puede no ser saludable para los seres humanos y el medio ambiente. Los contaminantes se pueden encontrar en forma de partículas sólidas, gotas líquidas o gases. Además, pueden ser debido a la acción del hombre o a causas de origen natural. Los contaminantes pueden clasificarse como primarios o secundarios. Los contaminantes primarios provienen de un proceso, como en el caso del monóxido de carbono de un escape de vehículos a motor, el dióxido de azufre de las fábricas o la ceniza que puede ser debida a una erupción volcánica. Los contaminantes secundarios se forman en el aire cuando los contaminantes primarios interactúan o reaccionan y, por lo tanto, dichos contaminantes no se emiten directamente. Por ejemplo, un importante contaminante secundario es el ozono troposférico o a nivel del suelo, que es uno de los muchos contaminantes secundarios que constituyen el smog fotoquímico [103,134-136,139]. Algunos contaminantes pueden ser tanto primarios como secundarios, es decir, pueden haber sido emitidos directamente y también formados a partir de otros contaminantes primarios.

Principales contaminantes primarios debidos a la actividad humana [100-111,134-136,139-141]:

- Material Particulado (PM): también llamado material particulado atmosférico o partículas finas. Son pequeñas partículas de sólido o líquido suspendidas en un gas. Por otra parte, un aerosol indica partículas y gas juntos.
- Óxidos de azufre (SO_x): en particular, dióxido de azufre, un compuesto químico de fórmula SO_2 . La combustión del carbón y del petróleo genera dióxido de azufre, ya que a menudo contienen compuestos de azufre.

- Óxidos de nitrógeno (NO_x): principalmente dióxido de nitrógeno (NO_2) que se emite durante la combustión a alta temperatura. El primer producto formado es el óxido nítrico (NO) y cuando el NO se oxida aún más en la atmósfera se convierte en dióxido de nitrógeno (NO_2).
- Monóxido de carbono (CO): es producido por la combustión incompleta de combustibles tales como carbón, madera o gas natural.

Los contaminantes secundarios incluyen [100-111,134-136,139-141]:

- Material Particulado: se compone de contaminantes primarios gaseosos y compuestos en ‘smog’ fotoquímico. El smog es un tipo particular de contaminación del aire. El smog típico resulta de grandes cantidades de carbón quemado en una zona y es causado por una mezcla de humo y dióxido de azufre.
- Ozono troposférico o al nivel del suelo (O_3): se desarrolla a partir de NO_x y compuestos orgánicos volátiles (COV). La exposición a corto plazo a niveles elevados de ozono puede ser el origen de las irritaciones oculares y pulmonares.

Con respecto a las tendencias en la calidad del aire, la Ley de Aire Limpio de 1970 estableció criterios de normas para cuatro de los contaminantes primarios (aerosoles, dióxido de azufre, monóxido de carbono y óxidos de nitrógeno) y para el contaminante secundario ozono. Por aquel entonces, en 1970, estos cinco contaminantes fueron identificados como los más extendidos y menos deseables. Hoy en día, se ha añadido el plomo y se conocen por *contaminantes criterio*, estando contemplados en los estándares nacionales de calidad del aire ambiental de los Estados Unidos (Tabla 6.1) [100-111].

Tabla 6.1. Normas nacionales de calidad del aire ambiental de la Agencia de Protección Ambiental de los Estados Unidos (US EPA) [1-12,40-42]. Elaboración a partir de US EPA.

Contaminante	Máximas concentraciones permitidas
Monóxido de carbono (CO)	
8-h promedio	9 ppm (10 mg/m ³)
1-h promedio	35 ppm (40 mg/m ³)
Dióxido de nitrógeno (NO₂)	
Media aritmética anual	0.053 ppm (100 µg/m ³)
Ozono (O₃)	
1-h promedio	0.12 ppm (235 µg/m ³)
8-h promedio	0.08 ppm (157 µg/m ³)
Partículas<10µm/m³ (PM₁₀)	
Media aritmética anual	50 µg/m ³
24-h promedio	150 µg/m ³
Dióxido de azufre (SO₂)	
Media aritmética anual	0.03 ppm (80 µg/m ³)
24-h average	0.14 ppm (365 µg/m ³)

6.3.2. Datos experimentales.

6.3.2.1. Obtención del conjunto de datos

El Gobierno de Asturias, en particular su Sección de Industria y Energía, cuenta con tres estaciones de control de la calidad del aire distribuidas a lo largo de la ciudad de Oviedo (Figura 6.1). Cada quince minutos se toman medidas de los siguientes contaminantes primarios y secundarios: dióxido de azufre (SO₂), óxidos de nitrógeno (NO y NO₂), monóxido de carbono (CO), partículas de menos de 10 micrómetros (PM₁₀) y ozono (O₃).

Los seis contaminantes medioambientales estudiados con la ayuda de estas estaciones automatizadas de monitoreo se midieron con los siguientes sensores: (a) analizador API 100A para gas SO₂; (b) analizador API 200A para gases NO_x; (c) analizador TELEDYNE 300E para gas CO; (d) analizador TELEDYNE 400E para gas O₃; (e) analizador DASIBI 7001 para aerosol PM₁₀.

Tabla 6.2. Concentración media mensual de contaminación atmosférica en el área urbana de Oviedo entre enero de 2013 y diciembre de 2015. Elaboración propia.

Mes del año	SO₂	NO	NO₂	CO	PM₁₀	O₃
Enero2013	15	64	51	0.48	31	32
Febrero2013	14	43	41	0.47	29	36
Marzo2013	10	30	34	0.40	27	47
Abril 2013	8	9	22	0.37	33	52
Mayo2013	9	6	17	0.29	29	54
Junio 2013	5	8	17	0.35	31	48
Julio2013	9	9	19	0.37	36	52
Agosto 2013	13	3	23	0.37	31	48
Septiembre2013	12	20	31	0.38	34	40
Octubre 2013	9	28	33	0.46	28	26
Noviembre2013	13	42	44	0.35	25	22
Diciembre2013	22	71	50	0.43	37	28
Enero2014	11	41	40	0.37	27	37
Febrero2014	8	24	37	0.42	25	46
Marzo2014	11	25	36	0.42	37	47
Abril 2014	7	12	29	0.36	25	54
Mayo2014	7	11	26	0.37	20	59
Junio 2014	7	9	22	0.36	27	64
Julio2014	9	13	18	0.24	24	47
Agosto 2014	7	6	21	0.24	25	35
Septiembre2014	9	17	24	0.26	30	32
Octubre2014	9	15	21	0.26	28	40
Noviembre2014	7	13	23	0.27	24	32
Diciembre2014	8	32	27	0.32	31	28
Enero2015	16	56	40	0.41	29	22
Febrero2015	9	32	40	0.30	22	32
Marzo 2015	5	17	34	0.31	35	35
Abril 2015	8	12	31	0.29	39	42
Mayo2015	5	6	29	0.19	30	42
Junio 2015	7	31	26	0.22	34	43
Julio2015	5	22	31	0.24	33	46
Agosto 2015	7	8	27	0.32	30	41
Septiembre 2015	5	11	25	0.47	29	34
Octubre 2015	9	29	39	0.47	28	21
Noviembre2015	15	64	44	0.56	34	26
Diciembre2015	13	68	46	0.62	32	27

El procedimiento de medición de material particulado, se basa en la atenuación de los rayos beta para medir la concentración de las partículas en suspensión en el aire con un diámetro inferior a $10 \mu\text{m}$.

Estos sensores recogen los datos que se procesan y los entregan en promedio, para toda la ciudad, todos los días. Por lo tanto, tenemos datos de los contaminantes mencionados anteriormente cada día desde enero de 2013 a diciembre de 2015. Las concentraciones medias mensuales se muestran en la Tabla 6.2.

6.3.2.2. Análisis de la tendencia de las concentraciones de los contaminantes

Entonces, es posible estudiar la tendencia en las concentraciones de los contaminantes anteriores en los años 2013, 2014 y 2015 [100-111,134-136,139-141].

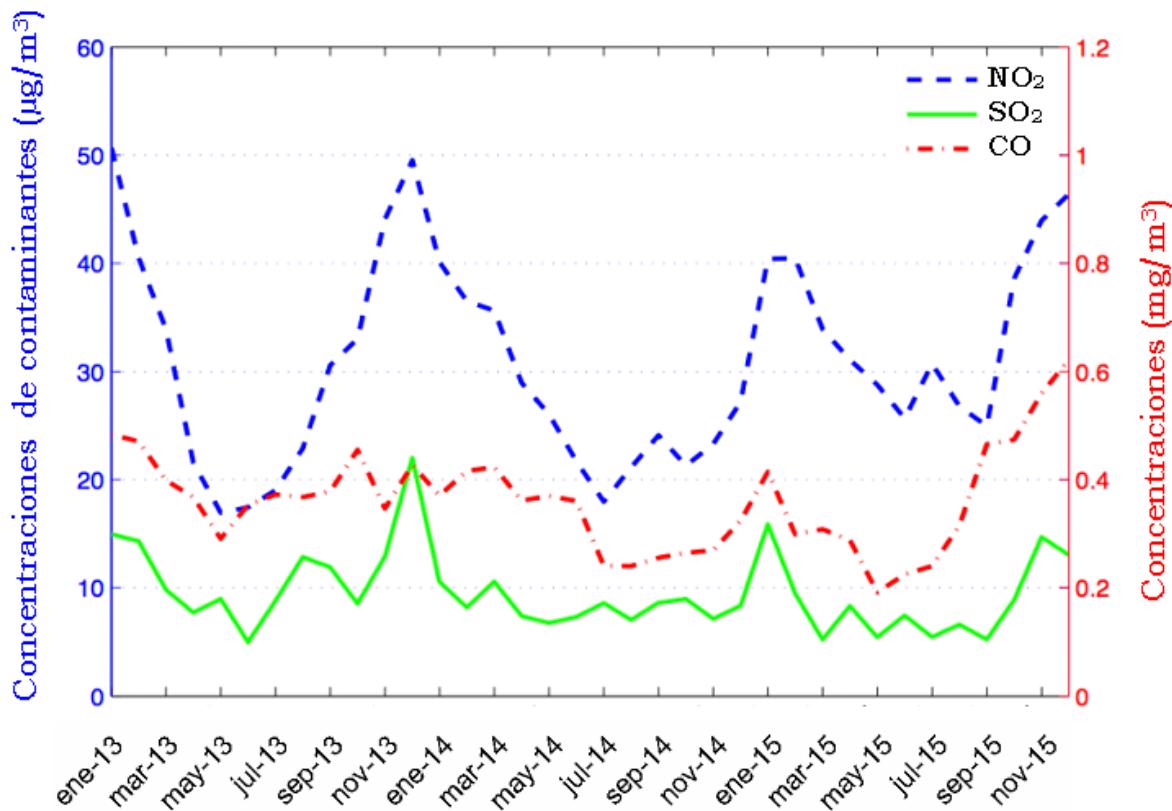


Figura 6.2. Tendencia mensual de las concentraciones de dióxido de nitrógeno (NO_2), dióxido de azufre (SO_2) y monóxido de carbono (CO) durante los años 2013, 2014 y 2015 en el área urbana de Oviedo. Elaboración propia.

La Figura 6.2 muestra las concentraciones mensuales de dióxido de nitrógeno (NO_2), dióxido de azufre (SO_2) y monóxido de carbono (CO) en tres años consecutivos (entre 2013 y 2015).

La cantidad de dióxido de nitrógeno (NO_2) fluctuó significativamente con varios máximos de $51 \mu\text{g}/\text{m}^3$ en enero de 2013, $50 \mu\text{g}/\text{m}^3$ en diciembre de 2013, $40 \mu\text{g}/\text{m}^3$ en enero y febrero de 2015 y $46 \mu\text{g}/\text{m}^3$ en diciembre de 2015, respectivamente. Estos máximos correspondieron a los meses de mayor consumo de energía en los hogares debido a las calefacciones y a la mayor densidad de automóviles en las carreteras durante la temporada de invierno. Asimismo, los mínimos de la concentración correspondieron a los meses de verano. De acuerdo con las Normas de Calidad del Aire de la US EPA (Tabla 6.1), la concentración máxima permisible de NO_2 expresada como media aritmética anual es de $100 \mu\text{g}/\text{m}^3$. Las medias aritméticas anuales para este gas durante los años 2013, 2014 y 2015 fueron 31.8 , 27.0 y $34.3 \mu\text{g}/\text{m}^3$ respectivamente. Por lo tanto, las concentraciones de dióxido de nitrógeno (NO_2) también están por debajo del máximo permitido y cumplen con los estándares de calidad del aire para una persona sana durante estos tres años, incluyendo los picos de emisión.

De forma similar, la concentración de dióxido de azufre (SO_2) también fluctuó ligeramente con máximos de $20 \mu\text{g}/\text{m}^3$ en diciembre de 2013, $16 \mu\text{g}/\text{m}^3$ en enero de 2015 y $15 \mu\text{g}/\text{m}^3$ en noviembre de 2015, respectivamente. Una vez más, estos máximos correspondieron a los meses de invierno. También es posible observar que la concentración de SO_2 siguió aproximadamente un comportamiento similar a la concentración de NO_2 , excepto que la concentración de SO_2 fue mucho menor. Esta tendencia es general a lo largo de los años estudiados, y está dentro de la lógica, dado que cerca de esta área hay una planta de carbón (Figura 6.1). Finalmente, la concentración de CO también subió y bajó ligeramente, pero tuvo un comportamiento más irregular correspondiendo también los máximos a los meses de invierno.

De manera similar, siguiendo las Normas de Calidad del Aire de la US EPA [100-111,139-141] (Tabla 6.1), la concentración máxima permisible de CO expresada como media aritmética anual es 3,33 mg/m³. Las medias aritméticas anuales para este gas durante los años 2013, 2014 y 2015 fueron de 0.39, 0.32 y 0.37 mg/m³ respectivamente.

Por lo tanto, las concentraciones de CO estuvieron por debajo del nivel más alto que puede ser tolerado por los seres humanos, de acuerdo con los estándares de calidad del aire de la US EPA, durante estos tres años, incluyendo picos de emisión [100-111].

Análogamente, la Figura 6.3 muestra las concentraciones mensuales en la zona urbana de Oviedo de material particulado (PM₁₀), ozono (O₃) y óxido nítrico (NO) desde enero de 2013 hasta diciembre de 2015.

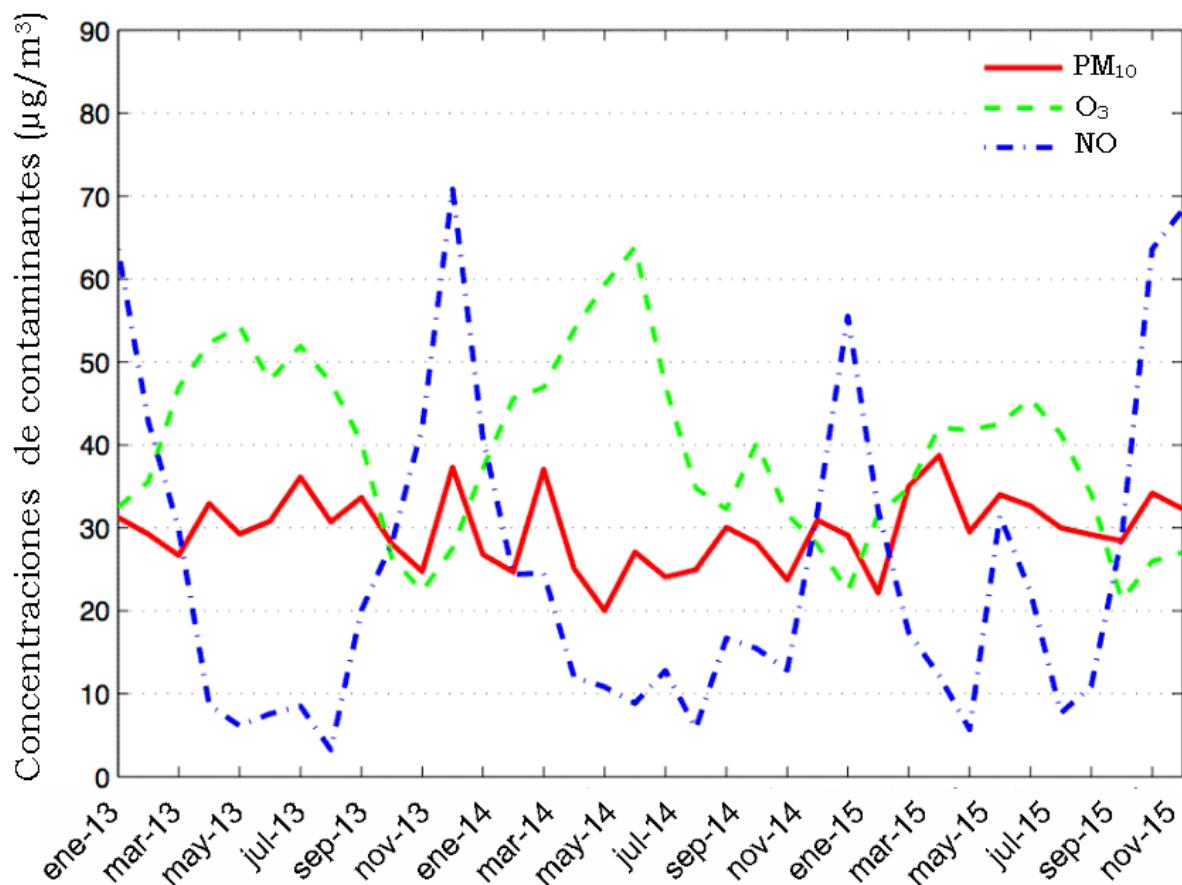


Figura 6.3. Tendencia mensual de las concentraciones de partículas (PM₁₀), ozono (O₃) y óxido nítrico (NO) durante los años 2013, 2014 y 2015, en el área urbana de Oviedo. Elaboración propia.

La concentración de materia particulada (PM_{10}) subió y bajó ligeramente, pero se mantuvo bastante estable en alrededor de $30 \mu\text{g}/\text{m}^3$, con dos picos a $37 \mu\text{g}/\text{m}^3$ en diciembre de 2013 y marzo de 2014; con un mínimo de $20 \mu\text{g}/\text{m}^3$ en mayo 2014 y un máximo de $39 \mu\text{g}/\text{m}^3$ en abril de 2015, respectivamente. Desde el punto de vista de la calidad del aire estándar, siguiendo las normas de calidad del aire de la US EPA (Tabla 6.1), la concentración máxima permisible de PM_{10} , expresada como media aritmética anual, es de $50 \mu\text{g}/\text{m}^3$.

Las medias aritméticas anuales para este contaminante durante los años 2013, 2014 y 2015 fueron de 30.9 , 26.9 y $31.3 \mu\text{g}/\text{m}^3$, respectivamente. Por lo tanto, las concentraciones de aerosol estuvieron por debajo del máximo permitido para una persona sana, durante estos tres años, aunque los picos de emisión fueron cercanos a este valor. Este comportamiento puede dar lugar a graves problemas de salud para la población, pudiendo dar lugar al incremento de enfermedades crónicas y muertes. De forma similar, la concentración de óxido nítrico (NO) fluctuó enormemente con máximos de $64 \mu\text{g}/\text{m}^3$ en enero de 2013, $71 \mu\text{g}/\text{m}^3$ en diciembre de 2013 (el pico más alto), $56 \mu\text{g}/\text{m}^3$ en enero de 2015 y $68 \mu\text{g}/\text{m}^3$ en diciembre de 2015, respectivamente. Una vez más, estos máximos correspondieron a los meses de invierno. Además, de forma análoga los mínimos de concentración de NO se produjeron durante los meses de verano. Los valores registrados fueron $3 \mu\text{g}/\text{m}^3$ en agosto de 2013, $6 \mu\text{g}/\text{m}^3$ en agosto de 2014 y $8 \mu\text{g}/\text{m}^3$ en agosto de 2015, respectivamente. Aunque el producto inicial de la combustión es el óxido nítrico (NO), este gas se oxida rápidamente y se convierte en NO_2 . Su tiempo de residencia en la atmósfera es muy corto y por ese motivo los estándares de calidad del aire de la US EPA no lo consideran [100-111].

Finalmente, la concentración de O_3 también fluctuó considerablemente, pero su comportamiento es justo lo contrario al del óxido nítrico (NO), es decir, los máximos de O_3 correspondían a mínimos de NO y viceversa. Esta tendencia es general a lo largo de los años estudiados, ya que el ozono se asocia con reacciones fotoquímicas, que requieren la presencia de luz solar fuerte que actúe como catalizador. La Ley de Aire Limpio obliga a la US EPA a establecer Normas Nacionales de Calidad del Aire Ambiental para varios contaminantes, incluyendo el ozono a nivel del suelo, y las ciudades que no cumplen con estas normas deben tomar medidas para reducir sus niveles.

En mayo de 2008, la US EPA bajó el nivel fijado para el ozono de $80 \mu\text{g}/\text{m}^3$ a $75 \mu\text{g}/\text{m}^3$. Esto resultó controvertido; ya que los propios científicos y el consejo asesor del Organismo, habían recomendado bajar la norma a $60 \mu\text{g}/\text{m}^3$, y la Organización Mundial de la Salud recomienda $51 \mu\text{g}/\text{m}^3$.

Muchos grupos de salud pública y ambiental también apoyaron la iniciativa de reducir el estándar a $60 \mu\text{g}/\text{m}^3$. Las medias aritméticas anuales para este gas en la zona urbana de Oviedo durante los años 2013, 2014 y 2015 fueron de 40.4, 43.4 y $34.3 \mu\text{g}/\text{m}^3$, respectivamente. Por lo tanto, las concentraciones de este gas estuvieron por debajo del máximo permitido y cumplen con los estándares de calidad del aire durante estos tres años, incluyendo picos de emisión. Sin embargo, en junio de 2014, se alcanzó un máximo de $68 \mu\text{g}/\text{m}^3$ y por lo tanto superior a $60 \mu\text{g}/\text{m}^3$, que es el valor recomendado. Este hecho puede ser peligroso para la salud de la población de Oviedo. Hay evidencias que demuestran que altas concentraciones de ozono, creadas por altas concentraciones de contaminación y rayos UV diurnos en la superficie de la Tierra, pueden dañar la función pulmonar e irritar el sistema respiratorio. La exposición al ozono y a los contaminantes que la producen se ha relacionado con muertes prematuras, asma, bronquitis, ataque cardíaco y otros problemas cardiopulmonares.

6.3.3. Método de las máquinas de soporte vectorial (SVM)

Las Máquinas de vectores de soporte (SVMs), son un conjunto de algoritmos de aprendizaje supervisado que están estrechamente relacionados con los problemas de clasificación y de regresión [112-119].

Este último método se denomina máquinas de vectores de soporte para regresión (SVR). Ahora, queremos predecir una salida de valor real y' . La función de regresión $y = f(\mathbf{x})$ para nuestros datos de entrenamiento

$T = \{(\mathbf{x}_i, y_i)\}_{i=1}^L$ donde $y_i \in \Re$ y $\mathbf{x}_i \in \Re^D$, es como sigue:

$$f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b \quad (6.1)$$

Donde w y b son, respectivamente, el vector de peso y la intersección del modelo. En general, usará una sofisticada función de penalización, no asignando penalización si el valor estimado y_i es menor que la distancia ε lejos del valor real t_i , es decir, si $|t_i - y_i| < \varepsilon$.

La región limitada por $y_i \pm \varepsilon$ para todo i se denomina tubo ε -insensible (Fig. 4). Otra modificación a la función de penalización es que las variables de salida que caen fuera del tubo se dan a través de dos variables de holgura de penalización dependiendo de si están por encima (ξ^+) o por debajo (ξ^-) del tubo (donde $\xi^+, \xi^- > 0$ para todo i):

$$t_i \leq y_i + \varepsilon + \xi^+ \quad (6.2)$$

$$t_i \geq y_i - \varepsilon - \xi^- \quad (6.3)$$

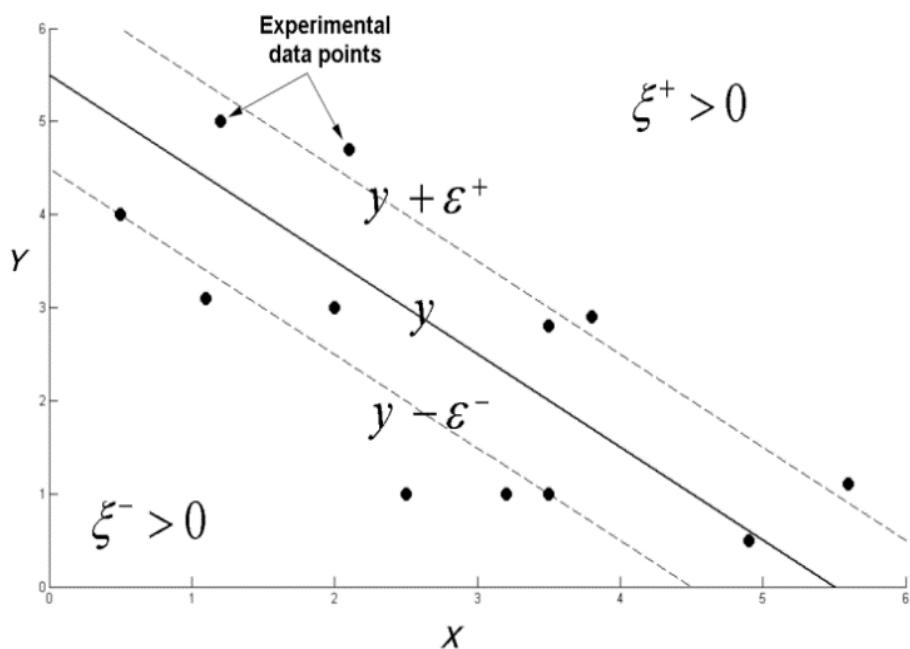


Figura 6.4. Regresión con tubo ε -insensible para problemas unidimensionales.
Fuente: J.C. Álvarez Antón et al./Applied Mathematical Modelling 37 (2013) 6244-6253

La función de error para la SVR puede escribirse como [112-119]:

$$C \sum_{i=1}^L (\xi_i^+ + \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2 \quad (6.4)$$

Donde C denota el parámetro de penalización o costo entre los errores empíricos y de generalización y ξ_i^+, ξ_i^- son las variables de holgura definidas en la Figura 6.4. Para minimizar esta función de error, es obligatorio tener en cuenta conjuntamente las restricciones (2) y (3).

Para ello, se aplican las condiciones de optimización de Karush-Kuhn-Tucker (KKT) [13-20], que son las condiciones necesarias de primer orden para que una solución en la programación no lineal sea óptima y que permita restricciones de desigualdad. Si introducimos multiplicadores de Lagrange $\alpha_i^+ \geq 0, \alpha_i^- \geq 0$, para todo i, el problema de optimización para identificar el modelo de regresión puede formularse de la siguiente manera [112-119, 142,143]:

$$\begin{aligned} & \max_{\alpha^+ \alpha^-} \left[\sum_{i=1}^L (\alpha_i^+ - \alpha_i^-) t_i - \varepsilon \sum_{i=1}^L (\alpha_i^+ - \alpha_i^-) - \frac{1}{2} \sum_{i=1}^L (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) x_i \cdot x_j \right] \\ & \text{s.t.} \left\{ \begin{array}{l} 0 \leq \alpha_i^+ \leq C \\ 0 \leq \alpha_i^- \leq C \\ \sum_{i=1}^L (\alpha_i^+ - \alpha_i^-) = 0 \end{array} \right\} \text{ Para todo } i \end{aligned} \quad (6.5)$$

Por lo tanto, se pueden obtener nuevas predicciones \hat{y}' como:

$$\hat{y}' = \sum_{i=1}^L (\alpha_i^+ - \alpha_i^-) \mathbf{x}_i \cdot \mathbf{x}' + b \quad (6.6)$$

En casos no lineales, tenemos que proceder a mapear los vectores de entrada de baja dimensional a través de una función no lineal $\Phi: \mathbb{R}^p \rightarrow F$, donde F es la función de espacio de Φ [112-119, 142,143]. Después de la cartografía no lineal, la función de regresión tiene la siguiente forma:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b \quad (6.7)$$

La solución de este problema de optimización cuadrática por el método dual de Lagrange [13-20] proporciona el método numérico con el valor de predicción:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b = \sum_{i=1}^L (\alpha_i^+ - \alpha_i^-) K(\mathbf{x}, \mathbf{x}_i) + b \quad (6.8)$$

Donde α_i^+, α_i^- son de nuevo los multiplicadores de Lagrange de la forma dual del problema de optimización y $K(\mathbf{x}_i, \mathbf{x}_j)$ es la función del núcleo (kernel) que satisface la condición de Mercer [112-119, 142,143], y puede describirse como:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (6.9)$$

Las funciones típicas del kernel descritas en la bibliografía [112-119, 142,143] son:

- Función de base radial (núcleo RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \quad (6.10)$$

- Núcleo polinomial:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (S \mathbf{x}_i \times \mathbf{x}_j + a)^b \quad (6.11)$$

- Núcleo sigmoide:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(S \mathbf{x}_i \times \mathbf{x}_j + a) \quad (6.12)$$

Donde a , b y σ son parámetros que definen el comportamiento del núcleo o ‘kernel’.

Resumiendo, para usar un SVM con el fin de resolver un problema de regresión con datos que no son linealmente separables, en primer lugar tenemos que elegir un kernel y parámetros relevantes esperados para mapear los datos no linealmente separables, en un espacio de características donde sea linealmente separable.

6.3.4. El algoritmo de optimización de la nube de partículas (PSO)

La optimización por nube o por enjambre de partículas (en inglés, «*particle swarm optimization*» PSO), es una técnica de optimización/búsqueda matemática. El PSO se utiliza en espacios de búsqueda con muchas dimensiones. Los métodos de PSO se atribuyeron originalmente a los investigadores Kennedy, Eberhart y Shi [120-121]. En un principio fueron concebidos para elaborar modelos de comportamiento social, como el movimiento descrito por organismos vivos en una bandada de aves o un banco de peces. El algoritmo se simplificó y resultó ser adecuado para resolver problemas de optimización. PSO permite la optimización de un problema matemático utilizando una población de soluciones candidatas, denotadas como partículas, moviéndose a través del espacio de búsqueda según reglas matemáticas que toman en cuenta la posición y velocidad de las partículas. El movimiento de cada partícula está influenciado por su mejor posición local hasta el momento, así como por las mejores posiciones globales encontradas por otras partículas a medida que las partículas viajan a través del espacio de búsqueda. La base teórica de este rendimiento es hacer que la nube de partículas converja rápidamente hacia las mejores soluciones. Además, el método PSO es una técnica metaheurística, ya que no asume hipótesis sobre el problema a optimizar y puede aplicarse en grandes espacios de soluciones examinadas.

Sea S el número de partículas en la nube, cada una de las cuales tiene una posición $\mathbf{x}_i \in \Re^n$, en el espacio de búsqueda y una velocidad $\mathbf{v}_i \in \Re^n$.

Del mismo modo, vamos a representar la posición inicial de la partícula como \mathbf{x}_i^0 y su velocidad como \mathbf{v}_i^0 , ambos elegidos al azar. Las mejores posiciones corresponden a los mejores valores de la función de ajuste o aptitud evaluados para cada partícula. Las posiciones y velocidades de cada partícula se actualizan teniendo en cuenta estos valores, como sigue:

$$\mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_1 (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_2 (\mathbf{l}_i^k - \mathbf{x}_i^k) \quad (6.13)$$

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1} \quad (6.14)$$

La velocidad de cada partícula, i , a la iteración k , se basa en tres componentes: (a) el término de velocidad en iteración k , \mathbf{v}_i^k , referido al peso de inercia constante, ω ; (b) el término denominado '*aprendizaje cognitivo*', que es la diferencia entre la mejor posición de la partícula encontrada hasta ahora (llamada \mathbf{l}_i^k , mejor localizada) y la posición actual de la partícula \mathbf{x}_i^k ; y (c) el término de '*aprendizaje social*', que es la diferencia entre la mejor posición global encontrada hasta ahora, en toda la nube o enjambre (llamada \mathbf{g}^k , mejor global) y la posición actual de la partícula \mathbf{x}_i^k . Los dos últimos términos están relacionados en la ecuación (6.19) por factores $\phi_1 = c_1 r_1$ y $\phi_2 = c_2 r_2$. En estos dos multiplicadores, c_1 y c_2 son constantes, mientras que r_1 y r_2 son números aleatorios distribuidos uniformemente en el intervalo $[0,1]$. Además, en este estudio, se ha utilizado la norma PSO 2011 [144]. Esto implica algunas mejoras con respecto a las implementaciones preliminares [120-123, 144]. Por lo tanto, los parámetros PSO se eligen aquí como:

$$\omega = \frac{1}{2 \ln 2} \quad y \quad c_1 = c_2 = 0.5 + \ln 2 \quad (6.15)$$

La topología de la nube o enjambre define cómo las partículas N_p del enjambre están conectadas entre ellas para intercambiar información con la de mejor posición global. En la PSO estándar actual cada partícula informa sólo K partículas, normalmente tres, elegidas aleatoriamente. A continuación, se ilustra un pseudo-código puro del algoritmo PSO, en el Algoritmo 1.

Algorithm 1. Pseudo-código del algoritmo PSO

Input: PSO población de partículas $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,n})^T$ para $i = 1, \dots, N_p$; N_p es el número de partículas de la muestra

Output: La mejor solución \mathbf{g} y su correspondiente valor de la función objetivo $f_{\min} = \min(f(\mathbf{x}))$

```
1: initialize_particles;  
2: eval = 0 ;  
3: while termination_condition_not_met do  
4:   for i = 1 to Np do  
5:     fi = evaluate_the_new_solution_objective_function(x_i);  
6:     eval = eval + 1 ;  
7:     if fi ≤ IBest_i then  
8:       I_i = x_i; IBEST_i = fi; // save the local best solution  
9:     end if  
10:    if fi ≤ f_min then  
11:      g = x_i; f_min = fi; // save the global best solution  
12:    end if  
13:    x_i = generate_new_solution_with_equations_13_and_14(x_i);  
14:  end for  
15: end while
```

6.3.5. Método de la Red neuronal artificial MLP

Las redes neuronales artificiales (ANNs) son un modelo computacional basado en un gran conjunto de unidades neuronales simples (neuronas artificiales), aproximadamente similar al comportamiento observado en los axones o prolongaciones de las neuronas en los cerebros biológicos [124-130].

El perceptrón de múltiples capas (MLP) es una especie de red neuronal artificial (ANN) compuesta de múltiples capas que permite resolver problemas que no son linealmente separables. De hecho, el perceptrón multicapa (MLP) consta de una capa de entrada y una capa de salida y una o más capas ocultas de nodos activadores no lineales [124, 125, 145]. Está basado en una modificación del perceptrón lineal estándar, para que utilice tres o más capas de neuronas (nodos) con funciones de activación no lineales (Fig. 6.5).

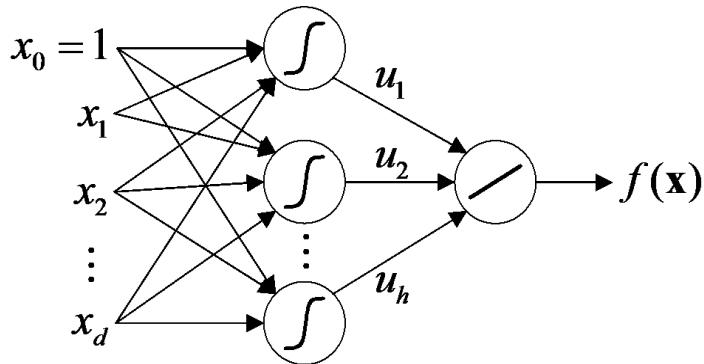


Figura 6.5. Diagrama de una red MLP con h neuronas en la capa oculta, d neuronas en la capa de entrada y una sola neurona en la capa de salida. Fuente: P.J García Nieto et al./Applied Mathematical Modelling 36 (2012) 6137-6145.

La red neuronal MLP introduce la función $\mathbf{f} : \mathbf{X} \subset \mathbb{R}^d \rightarrow \mathbf{Y} \subset \mathbb{R}^c$, que puede escribirse de la siguiente manera [124-130]:

$$\begin{aligned}\mathbf{f}(\mathbf{x}) &= \phi(\psi(\mathbf{x})) = (\phi \circ \psi)(\mathbf{x}) \\ \phi : \mathbf{X} \subset \mathbb{R}^d &\rightarrow \mathbf{U} \subset \mathbb{R}^h \\ \psi : \mathbf{U} \subset \mathbb{R}^h &\rightarrow \mathbf{Y} \subset \mathbb{R}^c\end{aligned}\tag{6.16}$$

En la ecuación (16), \mathbf{u} es el espacio de variables ocultas, denominado espacio de características. Dependiendo de la arquitectura establecida, tenemos [124-130]:

- $\psi_j(\mathbf{x}) = \psi(\mathbf{w}_j^T \mathbf{x} + w_{j0})$: ψ es la función de activación de las neuronas de la capa oculta, $\mathbf{w}_j \in \Re^d$ es el vector de parámetros de las diferentes neuronas y $w_{j0} \in \Re$ es el valor umbral. Los tres tipos de función de activación ψ son sigmoidal, logística y tangente hiperbólica.
- $\phi_j(\mathbf{u}) = \phi(\mathbf{c}_j^T \mathbf{u} + c_{j0})$: ϕ es la función de activación de las neuronas de la capa de salida, $\mathbf{c}_j \in \Re^h$ es el vector de pesos de las neuronas y $c_{j0} \in \Re$ es el valor umbral. Normalmente, ϕ es la función de identidad, la *función escalón de Heaviside*, también llamada *función escalón unitario*, o una función dicotómica.

La función afectada por la red neuronal MLP se escribe como [124-130]:

$$f(\mathbf{x}) = \sum_{j=1}^h c_j \psi_j(\mathbf{w}_j^T \mathbf{x} + w_{j0}) + c_0 \quad (6.17)$$

6.3.6. Método del árbol de regresión M5

El algoritmo original árbol modelo M5, fue inventado por Quinlan [131]. El árbol modelo M5 (M5Tree) combina un árbol de decisión convencional con la posibilidad de funciones de regresión lineal en los nodos (hoja) [132-133]. La creación del árbol de regresión M5 requiere dos fases diferentes [145-146]. Durante el primer período, el conjunto de datos se divide en subconjuntos para que se construya un árbol de decisión. El criterio de división utiliza la desviación estándar de los valores de la clase y la reducción esperada en este error. La reducción de la desviación estándar (SDR) se puede calcular como [131-133, 147]:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (6.18)$$

Donde T es el conjunto de peticiones o instancias que alcanzan este nodo, T_i son los conjuntos que resultan de dividir el nodo de acuerdo con el atributo elegido y sd es la desviación estándar de los valores de clase [146-147]. El proceso de división termina cuando los valores de clase de las instancias que alcanzan un nodo sólo varían ligeramente, es decir, cuando su desviación estándar es sólo una pequeña fracción (por ejemplo, menos del 5%) de la desviación estándar del conjunto de instancias original. Como resultado del proceso de división, los datos sobre los nodos secundarios tienen menos desviación estándar en comparación con los nodos padre y por lo tanto son hijos más puros. El árbol M5 elige la que maximiza la reducción de error esperado después de escanear todas las divisiones posibles. Esta división a menudo da lugar a una estructura arbórea extremadamente grande y puede producir un rendimiento insatisfactorio. Para resolver este problema, el árbol enorme se poda y estos nodos del árbol se sustituyen con funciones de regresión lineal en la segunda fase [148].

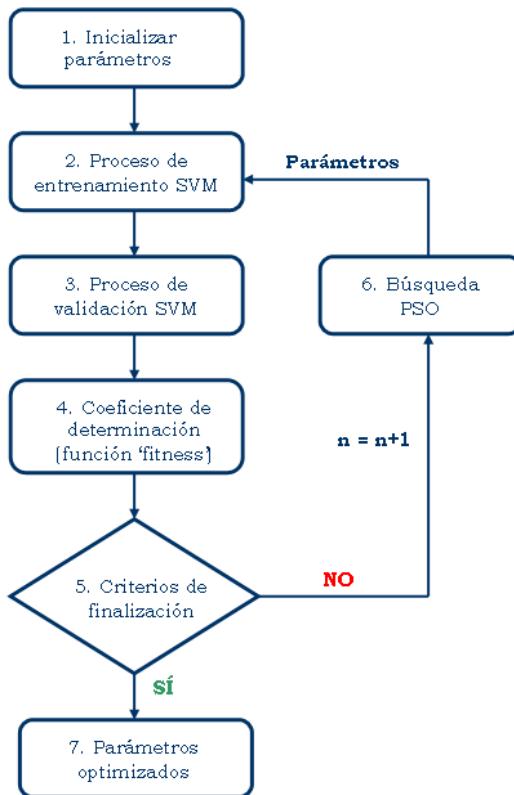
6.4. Análisis de resultados y discusión

Las variables físicoquímicas de entrada tomadas en cuenta en este trabajo de investigación, se muestran en la Tabla 6.3 [100-111, 134-136,139-141]. El número total de *variables predictoras* utilizadas para llevar a cabo la regresión del modelo híbrido basado en PSO-SVM, el enfoque MLP y el árbol modelo M5, fue 5. Además, el número total de *variables dependientes* de salida pronosticadas fue de 3: dióxido de nitrógeno (NO_2) Dióxido de azufre (SO_2) y material particulado inferior a 10 mm (PM_{10}). De hecho, hemos construido tres modelos diferentes tomando como *variables dependientes* NO_2 , SO_2 y PM_{10} , respectivamente. Adicionalmente, como variables de entrada independientes (*variables predictoras*), se consideraron las otras variables restantes que figuran en la Tabla 6.3.

Tabla 6.3. Conjunto de variables de entrada físico-químicas utilizadas en este estudio y sus nombres junto con su media y desviación estándar. Elaboración propia.

Variables de entrada	Nombre de la variable	Media	Desviación estándar
SO₂ (μg/m³)	Dióxido de azufre	26.50	9.42
NO (μg/m³)	Óxido nítrico	68.32	12.47
NO₂ (μg/m³)	Dióxido de nitrógeno	264.33	28.79
CO (mg/m³)	Monóxido de carbono	48.21	34.83
PM₁₀ (μg/m³)	Partículas de aerosol de menos de 10 μm	0.47	0.13
O₃ (μg/m³)	Ozono	1.24	0.23

Por un lado, las técnicas SVM son muy dependientes de los valores de sus hiperparámetros. Además, el número de hiperparámetros depende del tipo de núcleo (kernel) elegido. Entre ellos podemos mencionar: el factor de regularización C (ecuación 6.4); el valor de ε que define el ancho del tubo insensible (error permitido); y los hiperparámetros restantes comúnmente llamados a , b y σ . Por ejemplo, la búsqueda en la red, los algoritmos genéticos y la colonia de abejas artificiales (ABC), son métodos de optimización habitualmente utilizados para determinar los parámetros SVR apropiados para cada núcleo [118,119]. El método de búsqueda de cuadricula utilizado por la mayoría de los códigos computacionales es un método de fuerza bruta y, como tal, casi cualquier método de optimización mejora su eficiencia. En concreto, en este estudio hemos utilizado la técnica de optimización PSO [120-123] para poner a punto los parámetros SVR, para que el modelo híbrido basado en PSO-SVM se ajuste a un conjunto de datos experimentales, para predecir las variables de salida dependientes (NO_2 , SO_2 y PM_{10}), de las otras variables restantes (variables de entrada) en un análisis de la calidad del aire [134-136,139-141] con éxito. Como estimador estadístico de la bondad de ajuste, se utilizó con éxito el coeficiente de determinación R^2 . La Figura 6.5 muestra el diagrama de flujo de este nuevo modelo híbrido implementado en este estudio, basado en PSO-SVM.

**Figura 6.6.** Diagrama de flujo del nuevo modelo híbrido PSO-RBF-SVM.

Fuente: P.J. García Nieto et al. Int J Adv Manuf Technol, 2015.

Si ahora aplicamos la técnica de PSO, las llamadas partículas \mathbf{x}_i incluyen los parámetros de puesta a punto. Por ejemplo, si elegimos la función de base radial (RBF) como núcleo, entonces los componentes de la partícula se escriben como $\mathbf{x}_i = (C_i, \varepsilon_i, \sigma_i)$. De acuerdo con el algoritmo PSO, inicializamos estos parámetros aleatoriamente en la primera etapa. Para las iteraciones siguientes, las partículas evolucionan siguiendo las ecuaciones (6.13) y (6.14). A continuación, se determina el valor de la función objetivo para todas las partículas en cada iteración. Específicamente, el valor de la función objetivo es el coeficiente negativo de validación cruzada diez veces para cada partícula. Si se cumplen los criterios de determinación, el mejor \mathbf{x}_i global contiene los parámetros optimizados. Por lo tanto, la validación cruzada diez veces fue la técnica estándar utilizada aquí para encontrar el coeficiente real de determinación (R^2) [149-152]. La combinación de los hiperparámetros con la mejor eficiencia es denominada hiperparámetros óptimos [112-119,151,152].

El vector de soporte para la regresión se ha llevado a cabo con el método SVR- ϵ , utilizando la biblioteca LIBSVM [54] y los hiperparámetros se han optimizado con PSO, utilizando la versión estándar PSO 2011 [144, 154, 155]. La búsqueda en el espacio de parámetros se ha realizado teniendo en cuenta que el algoritmo SVM cambia significativamente sus resultados cuando sus parámetros aumentan o disminuyen en una potencia de 10. Por ejemplo, en el caso del núcleo RBF hemos considerado $[-6,2] \times [-10,2] \times [-6,2]$; es decir, los valores C (parámetro de regularización) varían dentro del intervalo $[10^{-6}, 10^2]$, los valores σ dentro de $[10^{-10}, 10^2]$ y los valores σ dentro de $[10^{-6}, 10^2]$ en la etapa de optimización. El criterio de parada se cumple si no hay mejoría en las diez iteraciones posteriores, en combinación con un número máximo de iteraciones igual a 500. El criterio de parada se cumple si no hay mejora en el R^2 después de diez iteraciones, en combinación con un número máximo de iteraciones igual a 500.

Tabla 6.4. Hiperparametros óptimos del modelo ajustado PSO-RBF-SVM encontrado con la técnica de optimización de enjambre de partículas (PSO), para el dióxido de nitrógeno (NO_2) en el área urbana de Oviedo. Elaboración propia.

Kernel	Valores de hiperparámetros óptimos	
<i>Lineal</i>	Factor de regularización	$C = 1.4206 \times 10^{-1}$, $\epsilon = 6.0791 \times 10^{-2}$
<i>Cuadrático</i>	Factor de regularización	$C = 6.8125 \times 10^1$, $\epsilon = 2.9739 \times 10^{-6}$, $\sigma = 1.4150 \times 10^{-2}$, $a = 4.7847 \times 10^{-2}$, $b = 2$
<i>Cúbico</i>	Factor de regularización	$C = 3.4989 \times 10^1$, $\epsilon = 3.1132 \times 10^{-7}$, $\sigma = 5.1929 \times 10^0$, $a = 8.6717 \times 10^{-2}$, $b = 3$
<i>Sigmoide</i>	Factor de regularización	$C = 1.5314 \times 10^{-1}$, $\epsilon = 1.3675 \times 10^{-5}$, $\sigma = 2.1554 \times 10^{-1}$, $a = 1.3682 \times 10^{-5}$
<i>RBF</i>	Factor de regularización	$C = 8.7135 \times 10^0$, $\epsilon = 1.3221 \times 10^{-6}$, $\sigma = 3.2069 \times 10^1$

La Tabla 6.4 muestra los hiperparámetros óptimos del modelo ajustado basado en PSO-RBF-SVM encontrado con la técnica de optimización de enjambre de partículas (PSO), para el dióxido de nitrógeno (NO_2) en el área urbana de Oviedo a escala local.

Se utilizó un ordenador iMac con una CPU Intel Core i5 de 3.2 GHz con 8 Gb de RAM y Mavericks como sistema operativo. Se cumplieron las condiciones de parada, diez iteraciones sin mejora o un máximo de 300 iteraciones, que se cumplieron después de 75 iteraciones en 4 horas y 22 minutos para el dióxido de nitrógeno (NO_2).

De forma similar, y a efectos de comparación, se ha adaptado un perceptrón multicapa (MLP) y un árbol modelo M5 a los datos experimentales correspondientes al dióxido de nitrógeno (NO_2) para predecir su valor a escala local, en el área urbana de Oviedo. En este sentido, una red neuronal artificial (ANN) se define típicamente por tres tipos de parámetros [124,130]: el patrón de interconexión entre las diferentes capas de neuronas (Figura 6.5), del proceso de aprendizaje para actualizar los pesos de las interconexiones, el factor de impulso para evitar cambios de peso oscilantes de la ANN y la función de activación que convierte la entrada ponderada de una neurona a su activación de la salida. Los parámetros óptimos de ANN para el perceptrón multicapa (MLP), en este trabajo de investigación, se muestran en la Tabla 6.5.

Tabla 6.5.Los parámetros de ANN del perceptrón multicapa (MLP) para el dióxido de nitrógeno (NO_2) en el área urbana de Oviedo. Elaboración propia.

Parámetros	Valores
<i>Número de neuronas ocultas</i>	9
<i>Tasa de aprendizaje</i>	0.1
<i>Factor de impulso</i>	1.0×10^{-10}
<i>Función de activación</i>	Función de transferencia tangente sigmoidal

La Tabla 6.6 muestra los coeficientes de determinación y correlación para los modelos basados en PSO-SVM, para los cinco núcleos (kernel) (lineal, cuadrático, cúbico, sigmoide y RBF, respectivamente), perceptrón multicapa y modelo de árbol M5, ajustados aquí para el dióxido de nitrógeno (NO_2), en el área urbana de Oviedo.

Tabla 6.6. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol M5 ajustado en este estudio Para el dióxido de nitrógeno (NO_2) en la zona urbana de Oviedo. Elaboración propia.

Modelo	Coeficientes de determinación (R^2)/coeficientes de corelación(r)
SVM- <i>Lineal</i>	0.71/0.84
SVM- <i>Cuadrático</i>	0.74/0.86
SVM- <i>Cúbico</i>	0.79/0.89
SVM- <i>Sigmoide</i>	0.71/0.84
SVM- <i>RBF</i>	0.98/0.99
Perceptrón Multicapa	0.76/0.87
Árbol modelo M5	0.79/0.89

De acuerdo con los cálculos estadísticos, el SVM con la función de núcleo RBF es el mejor modelo para estimar la concentración del dióxido de nitrógeno (NO_2) a escala local en el área urbana de Oviedo, ya que la SVM ajustada con función núcleo RBF tiene un coeficiente de determinación igual A 0,9802 y un coeficiente de correlación igual a 0,9900. Estos resultados indican una importante bondad de ajuste, es decir, se obtiene un muy buen acuerdo entre nuestro modelo y los datos observados. Por otra parte, la importancia de la clasificación de los cinco restantes variables de entrada (Tabla 6.3), con el fin de predecir el valor del dióxido de nitrógeno (NO_2) (variable de salida), en este problema complejo no lineal, se muestran en la Tabla 6.7 y Figura 6.7.

Tabla 6.7. Peso de las variables en el modelo de PSO-RBF-SVM ajustado para el valor del dióxido de nitrógeno (NO_2) en el área urbana de Oviedo. Elaboración propia.

Variable	Peso
<i>Óxido nítrico (NO)</i>	4.6232
<i>Dióxido de azufre (SO_2)</i>	2.7424
<i>Ozono (O_3)</i>	-2.7399
<i>Monóxido de carbono (CO)</i>	0.0172
<i>Partículas de aerosol menores de 10 μm (PM_{10})</i>	-0.0426

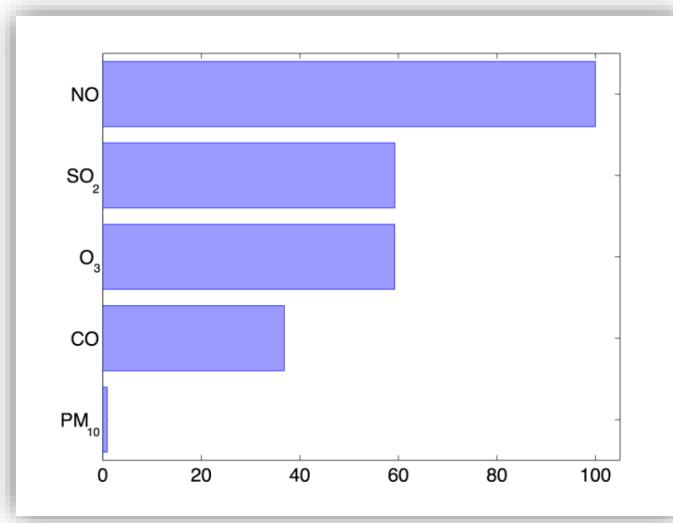


Figura 6.7. Importancia relativa de las variables de entrada para predecir el valor del dióxido de nitrógeno (NO_2) en el área urbana de Oviedo en el modelo basado en PSO-RBF-SVM. Elaboración propia.

Siguiendo la misma metodología, también se hicieron ajustes para el dióxido de azufre (SO_2) y material particulado inferior a 10 μm (PM_{10}) como variables dependientes usando el modelo basado en PSO-RBF-SVM, la técnica de perceptrón multicapa (MLP) y el árbol de regresión modelo M5, cuyos resultados se muestran a continuación. En efecto, la Tabla 6.8 muestra los hiperparámetros óptimos del modelo de PSO-RBF-SVM ajustado mediante la técnica de optimización por enjambre de partículas (PSO) para el dióxido de azufre (SO_2) en el área urbana de Oviedo, a escala local.

Tabla 6.8. Hiperparametros óptimos del modelo de PSO-RBF-VM ajustado encontrado con la técnica de optimización de enjambres de partículas (PSO) para el dióxido de azufre (SO_2) en el área urbana de Oviedo. Elaboración propia.

Kernel	Valores de hiperparámetros óptimos	
<i>Lineal</i>	Factor de regularización	$C = 1.8911 \times 10^0$, $\varepsilon = 8.9098 \times 10^{-2}$
<i>Cuadrático</i>	Factor de regularización	$C = 3.5705 \times 10^1$, $\varepsilon = 8.2453 \times 10^{-2}$, $\sigma = 2.7744 \times 10^{-1}$, $a = 2.7525 \times 10^{-2}$, $b = 2$
<i>Cúbico</i>	Factor de regularización	$C = 2.3067 \times 10^1$, $\varepsilon = 2.6629 \times 10^{-2}$, $\sigma = 1.6104 \times 10^0$, $a = 7.0757 \times 10^{-1}$, $b = 3$
<i>Sigmoide</i>	Factor de regularización	$C = 8.6787 \times 10^1$, $\varepsilon = 9.0458 \times 10^{-2}$, $\sigma = 2.0674 \times 10^{-3}$, $a = 2.1430 \times 10^{-3}$
<i>RBF</i>	Factor de regularización	$C = 3.1955 \times 10^0$, $\varepsilon = 3.1654 \times 10^{-8}$, $\sigma = 3.9841 \times 10^1$

Se cumplieron las condiciones de parada para el dióxido de azufre (SO_2), diez iteraciones sin mejora o un máximo de 300 iteraciones, después de 84 iteraciones en 5 horas y 41 minutos.

En la Tabla 6.9 de este trabajo de investigación, se presentan los parámetros de la red neuronal artificial (RNA) del perceptrón multicapa (MLP), para el dióxido de azufre (SO_2) en el área urbana de Oviedo.

Tabla 6.9. Parámetros de RNA del perceptrón multicapa (MLP) para el dióxido de azufre (SO_2) en el área urbana de Oviedo. Elaboración propia.

Parametros	Valores
<i>Número de neuronas ocultas</i>	9
<i>Tasa de aprendizaje</i>	0.1
<i>Factor de impulso</i>	0.001
<i>Función de activación</i>	Función de transferencia tangente sigmoidal

Del mismo modo, la Tabla 6.10 muestra los coeficientes de determinación y correlación de los modelos basados en PSO-SVM para los cinco núcleos (lineal, cuadrático, cúbico, sigmoide y RBF, respectivamente), perceptrón multicapa y modelo de árbol de regresión M5, ajustado aquí para el dióxido de azufre (SO_2) en la zona urbana de Oviedo.

Tabla 6.10. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol de regresión M5 ajustado en este estudio para el dióxido de azufre (SO_2) en la zona urbana de Oviedo. Elaboración propia.

Modelo	Coefficients of determination (R^2)/correlation coefficients (r)
SVM-Lineal	0.50/0.71
SVM-Cuadrático	0.64/0.80
SVM-Cúbico	0.69/0.83
SVM-sigmoide	0.50/0.71
SVM-RBF	0.94/0.97
Perceptrón Multicapa	0.67/0.82
Árbol modelo M5	0.65/0.81

De acuerdo con los cálculos estadísticos, el SVM con la función de núcleo RBF es el mejor modelo para estimar la concentración del dióxido de nitrógeno (NO_2) en el área urbana de Oviedo a escala local, ya que la SVM ajustada con función núcleo RBF tiene un coeficiente de determinación igual a 0,9499 y un coeficiente de correlación igual a 0,9746. Estos resultados indican una importante bondad de ajuste, es decir, se obtiene un muy buen acuerdo entre nuestro modelo y los datos observados. Por otra parte, la importancia de la clasificación de las cinco variables de entrada restantes (Tabla 6.3) con el fin de predecir el valor (variable de salida) del dióxido de nitrógeno (NO_2) en este problema complejo no lineal, se muestra en la Tabla 6.11 y Figura 6.8.

Tabla 6.11. Peso de las variables en el modelo de PSO-RBF-SVM ajustado para el valor de dióxido de azufre (SO_2) en el área urbana de Oviedo. Elaboración propia.

Variable	Peso
<i>Óxido nítrico (NO)</i>	3.3559
<i>Dióxido de nitrógeno (NO_2)</i>	2.5972
<i>Monóxido de carbono (CO)</i>	1.6660
<i>Ozono (O_3)</i>	-1.6075
<i>Partículas de aerosol menores de μm (PM_{10})</i>	1.4966

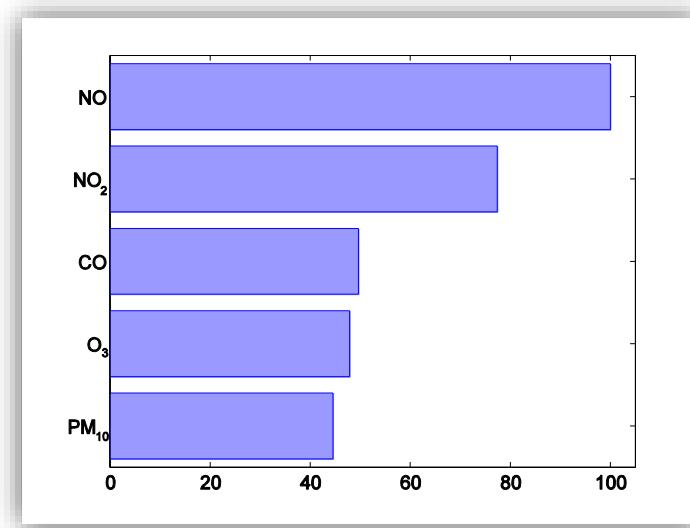


Figura 6.8. Importancia relativa de las variables de entrada para predecir el valor del dióxido de azufre (SO_2) en la zona urbana de Oviedo en el modelo PSO-RBF-SVM ajustado. Elaboración propia.

A continuación, la Tabla 6.12 muestra los hiperparámetros óptimos del modelo de PSO-RBF-SVM ajustado que se encuentra con la técnica de optimización de enjambres de partículas (PSO) para la materia particulada inferior a 10 μm (PM_{10}) en el área urbana de Oviedo a escala local.

Tabla 6.12. Hiperparametros óptimos del modelo de PSO-RBF-SVM ajustado encontrado con la técnica de optimización de enjambre de partículas (PSO) para partículas de menos de 10 μm (PM_{10}) en el área urbana de Oviedo. Elaboración propia.

Kernel	Valores de los hiperparámetros óptimos	
<i>Lineal</i>	Factor de regularización	$C = 1.3130 \times 10^{-1}$, $\varepsilon = 1.3493 \times 10^{-1}$
<i>Cuadrático</i>	Factor de regularización	$C = 6.6965 \times 10^1$, $\varepsilon = 8.0252 \times 10^{-2}$, $\sigma = 1.2451 \times 10^0$, $a = 7.5358 \times 10^{-1}$, $b = 2$
<i>Cúbico</i>	Factor de regularización	$C = 8.8805 \times 10^2$, $\varepsilon = 5.7101 \times 10^{-2}$, $\sigma = 3.2484 \times 10^{-1}$, $a = 9.6975 \times 10^{-1}$, $b = 3$
<i>Sigmide</i>	Factor de regularización	$C = 3.7271 \times 10^1$, $\varepsilon = 1.2589 \times 10^{-7}$, $\sigma = 2.6951 \times 10^{-3}$, $a = 1.1634 \times 10^{-5}$
<i>RBF</i>	Factor de regularización	$C = 3.7120 \times 10^0$, $\varepsilon = 1.1093 \times 10^{-9}$, $\sigma = 5.7759 \times 10^1$

Las condiciones de parada, para las partículas de menos de 10 μm (PM_{10}), diez iteraciones sin mejora o un máximo de 300 iteraciones, se cumplieron después de 69 iteraciones y 4 horas y 6 minutos.

En este trabajo de investigación se presentan en la Tabla 6.13 los parámetros de la RNA del perceptrón multicapa (MLP) para el aerosol de menos de 10 μm (PM_{10}) en el área urbana de Oviedo.

Tabla 6.13. Los parámetros de la RNA del perceptrón multicapa (MLP) para el aerosol de menos de 10 μm (PM_{10}) en el área urbana de Oviedo. Elaboración propia.

Parámetros	Valores
<i>Número de neuronas ocultas</i>	11
<i>Tasa de aprendizaje</i>	0.1
<i>Factor de impulso</i>	0.1
<i>Función de activación</i>	Función de transferencia tangente sigmoidal

De forma similar, la Tabla 6.14 muestra los coeficientes de determinación y correlación para los modelos basados en PSO-SVM para los cinco núcleos ‘kernel’ (lineal, cuadrático, cúbico, sigmoide y RBF, respectivamente), perceptrón multicapa y modelo de árbol de regresión M5 ajustado aquí para el aerosol menor que 10 μm (PM_{10}) en el área urbana de Oviedo.

Tabla 6.14. Coeficiente de determinación (R^2) y coeficiente de correlación (r) para los modelos híbridos basados en PSO-SVM (con núcleos lineales, cuadráticos, cúbicos, sigmoides y RBF), perceptrón multicapa (MLP) y modelo de árbol de regresión M5 ajustado en este estudio Para el aerosol de menos de 10 μm (PM_{10}) en el área urbana de Oviedo. Elaboración propia.

Modelo	Coeficientes de determinación (R^2)/coeficientes de correlación (r)
<i>SVM-Lineal</i>	0.12/0.35
<i>SVM-Cuadrático</i>	0.20/0.45
<i>SVM-Cúbico</i>	0.36/0.60
<i>SVM-sigmoide</i>	0.11/0.33
<i>SVM-RBF</i>	0.85/0.92
<i>Perceptrón Multicapa</i>	0.32/0.57
<i>Árbol modelo M5</i>	0.33/0.58

Además, de acuerdo con los cálculos estadísticos, el SVM con la función núcleo o ‘kernel’ RBF es el mejor modelo para estimar la concentración de la materia particulada inferior a 10 μm (PM_{10}) en el área urbana de Oviedo a escala local; ya que el SVM ajustado con función núcleo o ‘kernel’ RBF tiene un coeficiente de determinación igual a 0,8458 y un coeficiente de correlación igual a 0,9197.

Por otra parte, la importancia de la clasificación de las cinco variables de entrada restantes (Tabla 6.3) con el fin de predecir el valor (variable de salida) de la materia particulada inferior a 10 μm (PM_{10}), en este problema complejo no lineal, se muestra en la Tabla 6.15 y Figura 6.9.

Tabla 6.15. Pesos de las variables del modelo de PSO-RBF-SVM ajustado para el material particulado inferior a 10 μm (PM_{10}) en el área urbana de Oviedo. Elaboración propia.

Variable	Peso
Óxido nítrico (NO)	2.2389
Dióxido de azufre (SO_2)	1.7814
Monóxido de carbono (CO)	1.4004
Dióxido de nitrógeno (NO_2)	0.6798
Ozono (O_3)	-0.5785

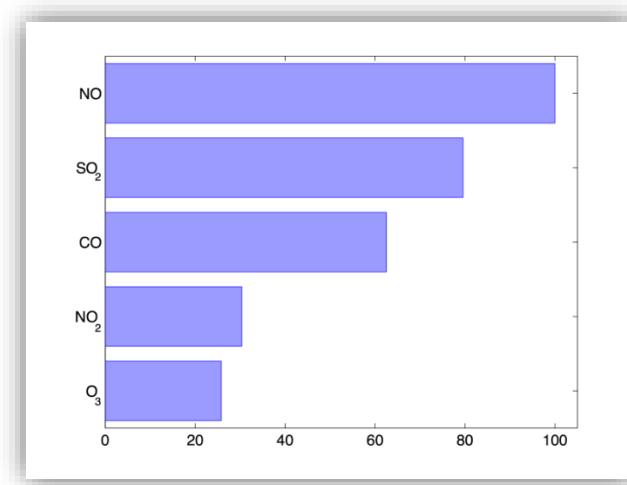


Figura 6.9. Importancia relativa de las variables de entrada para predecir el valor de las partículas (PM_{10}) en el área urbana de Oviedo en el modelo PSO-RBF-SVM ajustado. Elaboración propia.

A partir de los resultados representados en la Tabla 6.7 y en la Figura 6.7, es posible observar que las variables más importantes para la predicción del NO₂ (variable de salida) según el modelo PSO-RBF-SVM están en orden jerárquico: NO, SO₂, O₃, CO y PM₁₀. La influencia de la variable PM₁₀ fue insignificante según los cálculos. La variable más significativa en la predicción de NO₂ es el óxido nítrico (NO). Este resultado es lógico ya que el dióxido de nitrógeno (NO₂) típicamente surge a través de la oxidación por el oxígeno en el aire del óxido nítrico (NO). El dióxido de nitrógeno se forma en la mayoría de los procesos de combustión usando aire como oxidante.

De manera similar, los resultados mostrados en la Tabla 6.11 y la Figura 6.8 indican que las variables más importantes para la predicción del SO₂ (variable de salida) son: NO, NO₂, CO, O₃ y PM₁₀. De nuevo, la influencia de la variable PM₁₀ fue la más pequeña según los cálculos. El dióxido de azufre (SO₂) es el producto de la combustión de azufre o de la combustión de materiales que contienen azufre. Además, las emisiones de dióxido de azufre son precursoras de la lluvia ácida y de las partículas atmosféricas.

A partir de los resultados mostrados en la Tabla 6.15 y Figura 6.9, las variables más importantes para la predicción de la materia particulada menor que 10 μm (PM₁₀) (variable de salida) son: NO, SO₂, CO, NO₂ y O₃. La influencia de las variables NO₂ y O₃ fue despreciable según los cálculos. Algunas partículas se producen de forma natural, originadas en volcanes, tormentas de polvo, incendios de bosques y praderas, vegetación viva y rocío marino. Las actividades humanas, como la quema de combustibles fósiles en vehículos, centrales eléctricas y diversos procesos industriales, también generan cantidades significativas de partículas (aerosoles antropogénicos). De esta manera, las partículas secundarias derivan de la oxidación de gases primarios tales como azufre y óxidos de nitrógeno en ácido sulfúrico (líquido) y ácido nítrico (gaseoso). Los precursores de estos aerosoles (es decir, los gases de los que se originan) pueden tener un origen antropogénico (de combustión fósil o combustión de carbón) y un origen biogénico natural.

Por último, este trabajo de investigación permite predecir las concentraciones de NO₂ de 2013 a 2015 de acuerdo con las concentraciones experimentales reales de NO₂ observadas, utilizando el modelo basado en PSO-RBF-SVM, con gran exactitud y éxito. De hecho, la Figura 6.10 muestra la comparación entre los valores de NO₂ observados y predichos usando el modelo de árbol de regresión M5 (Figura 6.10 a), MLP (Figura 6.10 b) y el modelo basado en PSO-SVM con función núcleo o kernel RBF (Figura 6.10 c). Es necesario el uso de un modelo SVM con función núcleo o kernel RBF, con el fin de lograr el enfoque más eficaz frente a las no linealidades presentes en este problema de regresión. Obviamente, estos resultados coinciden de nuevo con el criterio de resultado de 'bondad de ajuste' (R^2) para el modelo basado en PSO-SVM con una función de núcleo o kernel RBF, que sido el mejor ajuste.

De forma similar, este estudio también fue capaz de predecir las concentraciones de SO₂ y PM₁₀ de 2013 a 2015 de acuerdo con las concentraciones experimentales reales de SO₂ y PM₁₀ observadas usando el modelo basado en PSO-RBF-SVM con gran exactitud y éxito. De hecho, la Figuras 6.11 y la Figura 6.12 muestran, a continuación, la comparación entre los valores de SO₂ y PM₁₀ observados y predichos utilizando el árbol de regresión modelo M5 (Figura 6.11a y Figura 6.12a), el MLP (Figura 6.11b y Figura 6.12b) y el modelo basado en PSO-SVM con función núcleo RBF (Figura 6.11c y Figura 6.12c), respectivamente. Es obligatorio el uso de un modelo SVM con función núcleo o kernel RBF con el fin de lograr el enfoque más eficaz a las no linealidades presentes en este problema de regresión. Obviamente, estos resultados coinciden de nuevo con el criterio de resultado de 'bondad de ajuste' (R^2) para que el modelo basado en PSO-SVM, con una función de núcleo o kernel RBF, haya sido el mejor ajuste.

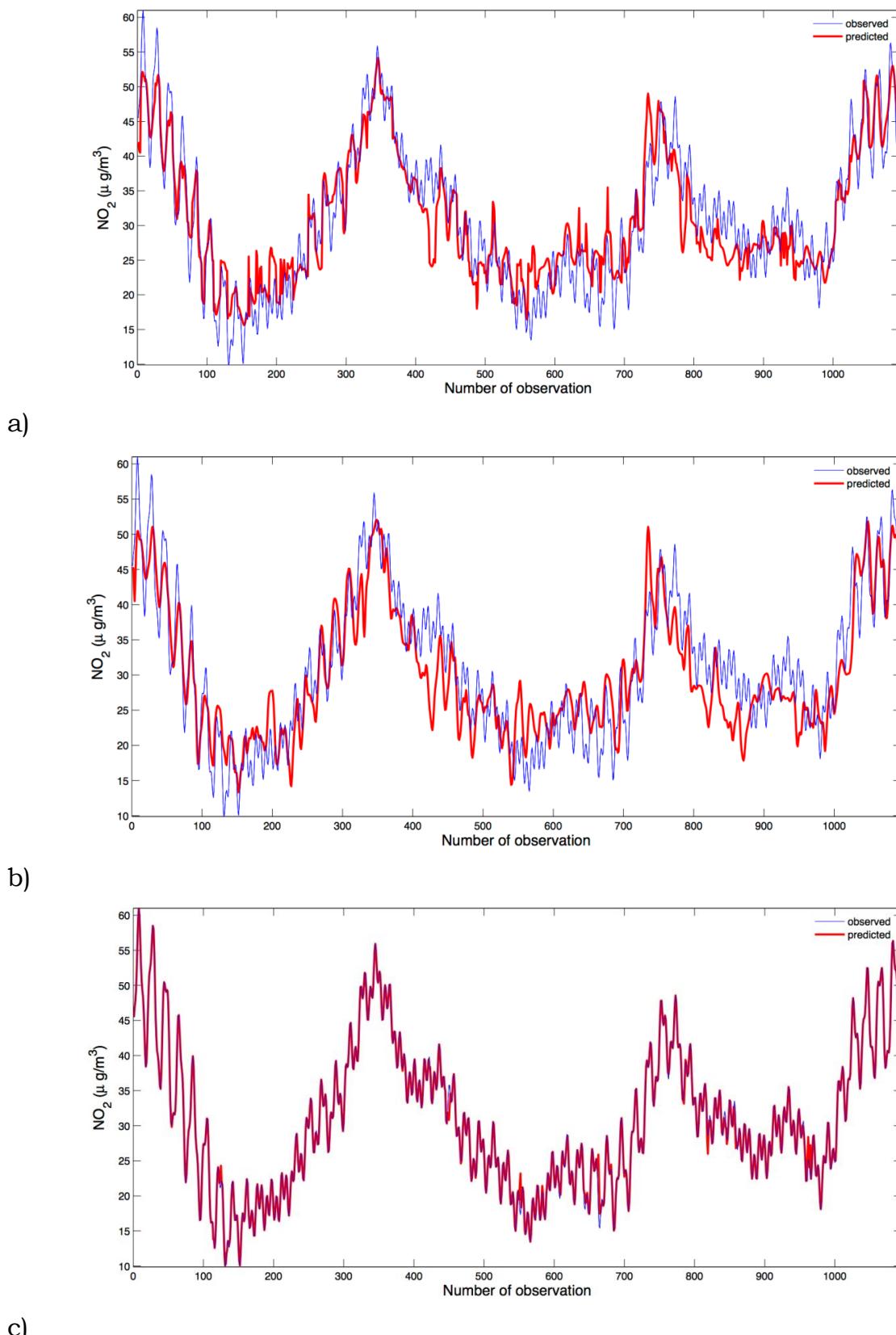


Figura 6.10. Comparación entre los valores de NO₂ observados y predichos por el árbol de regresión modelo M5, enfoque MLP y el modelo basado en PSO-SVM: a) árbol de regresión modelo M5 ($R^2=0.75$); b) red MLP ($R^2=0.80$); y c) Modelo PSO-SVM con núcleo RBF ($R^2=0.9802$). Elaboración propia.

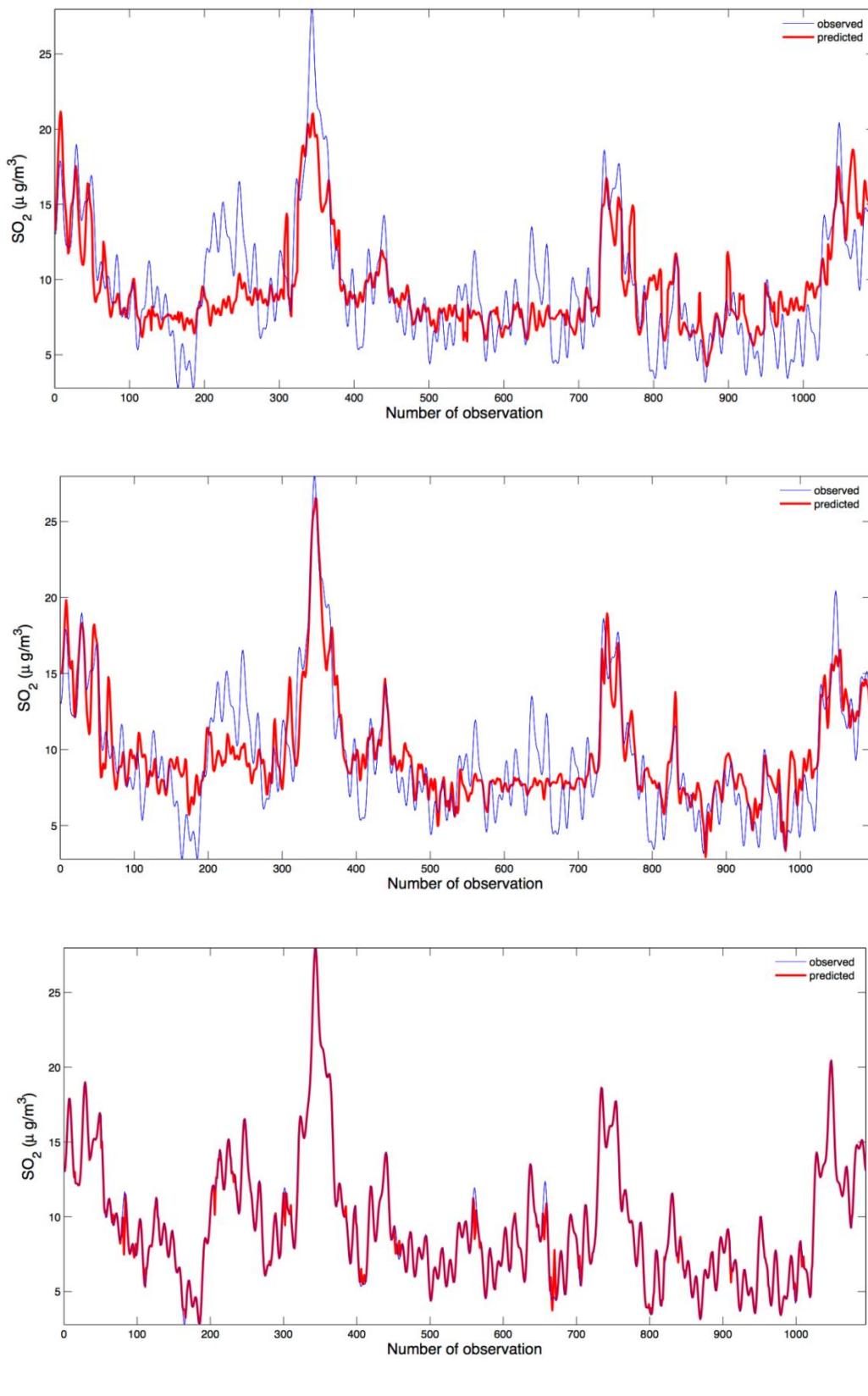


Figura 6.11. Comparación entre los valores de SO_2 observados y predichos por el árbol de regresión modelo M5, el enfoque MLP y el modelo basado en PSO-SVM: a) árbol modelo M5 ($R^2 = 0.75$); b) red MLP ($R^2 = 0.80$); Y c) Modelo PSO-SVM con núcleo RBF ($R^2 = 0.9499$). Elaboración propia.

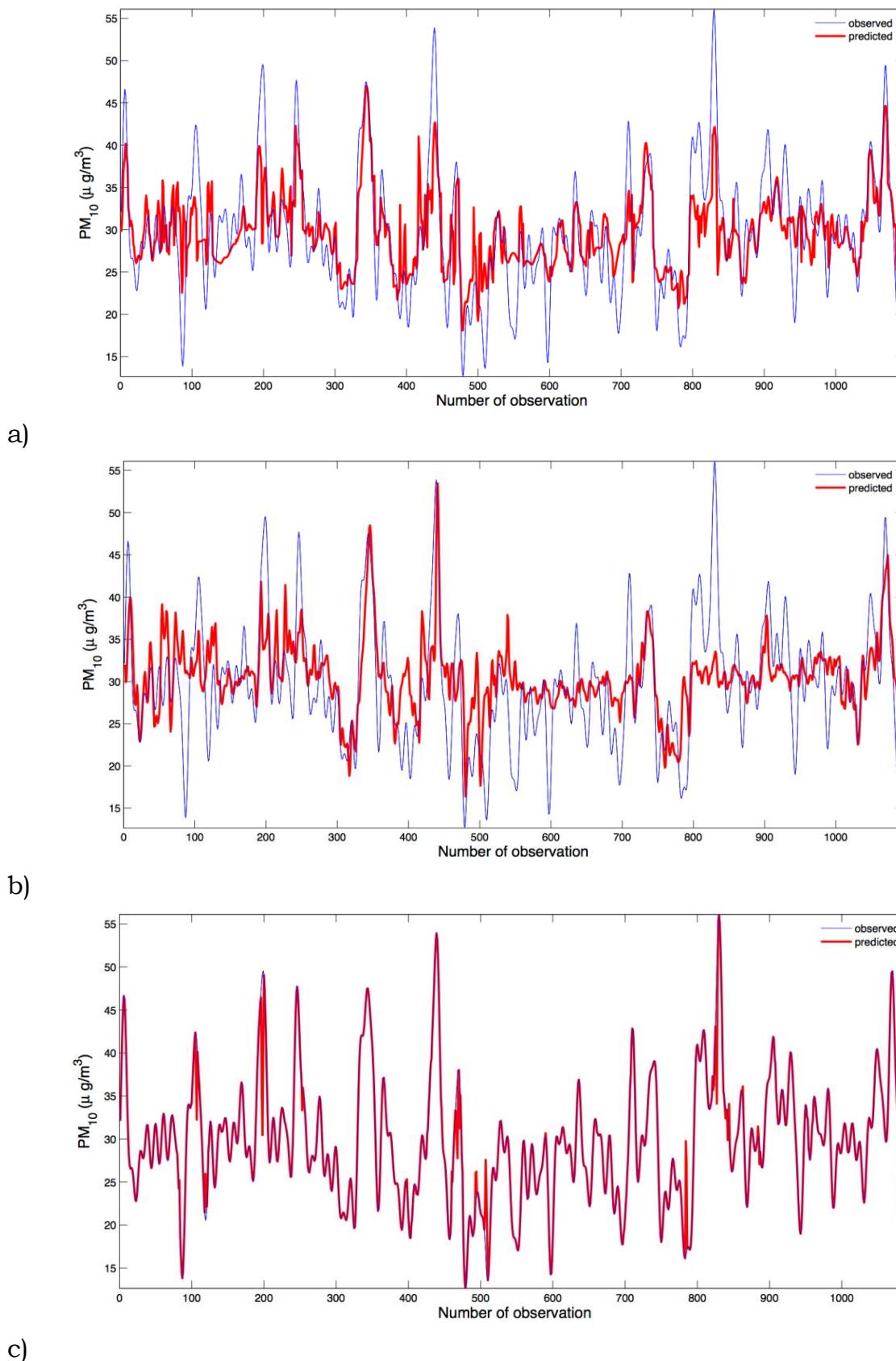


Figura 6.12. Comparación entre las concentraciones de PM₁₀ observadas y predichas por el árbol de regresión modelo M5, el enfoque MLP y el modelo basado en PSO-SVM: a) árbol modelo M5 ($R^2 = 0.75$); b) red MLP ($R^2 = 0.80$); Y c) Modelo PSO-SVM con núcleo RBF ($R^2 = 0.8458$). Elaboración propia.

6.5. Conclusiones

A pesar del éxito generalizado, los retos para la gestión de la calidad del aire siguen sin resolverse hoy en día. Sobre la base de los resultados experimentales y numéricos, los principales hallazgos de este trabajo de investigación pueden resumirse como sigue:

- *En primer lugar*, hoy todos los gobiernos anuncian planes para mejorar la calidad del aire en las ciudades y minimizar el impacto en las empresas. Sin embargo, las estrategias de reducción de emisiones para evitar litigios y satisfacer al público y a otras partes interesadas son muy difíciles de llevar a cabo en la práctica, requiriendo tal vez años para su implementación. Además, las técnicas de diagnóstico comúnmente utilizadas, basadas en métodos tradicionales (por ejemplo, la monitorización de contaminantes a través de estaciones automáticas) son caras, tanto desde el punto de vista material como humano. En consecuencia, es necesario el desarrollo de técnicas alternativas de diagnóstico. En este sentido, el nuevo método híbrido basado en PSO-SVM con una función de núcleo o ‘kernel’ RBF, empleada en este trabajo de investigación, constituye una buena elección para evaluar la calidad del aire en las ciudades a escala local.
- *En segundo lugar*, se confirmó la hipótesis de que el diagnóstico de la calidad del aire en el área urbana de Oviedo, a escala local, puede modelarse con precisión utilizando un modelo híbrido basado en PSO-SVM con una función de núcleo o kernel RBF.
- *En tercer lugar*, se desarrolló con éxito un modelo híbrido PSO-SVM con una función de núcleo o ‘kernel’ RBF, para predecir las concentraciones de NO₂, SO₂ y PM₁₀ en función de los contaminantes de operación de entrada medidos, para reducir los costos en la evaluación de la calidad del aire, en la zona urbana de Oviedo.

- *En cuarto lugar*, se obtuvieron altos coeficientes de determinación iguales a 0,9802, 0,9499 y 0,8458 cuando se aplicó este modelo híbrido PSO-SVM con una función de núcleo o ‘kernel’ RBF, al conjunto de datos experimentales, correspondientes respectivamente a los contaminantes presentes en el área urbana de Oviedo. De hecho, los resultados previstos para este modelo han demostrado ser consistentes con el conjunto de datos históricos, de los valores reales observados entre los años 2013 y 2015, para dichos contaminantes (Figura 6.10, Figura 6.11 y Figura 6.12).
- *En quinto lugar*, se estableció el orden de importancia de las variables de entrada implicadas en la predicción de las concentraciones de NO₂, SO₂ y PM₁₀. Este es uno de los principales hallazgos en este estudio.
- *En sexto lugar*, se estableció la influencia de la configuración de los parámetros del kernel de las SVM, sobre el rendimiento de la regresión del valor de la calidad del aire.
- *Por último*, los resultados de esta investigación sobre el desarrollo de modelos de concentraciones de contaminantes a escala local, son una valiosa herramienta para proyectos de mitigación de lluvia ácida y para la investigación de los efectos de las partículas en la salud humana. Por otra parte, existe un creciente interés por el uso de modelos matemáticos con buenas propiedades físicas para entender el comportamiento de los contaminantes en la atmósfera, con el fin de mejorar la calidad del aire y reducir el número de muertes. Los resultados verifican que el método de regresión de PSO-SVM híbrido mejora significativamente la capacidad de generalización alcanzable sólo con el regresor basado en SVM. En este sentido, este modelo puede ser ensamblado dentro de otros modelos de la atmósfera más generales.

En resumen, esta metodología innovadora podría aplicarse a otras ciudades o lugares con tipos de contaminantes, similares o diferentes, con éxito; pero siempre, es obligatorio tener en cuenta las especificidades de cada lugar. Por consiguiente, un modelo eficaz basado en PSO-SVM es una solución práctica al problema de la determinación de la calidad del aire en las ciudades.

Esta técnica hace posible etiquetar las zonas de aire limpio en cada ciudad, como áreas donde el problema de la calidad del aire es menos grave. Además, este trabajo presenta ejemplos de aplicaciones reales y explicaciones simples de cálculo estadístico para la selección de los modelos mejor ajustados.

Capítulo VII

APORTACION Y EXPERIENCIAS

Difusión, contribución
y posibles líneas de
investigación

7. Difusión de esta tesis, contribución y posibles líneas de investigación

7.1. Contribuciones de esta tesis

Es evidente que se han logrado grandes resultados en materia de control medioambiental y que han sido implementadas numerosísimas acciones en este sentido. No obstante, a pesar del éxito generalizado, los retos para la gestión de la calidad del aire siguen sin resolverse en nuestros días. En este sentido, todas las administraciones anuncian planes para mejorar la calidad del aire en las ciudades, y, minimizar el efecto de las emisiones que se originan en las empresas. Sin embargo, como ya hemos dicho, las estrategias de reducción de emisiones para evitar potenciales litigios y satisfacer al público en general, así como al resto de las partes interesadas, son muy difíciles de abordar en la práctica, requiriendo tal vez años para su implementación. Esto unido a que las técnicas de diagnóstico basadas en métodos tradicionales son muy costosas, desde el punto de vista material y humano, hace indispensable desarrollar técnicas alternativas de diagnóstico.

Hemos visto en este trabajo, que se puede modelar con precisión el diagnóstico de la calidad del aire en un área urbana a escala local, utilizando un modelo híbrido basado en PSO-SVM con una función de núcleo o kernel RBF; obteniendo muy altos coeficientes de determinación, cuando se aplicó este modelo híbrido PSO-SVM con función de núcleo o ‘kernel’ RBF, a un determinado conjunto de datos experimentales. Además, uno de los principales hallazgos en este estudio ha sido la posibilidad de establecer el orden de importancia de las variables de entrada implicadas en la predicción de las concentraciones de NO₂, SO₂ y PM₁₀. Así, estaríamos en condiciones de afirmar, que los resultados de esta investigación sobre el desarrollo de modelos de concentraciones de contaminantes, a escala local, son una valiosa herramienta para proyectos de mitigación de lluvia ácida y para la investigación de los efectos de las partículas en la salud humana.

Por todo ello, esta metodología innovadora podría emplearse en otras ciudades o lugares con tipos de contaminantes, similares o diferentes, con éxito. Esta técnica hace posible etiquetar las zonas de aire limpio en cada ciudad, como áreas donde el problema de la calidad del aire es menos grave. Además, este trabajo presenta ejemplos de aplicaciones reales y explicaciones simples de cálculo estadístico para la selección de los modelos mejor ajustados.

7.1.1. Publicación en revista científica

Los principales hallazgos de este trabajo de investigación, que se comentan de forma pormenorizada en el capítulo VI., Aplicación al caso de estudio: Modelado y análisis de la calidad de aire en la ciudad de Oviedo (norte de España), mediante los enfoques PSO-SVM, red neuronal MLP y árbol de regresión M5, han sido divulgados mediante la publicación en la revista científica Environmental Modeling & Assessment – Springer:

P.J. García Nieto, E. García-Gonzálo, A. Bernardo Sánchez & A. A. Rodríguez Miranda: ‘Air Quality Modeling Using the PSO-SVM Based Approach, MLP Neural Network, and M5 Model Tree in the Metropolitan Area of Oviedo (Northerm Spain).

ISSN 1420-2026. Environ Model Assess. DOI 10.1007/s10666-017-9578-y

La revista científica Environmental Modeling & Assessment sirve de puente entre la comunidad científica y los problemas ambientales clave, facilitando a los responsables argumentos para la toma de decisiones políticas y normativas relevantes. Se ofrecen artículos revisados por pares y proporciona un foro para publicar una descripción matemática completa de modelos ambientales junto con el análisis y los supuestos que los acompañan. La revisión por pares (en inglés: peer review) o arbitraje, es un método para validar trabajos escritos con el fin de evaluar su calidad, originalidad, factibilidad y rigor científico, antes de su publicación o aprobación. Este método deja abierto el trabajo al escrutinio, y la anotación o modificación, por autores de rango semejante o superior al del autor.

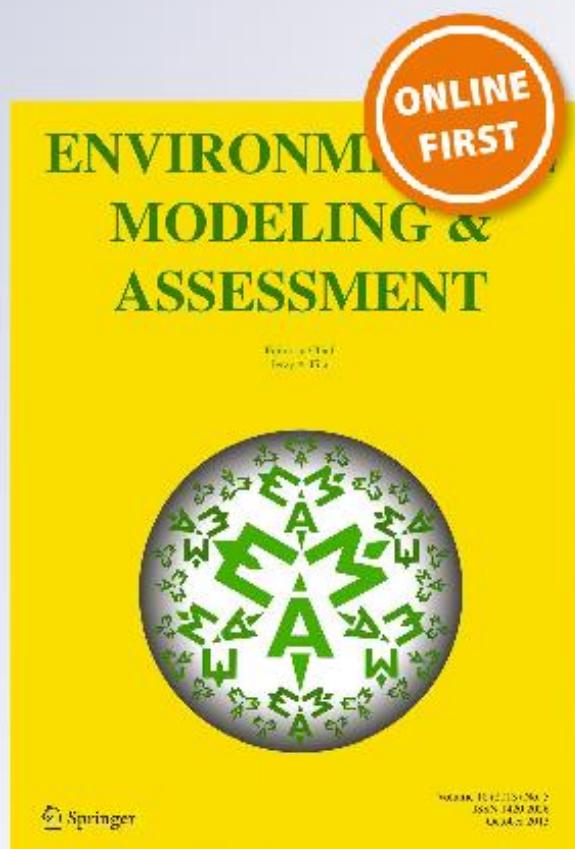
Air Quality Modeling Using the PSO-SVM-Based Approach, MLP Neural Network, and M5 Model Tree in the Metropolitan Area of Oviedo (Northern Spain)

**P. J. García Nieto, E. García-Gonzalo,
A. Bernardo Sánchez & A. A. Rodríguez
Miranda**

**Environmental Modeling &
Assessment**

ISSN 1420-2026

Environ Model Assess
DOI 10.1007/s10666-017-9578-y



 Springer

7.1.2. Participación en la IX Conferencia Internacional de Aprovechamiento de Recursos Minerales CINAREM'17

Los resultados obtenidos y los descubrimientos que han sido comentados en el capítulo VI y en especial los descritos en punto 6.3, *Análisis de resultados y discusión*, centrándose en la parte correspondiente a la aplicación del *nuevo método híbrido PSO-SVM*, respecto al conjunto de datos experimentales expuestos en este trabajo de investigación han sido presentados a la comunidad científica, mediante exposición oral y con el soporte gráfico preciso, en la Conferencia Internacional:



Este evento se desarrolló durante los días 14 al 16 de noviembre de 2017, en el Instituto Superior Minero Metalúrgico de Moa, provincia de Holguín, República de Cuba.

El Instituto Superior Minero Metalúrgico de Moa, "Dr. Antonio Núñez Jiménez" (ISMMM) cuenta con más de 40 años de experiencia en la actividad docente investigadora. Es el centro rector en la enseñanza de la Geología, la Minería y la Metalurgia en el país.

En la IX Conferencia Internacional de Aprovechamiento de Recursos Minerales "CINAREM", han participado Profesores, investigadores, profesionales, estudiantes y empresarios de la comunidad internacional. El objetivo principal de dicho evento, ha sido el de dar a conocer los avances científicos de alcance internacional en diversas áreas del saber, intercambiar experiencias, crear una red de colegas y egresados recientes e involucrar a las nuevas generaciones en este campo del conocimiento. En este evento se han incluido paneles, conferencias magistrales, mesas redondas y posters.



INSTITUTO SUPERIOR
MINERO METALÚRGICO DE MOA
Dr. ANTONIO NÚÑEZ JIMÉNEZ

IX Conferencia Internacional de Aprovechamiento de Recursos Minerales
CINAREM'2017

Moa, 25 de septiembre de 2017
“Año 59 de la Revolución”

Dr. Antonio Bernardo Sánchez
Ing. Alejandro Rodríguez Miranda
Dr. Paulino García Nieto
Dra. Esperanza García Gonzalo
Dra. Marta Menéndez Fernández

Ref.- aceptación de trabajo para presentar en CINAREM'2017

En nombre del Comité Organizador de la IX Conferencia Internacional de Aprovechamiento de Recursos Minerales “CINAREM’2017” a celebrarse en el Instituto Superior Minero Metalúrgico de Moa durante los días del 14 al 16 de noviembre del presente año me dirijo a ustedes para comunicarles que el trabajo titulado **TÉCNICAS MODERNAS DE MINERÍA DE DATOS PARA EL TRATAMIENTO DE PROBLEMAS RELACIONADOS CON LA CONTAMINACIÓN AMBIENTAL** ha sido aprobado por el Comité Científico para su presentación en el referido evento. La casa de altos estudios de la comunidad minera de Moa se alista para acoger a delegados de diferentes nacionalidades y se sentirá orgullosa de contar con su presencia; la que sin dudas prestigiará la ocasión.

A la espera de su confirmación, les saluda cordialmente,

Josefa Mestre Lamoru
Directora de Relaciones Internacionales
Secretaria Ejecutiva Comité Organizador
“CINAREM’2017”





INSTITUTO SUPERIOR
MINERO METALÚRGICO DE MOA
Dr. ANTONIO NÚÑEZ JIMÉNEZ

CINAREM'17

Se otorga el presente

CERTIFICADO

A: *Antonio Bernardo Sánchez*

Por su participación como conferencista

Conferencia Magistral

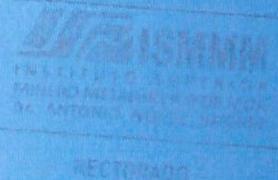
"TÉCNICAS MODERNAS DE MINERÍA DE DATOS PARA EL TRATAMIENTO
DE PROBLEMAS RELACIONADOS CON LA CONTAMINACIÓN AMBIENTAL"

Dado en Moa, a los 17 días del mes de noviembre de 2017

"Año 59 de la Revolución"

Dr. C. Ángel Oscar Columbié Navarro

Rector



Al margen de las contribuciones expuestas, se contemplan otras acciones experimentales que aún están en fase de elaboración; si bien, con el trabajo realizado y expuesto en este documento, se consideran alcanzados y suficientemente demostrados todos los objetivos formulados en el capítulo I, de esta tesis.

7.2. Posibles líneas de investigación

Tal y como expusimos anteriormente el modelo híbrido basado en PSO-SVM que se ha desarrollado en esta investigación, puede ser ensamblado dentro de otros modelos de la atmósfera más generales. Por ello, seguidamente se detallan las líneas generales de varios trabajos de investigación susceptibles de desarrollar mediante las técnicas descritas en esta tesis doctoral.

7.2.1 Mejorar la ubicación de los ‘Árboles urbanos’ inteligentes para reducir la contaminación del aire

Los denominados ‘Árboles urbanos’ son una estructura inteligente que tiene las propiedades de los árboles tradicionales. Según sus creadores, estas estructuras purifican el aire eliminando las partículas perjudiciales para la salud, tales como el óxido de nitrógeno y el ozono, de manera semejante a como lo harían un gran conjunto de árboles. En base a los resultados obtenidos en testeos y en despliegues del ‘CityTree’ en ciudades como Oslo (Noruega) y Dresden (Alemania), cada una de las estructuras allí emplazadas es capaz de reducir la contaminación local en un área de 50 metros a su alrededor, en un porcentaje que supera el 30%. La estructura tiene conexión wifi e incorpora un sistema de algoritmos que analiza los flujos de partículas. Además, genera datos de calidad del aire y del clima de la urbe, que pueden ser incorporados a las bases de datos abiertos de la ciudad, para su análisis a la hora de tomar decisiones relativas a cuestiones climáticas [CIO, Zhengliang Wu. 2016].

Mediante la integración del algoritmo desarrollado en esta tesis, adaptándolo y realizando las mejoras precisas, se podría evaluar la mejor ubicación y la orientación más conveniente para su emplazamiento.

Adaptando la herramienta diseñada en este trabajo, se podrían obtener modelos a partir de las diferentes variables de entrada, en las distintas áreas de inmersión históricamente contaminadas. El sistema desarrollado por Zhengliang Wu, CIO at Green City Solutions, podría completarse con esta herramienta basada en PSO-SVM que contribuiría a inferir las acciones posibilitando mejorar los tiempos de implementación de acciones.

7.2.2. Mejorar la precisión del cálculo los niveles de emisiones contaminantes en las ciudades, aplicando Técnicas de minería de datos al Big Data obtenido de una red de dispositivos móviles

El tratamiento y análisis del ‘Big Data’ mediante técnicas de minería de datos con la finalidad de obtener información, genera argumentos fundados para la toma de decisiones; especialmente en el ámbito de las políticas públicas. En este sentido, se está comprobando día a día que estas técnicas son muy versátiles y proporcionan ventajas comparativas a la hora de optimizar los procesos. Por otro lado, también sirve para realizar un análisis convencional más profundo, por lo que es un excelente recurso para extraer valor del ‘Big data’ y entender mejor los resultados obtenidos.

Un estudio con fines medioambientales, llevado a cabo en la ciudad alemana de Núremberg, otorga un nuevo valor al uso de los datos. Se trata de aprovechar los datos conseguidos de una red de dispositivos móviles, para estimar los niveles de contaminación del aire en las ciudades, a partir del análisis de los patrones de movilidad de los ciudadanos que portan y utilizan sus dispositivos móviles. De esta forma, las personas que circulan por la ciudad, en cada una de las posibles modalidades de transporte, actúan como ‘sensores’ que proporcionan una información de gran utilidad.

7.2.2.1. Desarrollo del estudio

En Alemania los datos procedentes de la red móvil, son generados por más de 44 millones de clientes. Cada vez que uno de ellos utiliza su teléfono móvil para comunicarse mediante voz, datos o consultar Internet, se establece un enlace con alguna de las estaciones base de Telefónica en Alemania, que posibilita la recepción de un conjunto de datos que, inmediatamente, son anonimizados. Teralytics, analizó estos datos agregados y anónimos (<https://www.teralytics.net/contact-us/>, con acceso en diciembre de 2017) y los refinó para obtener patrones de movilidad humana; para entender cómo utilizan los ciudadanos las distintas alternativas de transporte en Núremberg. Esta metodología permitió a los científicos calcular la concentración de contaminantes atmosféricos en las zonas urbanas con un 77% de precisión. Dado que cada forma de transporte produce una cantidad única de emisiones de CO₂ y NO_x, es importante conocer los patrones de movilidad urbana para saber cuáles son las principales fuentes de emisiones contaminantes.

Estos hallazgos “mostraron que estos datos pueden usarse para dar a los urbanistas una visión de cómo la movilidad humana contribuye a la contaminación”, y se convierten en “una parte vital para diseñar e implementar eficientemente estrategias de aire limpio y bajo carbono” [Georg Polzer, CEO de Teralytics]. Es sabido, que el coste de analizar e interpretar los datos, según los investigadores, es menor que el de la producción y mantenimiento de las estaciones de medición. Este estudio ha proporcionado interesantes conclusiones y por ello, el pasado mes de mayo, el proyecto fue reconocido con el ‘German Awards for Excellence’ 2017. Con este respaldo, los investigadores quieren ampliar y mejorar su metodología, estudiando rutas cortas y añadiendo otros factores locales causantes de emisiones, como aeropuertos o eventos a gran escala, así como otros factores que también influyen, como los atascos de tráfico. De esta forma esperan poder hacer estimaciones más precisas sobre los niveles de contaminación atmosférica en una ciudad.

En este sentido, tomando como referencia que la metodología empleada permitió a los científicos calcular la concentración de contaminantes atmosféricos, en las zonas urbanas, con un 77% de precisión, creemos que la aplicación de modelos similares a los desarrollados en esta tesis, podrían contribuir a mejorar la precisión de los datos. La aplicación de modelos híbridos optimizados sobre los datos registrados, podrían determinar el orden de importancia de las variables implicadas en la contaminación, simplificando y afinando la predicción en el cálculo de la concentración de los contaminantes atmosféricos.

7.2.2.2. Futuro del uso de Big Data de redes móviles para la investigación

Los resultados del piloto de Núremberg han permitido al consorcio seguir desarrollando esta metodología, ya que han obtenido apoyo financiero del Climate KIC's Low Carbon City Lab (LoCaL), iniciativa formada por empresas, académicos, ciudades y ONGs para estudios de impacto ambiental y social. Además, el pasado mes de mayo, el proyecto fue reconocido con el 'German Awards for Excellence' 2017.

Con este respaldo, los investigadores quieren ampliar y mejorar su metodología, estudiando rutas cortas y añadiendo otros factores locales causantes de emisiones, como aeropuertos o eventos a gran escala, así como otros factores que también influyen, como los atascos de tráfico. De esta forma esperan poder hacer estimaciones más precisas sobre los niveles de contaminación atmosférica en una ciudad. Es en este sentido, donde se podrían implementar nuevos métodos de estimación y obtención de modelos locales; ensamblando algoritmos más evolucionados, como los basados en enfoques PSO-SVM, dentro del sistema de captación de datos mediante dispositivos móviles, para modelizar y predecir situaciones futuras en función de los datos actuales y ensayar nuevas técnicas que también posibiliten dar respuestas inmediatas.

7.2.3. Generar modelos para estimar la contaminación del ambiente que nos rodea, mediante aplicación móvil (app)

El proyecto EVERYWARE permite a los ciudadanos conocer su grado de exposición a contaminación atmosférica y obtener la información necesaria para decidir cuáles son las horas más adecuadas para salir de casa. Esta app, disponible para Android, permite al usuario conocer las horas pautas en las que puede ser recomendable alejarse de una zona con alta contaminación. La aplicación AirProbe, desarrollada por el proyecto de investigación EVERYWARE, financiado por la Unión Europea, mide la contaminación del aire que nos rodea, mediante un sensor, que se puede cargar fácilmente en la mochila o bolso. El sensor envía señales sobre los niveles de ozono, carbón negro u otras partículas contaminantes que hay a nuestro alrededor mediante Bluetooth al teléfono móvil. La información recogida también permite a los científicos analizar diversas tendencias relacionadas con la contaminación y poner dicha información a disposición de los ciudadanos y de las autoridades públicas a través de Internet. Esto podría resultar útil para reducir el número de atascos, por ejemplo. «Aún es demasiado pronto para sacar ninguna



conclusión, pero será interesante ver de qué forma la gente cambia sus hábitos de vida a medida que aumente su conocimiento respecto al ambiente», [profesor Vittorio Loreto (coordinador del proyecto EVERYWARE, jefe de investigaciones en la Fundación ISI (Turín) y profesor de física en la Universidad La Sapienza de Roma)].

La aplicación AirProbe requiere un dispositivo Bluetooth externo (la caja del sensor EveryAware) para trabajar. Por supuesto, también es concebible la integración del dispositivo en smartphones, aunque a más largo plazo. Todo dependerá de si las compañías están interesadas en producir la caja y en cuánto estén dispuestos a invertir los fabricantes de teléfonos. Una vez que la caja esté disponible para el gran público, se espera que esto propicie la aparición de varias aplicaciones similares a AirProbe, que se sirvan de la tecnología de código libre de EVERYWARE.

Según lo que acabamos de exponer, se podrían desarrollar modelos matemáticos que sirvan para estimar concentraciones de agentes químicos, contenidos en el aire respirable. Así, mediante un captador y su correspondiente transductor de señal se enviarían los datos oportunos al teléfono móvil, para ser operados mediante una aplicación determinada. Para ello sería valorable la aplicación de técnicas de optimización tales como el PSO, enfocadas a la mejora de la precisión del clasificador empleado y al aumento de la velocidad de procesamiento de los datos registrados por los dispositivos móviles presentes en el área de acción. Con los datos recabados se procesarían elaborarían y se aplicarían a la obtención de modelos de los resultados esperados, que estarían relacionados con:

- La información sobre la calidad actual del aire, en el área donde se encuentra.
- La obtención de modelos de la evolución de la contaminación local para las próximas horas del día.
- La posibilidad de compartir datos de la aplicación en redes sociales.

7.2.4. Minimizar los posibles errores entre usuarios que comparten datos sobre el nivel de contaminación del aire respirable

En los cruces de calles y en las rutas más transitadas los índices de contaminación del aire son más altos que en las zonas con menos tráfico, y se intensifican en las horas punta. Unos investigadores de la Universidad de California en San Diego han diseñado un procedimiento denominado CitiSense, que permite al usuario controlar la calidad del aire que respira en cada momento, en tiempo real. La información generada puede llegar a todas las personas, no sólo a los portadores de los sensores, enviando datos a smartphones y ordenadores, que son mostrados en una escala de colores basada en los índices US EPA. El proyecto Citisense ha recibido financiación del Séptimo Programa Marco de la Unión Europea para investigación, desarrollo tecnológico y demostración en virtud del acuerdo de subvención nº 308524.

En esta fase, se buscaban voluntarios para portar un pequeño sensor de contaminación portátil, de 200gr de peso y fácil manejo, desarrollado con las últimas tecnologías. Los voluntarios que quieran llevar un sensor portátil de contaminación han de:

- residir en la ciudad donde se desarrolla el estudio
- ser mayores de edad
- usar el sensor en sus desplazamientos por un periodo de entre una semana y un mes. Ofreciendo la posibilidad de que el voluntario pueda retirarse del proyecto en cualquier momento.
- llevar el smartphone en un cinturón de corredor en la cintura
- cargar el sensor y el smartphone durante la noche.

El sensor recoge datos sobre: temperatura, humedad relativa, monóxido de carbono, ozono troposférico y dióxido de nitrógeno a tiempo real y de manera geo-referenciada.

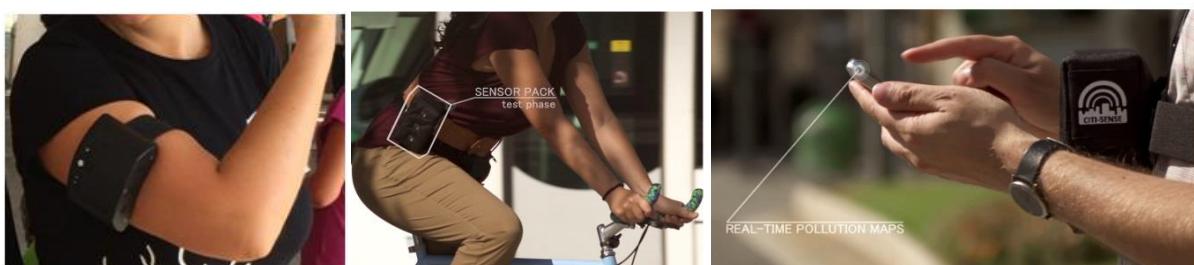


Figura 7.

Figura 7. 1. Ciudadanos, utilizando tecnologías colaborativas basadas en *smartphones* e internet, para la observación de su entorno. Fuente: herramientas, CitiSense.

Los datos se recogen mediante un smartphone (Android), que procesa la información mediante aplicaciones diseñadas a tal fin. La etapa inicial, de este proyecto involucra a ciudadanos con capacidad tecnológica que monitorizarán su entorno en Barcelona, Belgrado, Edimburgo, Haifa, Liubliana, Oslo, Ostrava y Viena. Todos los datos recogidos tanto de los sensores portátiles como de los sensores serán utilizados para crear un mapa modelo de contaminación a tiempo real de la ciudad objeto de estudio y de libre acceso.

Considerando la línea principal de esta investigación, no es descartable la idea de que se puedan producir errores o falsas alarmas a la hora de interpretar los datos captados, así como la información compartida entre usuarios debido a múltiples factores como, cambios de ubicación, interrupciones, errores de funcionamiento, etc. Por ello, se propone la elaboración de herramientas matemáticas de aplicación, mediante técnicas que posibiliten la clasificación optimizada de los datos recibidos, centrando el problema de investigación en minimizar los posibles errores en el trasvase de datos entre usuarios. Es decir, se trataría de desarrollar un sistema, aplicando técnicas de minería de datos e inteligencia artificial, para minimizar falsas alarmas que puedan causar alertas innecesarias, dado que cuando muchos dispositivos comparten información generada de forma rápida, en tiempo real, se corre el riesgo de interpretar de forma errónea o sesgada los datos. Haciendo un especial hincapié sobre aquellos datos que el usuario pueda compartir de forma personal.

Referencias bibliográficas





-
- [1] Instituto de Salud Carlos III. LA SITUACIÓN DEL CÁNCER EN ESPAÑA: INFORME 2015
 - [2] Informe Globocan, elaborado por la Organización Mundial de la Salud. <http://globocan.iarc.fr/Default.aspx> (Con acceso diciembre de 2017)
 - [3] Registro Estatal de Emisiones y Fuentes Contaminantes. <http://www.prtr-es.es/> (Con acceso diciembre de 2017)
 - [4] LNE, S.F. 02.10.2015. Arcelor pagará medio millón por la muerte de un trabajador de Baterías con leucemia.
http://www.fiteqa.ccoo.es/comunes/recursos/99927/2124447-Sentencia_benceno_ArcelorMittal_.pdf (Con acceso, diciembre de 2017)
 - [5] Estudio: 'Un oscuro panorama: los efectos en la salud de las centrales térmicas de carbón en España durante 2014', elaborado por profesionales del Instituto Internacional de Derecho y Medio Ambiente (IIDMA).
http://iidma.org/wp-content/uploads/2017/05/Un_Oscuro_Panorama_InformeCompleto_ES_opt_pro.pdf (Con acceso, diciembre de 2017)
 - [6] Estrategia institucional de Desarrollo Sostenible del Principado de Asturias.
<http://www.osasturias.es/Secciones/Publicaciones/Estrategia-de-Desarrollo-Sostenible-del-Principado/> (Con acceso, diciembre de 2017)
 - [7] Murre, JM; Sturdy, DP (1995). «The connectivity of the brain: multi-level quantitative analysis».
 - [8] Moravec, Hans (1988), *Mind Children*, Harvard University Press
 - [9] MICHAEL-TITUS, Andina. REVEST, Patricia. SHORTLAND, Peter. *The nervous system*. Churchill Livingstone. Inglaterra, 2007
 - [10] Paniagua, R.; Nistal, M.; Sesma, P.; Álvarez-Uría, M.; Fraile, B.; Anadón, R. y José Sáez, F. (2002). *Citología e histología vegetal y animal*. McGraw-Hill Interamericana de España, S.A.U.

Referencias bibliográficas

- [11] El estudio para la predicción de terremotos en Chile y en la Península Ibérica ha sido publicado en la revista Knowledge-Based Systems.
Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. F.Martínez-Álvarez, et al., 2013.
<http://www.sciencedirect.com/science/article/pii/S095070511300187>
1. Con acceso 12/2017
- [12] Neural Network Design. [Martin T. Hagan, Howard B, Demuth, Mark Beale – PWS Publishing Company]. 2nd edition, 2014.
- [13] <https://es.wiki2.org/wiki/Perceptrón>. Con acceso 12/2017.
- [14] Fayyad U., Piatesky-Shapiro G., Smyth P. (eds.) Advances in Knowledge Discovery and Data Mining, MIT Press, 1996.
- [15] https://es.wikipedia.org/wiki/Ingenier%C3%ADA_ambiental. Accesible el 12 de 2017.
- [16] Ingeniería de control de la contaminación del aire. Noel de Nevers, 1998 – McGraw Hill
- [17] Organización Panamericana de la Salud (OPS). 2005. Manual de Gestión Integrada de Residuos Sólidos Municipales en Ciudades de América Latina y Caribe. 1^a edición, 2006.
http://www.ibam.org.br/media/arquivos/estudos/girs_esp.pdf. Con acceso 12/2017
- [18] Guias de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre (2005). Air quality guidelines for Europe. Copenhagen. World Health Organization (WHO) Regional Office for Europe 2000.
- [19] Introducción a la química ambiental Autor Stanley E. Manahan. Traducido por Ivette Mora Leyva. Editor Reverte, 2007.
- [20] Contaminación atmosférica. Volumen 45 de Colección Ciencia y técnica / Ediciones de la Universidad de Castilla-La Mancha. Autores Ernesto Martínez Ataz, Yolanda Díaz de Mera Morales, Editor Univ de Castilla La Mancha
- [21] Pandis et al., 1992; Jacobson et al., 2000.
- [22] Bjorseth y Ramdahl ,1985; Jacobson et al., 2000.
- [23] Morawska and Zhang, 2002; Vouitsis et al., 2003; Fang et al., 2006.

- [24] USEPA, 1996; Seinfeld y Pandis, 1998; Qerol et al., 2001, Puigcerver y Dolors, 2008
- [25] RADIÁN INTERNACIONAL LLC. México: 1997. ACGIH. TLVs and BELs, threshold limit values for chemical substances and physical agents and biological exposure indices. Cincinatti: 2005.
- [26] Noel de Nevers. Ingeniería de Control de la Contaminación del Aire. Mexico. McGraw Hill.1997.
- [27] Health Effects of Fine Particulate Air Pollution: Lines that Connect. C. Arden Pope & Douglas Dockery 2006.
- [28] Observation of regional new particle formation in the urban atmosphere. TAREQ HUSSEIN 2004. Pope et al. ... 2002. Boreal Environment Research.
- [29] De Nevers, N. H. (1998). Ingeniería de control de la contaminación del aire. Mexico: McGraw - Hill.
- [30] Contaminación por Monóxido de Carbono: un Problema de Salud Ambiental, Jairo Téllez I, Alba Rodríguez II y Álvaro Fajardo III. Rev. salud pública. 8 (1): 108-117, 2006.
- [31] <http://www.20minutos.es/noticia/2922159/0/contaminacion-madrid-protocolo-tercer-escenario/>. Con acceso 12/2017
<http://www.madrid.es/UnidadWeb/UGNormativas/Normativa/2016/ficheros/Protocolocontaminacion21012016.pdf>. Con acceso 12/2017
- [32] http://www.crana.org/es/contaminacion/mas-informacion_3/axidos-nitrogeno-nox-no2. Con acceso 12-2017. Ver Agency for Toxic Substances and Disease Registry (ATSDR), Atlanta, Georgia. Department of Health and Human Services
- [33] <http://www.formulacionquimica.com/NO2/>. Con enlace el 12-2017
- [34] La Franchi, B. W.; Wolfe, G. M. (2009). «Closing the peroxy acetyl nitrate budget: observations of acyl peroxy nitrates (PAN, PPN, and MPAN) during BEARPEX 200». *Atmospheric Chemistry and Physics* (Copernicus Publications)
- [35] Joel Thornton, Department of Atmospheric Sciences, University of Washington (14 de noviembre de 2010). «PANS». Personal Website. N/A.
- [36] Peroxyacetyl nitrate (PAN) in the urban atmosphere. Jun-Bok Lee, (et al). 2013. Chemosphere journal homepage:
www.elsevier.com/locate/chemosphere. Con acceso 12-2017.

Referencias bibliográficas

- [37] Importance of Peroxyacetyl Nitrate (PAN) as a Phytotoxic Air Pollutant, O. C. Taylor 16 Mar 2012. Journal of the Air Pollution Control Association.
- [38] Phenomenology of the highest ozone episodes in NE Spain" by Xavier Querol et al. Atmos. Chem. Phys. Discuss., 2016.
- [39] Agencia Europea del Medioambiente (EEA - European Environment Agency), 2016. Air quality in Europe — 2016 report. Pag. 35,36.
<http://linkis.com/www.eea.europa.eu/pu/hxe56>. Con acceso 12/2017
- [40] Environment Protection Authority (EPA). Photochemical smog: what it means for us - March 2004
- [41] Galloway, J. N., Cowling, E. B., 1978, The effects of precipitation on aquatic and terrestrial ecosystems: a proposed precipitation Chemistry network, JAPCA, 28:229-35.
- [42] Robertson, N. & Wunder, S. 2005 Evaluación de Iniciativas Incipientes de Pagos por Servicios Ambientales en Bolivia.
- [43] Lee M & Thomas. 1986. "The Next Step: Acid Rain," EPA Journal, June/July 1986.
- [44] Krupa, S. U., 1977, Impact of Air Pollutants on Terrestrial Vegetation, Course notes, Department of Plant Pathology, University of Minnesota.
- [45] Lee, J. & Weber, D. (1982). Effects of Sulfuric Acid Rain on Major Cation and Sulfate Concentrations of Water
Percolating Through Two Model Hardwood Forests Journal of Environmental Quality.
- [46] Landsberg, H. E., 1984, The Resourceful Earth.
- [47] Galloway, J. N., Cowling, E. B., 1978, The effects of precipitation on aquatic and terrestrial ecosystems: a proposed precipitation Chemistry network, JAPCA, 28:229-35.
- [48] Manahan, S. 1994 "Environmental chemistry". lewis publishers. USA.
- [49] Eichler, M. & Fagundes, L. 2004. Conductas cognitivas relacionadas con problemas ambientales - Brasil.
- [50] Landsberg, H. E., 1984, The Resourceful Earth.

- [51] Por una Ciudad más Sostenible. *El Planeamiento Urbano frente al Paradigma de la Sostenibilidad. Departamento de Urbanismo y Ordenación del Territorio de la Escuela Técnica Superior de Arquitectura de Madrid* AGUSTÍN HERNÁNDEZ AJA. Ciudades para un futuro más sostenible. Ozono troposférico. Beatriz González Kirchner, Madrid (España), 2009.
- [52] U.S. Environmental Protection Agency. 2010. Clean Air Act. U.S. EPA, Office of Air and Radiation. Retrieved December 28, 2010 from <http://www.epa.gov/air/caa/>. Con acceso 12/2017
- [53] U.S. Environmental Protection Agency. 2006. Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final Report). Washington, DC: U.S. EPA, National Center for Environmental Assessment. EPA/600/R-05/004aF-cF. <http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=149923>. Con acceso 12/2017
- [54] U.S. Environmental Protection Agency. 2010. Integrated Science Assessment for Carbon Monoxide (Final Report). Washington, DC: U.S. EPA, National Center for Environmental Assessment. EPA/600/R-09/019F. <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=218686>. Con acceso 12/2017
- [55] U.S. Environmental Protection Agency. 2009. National Ambient Air Quality Standards (NAAQS). Retrieved May 21, 2009 from <http://www.epa.gov/ttn/naaqs/>. Con acceso 12/2017
- [56] Bateson, T.F., and J. Schwartz. 2008. Children's response to air pollutants. *Journal of Toxicology and Environmental Health* 71 (3):238-43.
- [57] Wigle, D.T., T.E. Arbuckle, M. Walker, M.G. Wade, S. Liu, and D. Krewski. 2007. Environmental hazards: Evidence for effects on child health. *Toxicology and Environmental Health Part B: Critical Reviews* 10 (1-2):3-39.
- [58] U.S. Environmental Protection Agency. 2010. *Integrated Science Assessment for Carbon Monoxide (Final Report)*. Washington, DC: U.S. EPA, National Center for Environmental Assessment. EPA/600/R-09/019F. <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=218686>. Con acceso 12/2017

Referencias bibliográficas

- [59] Real Decreto 785/2001, de 6 de julio, por el que se adelanta la prohibición de comercialización de las gasolinas con plomo y se establecen las especificaciones de las gasolinas que sustituirán a aquellas.
- [60] U.S. Environmental Protection Agency. 2006. *Air Quality Criteria for Lead (Final Report)*. Washington, DC: U.S. EPA, NationalCenter for Environmental Assessment. EPA/600/R-05/144aF-bF.
<http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=158823>. Con acceso 12/2017
- [61] Wigle, D.T., T.E. Arbuckle, M. Walker, M.G. Wade, S. Liu, and D. Krewski. 2007. Environmental hazards: Evidence for effects on child health. *Toxicology and Environmental Health Part B: Critical Reviews* 10 (1-2):3-39.
- [62] Kajekar, R. 2007. Environmental factors and developmental outcomes in the lung. *Pharmacology & Therapeutics* 114 (2):129-45.
- [63] U.S. Environmental Protection Agency. 2009. *Integrated Science Assessment for Particulate Matter*. Washington, DC: U.S. EPA. EPA/600/R-08/139F.
<http://cfpub.epa.gov/ncea/CFM/recordisplay.cfm?deid=216546>. Con acceso 12/2017
- [64] U.S. Environmental Protection Agency. 2008. *Integrated Science Assessment for Oxides of Nitrogen — Health Criteria (Final Report)*. Washington, DC.
<http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=194645>. Con acceso 12/2017
- [65] U.S. Environmental Protection Agency. 2008. *Integrated Science Assessment for Sulfur Oxides — Health Criteria (Final Report)*. Washington, DC: U.S. EPA, NationalCenter for Environmental Assessment. EPA/600/R-08/047F.
<http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=198843>.
- [66] Mortimer, K., R. Neugebauer, F. Lurmann, S. Alcorn, J. Balmes, and I. Tager. 2008. Early-lifetime exposure to air pollution and allergic sensitization in children with asthma. *Journal of Asthma* 45 (10):874-81.
- [67] Health Effects Institute. 2010. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. Boston, Massachusetts: Health Effects Institute.

-
- [68] Krzyzanowski, M., B. Kuna-Dibbert, and J. Schneider, eds. 2005. *Health Effects of Transport-related Air Pollution*. Copenhagen, Denmark: World Health Organization.
 - [69] Support Vector Machines for Classification and Regression. Steve R. Gunn, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science. UNIVERSITY OF SOUTHAMPTON, 1998.
 - [70] V. Vapnik, S. Golowich and A. Smola, "Support Method for Function Approximation Regression Estimation, and Signal Processing. Advances in Neural Information Processing Systems," MIT Press, Cambridge, 1997.
 - [71] Support-Vector Networks CORINNA CORTES & VLADIMIR VAPNIK. Machine Learning, 20, 273-297 (1995).
 - [72] Las condiciones Karush-Kuhn-Tucker (KKT) juegan un papel importante en la teoría y práctica de optimización de la restricción. Las condiciones de KKT se pueden encontrar en (Fletcher, 1987).
 - [73] Mitchell, T.M. Machine Learning (McGraw-Hill, 1997). Russell, S. y Norvig, P. Artificial Intelligence (A modern approach) (Prentice Hall, 2003).
 - [74] Steve R. Gunn, 1998. Support Vector Machines for Classification and Regression.
 - [75] A tutorial on support vector regresión. ALEX J. SMOLA and BERNHARD SCHOLKÖPF. Statistics and Computing 14: 199–222, 2004.
 - [76] Ivanciu, Ovidiu: Applications of Support Vector Machines in Chemistry. Reviews in Computational Chemistry, 2007.
 - [77] Evaluación de Modelos de Clasificación Carlos Hurtado L. Departamento de Ciencias de la Computación, U de Chile. 2013.
 - [78] N.T. van der Merwe, A.J. Hoffman. Developing an efficient cross validation strategy to determine classifier performance (CVCP). Neural Networks, 2001.
 - [79] Liu y col., 2008; Baumann, 2003.
 - [80] A Study of Croos-Validation and Bootstrap for Accuracy Estimation and Model Selection. Ron Kohavi, (IJCAI) 1995.Computer Science Department, Standford University
 - [81] Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.

Referencias bibliográficas

- [82] Vacca D., Gustavo. Definiendo Variables 1999
- [83] J. F. Díaz & J. C. Correa. Comparison between CART regression trees and linear regression. Comunicaciones en Estadística, diciembre 2013, Vol. 6, No. 2, pp. 175–195
- [84] Izenman, A.J.: Modern Multivariate Statistical Techniques. New York : Springer, 2008
- [85] Flexible and optimal M5 model trees with applications to flow predictions. Dimitri P. Solomatine et al., Institute for Water Education Netherland, 2004.
- [86] Predicción numérica Técnicas de regresión y predicción de datos. Jesús García Herrero Universidad Carlos III de Madrid. <http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/transparencias/regresion.pdf>. Con acceso 12-2017
- [87] J. R. Quinlan. Learning with continuous classes. pages 343-348. World Scienti_c, 1992.
- [88] Yong Wang and Ian H. Witten. Inducing model trees for continuous classes. In In Proc. of the 9th European Conf. on Machine Learning Poster Papers, pages 128-137, 1997.
- [89] Holmes, G., Hall, M., Frank, E. (1999). Generating Rule Sets from Model Trees. In: Twelfth Australian Joint Conference on Artificial Intelligence, 1-12, 1999.
- [90] E. Frank and I. H. Witten. Generating accurate rule sets without global optimization. In Proc. of the Fifteenth International Conference on Machine Learning, pages 144–151. Morgan Kaufmann, 1998.
- [91] L. Torgo. Data fitting with rule-based regression. In J. Zizka and P. Brazdil, editors, Proc. of the Workshop on Artificial Intelligence Techniques (AIT'95), Brno, Czech Republic, 1995.
- [92] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. Numerical Recipes in C (2Nd Ed.): The Art of Scienti_c Computing. Cambridge University Press, New York, NY, USA, 1992.
- [93] F. Glover. Future Paths for Integer Programming and Links to Artificial Intelligence. Computers & Operations Research, 13:533-549, 1986.
- [94] C.R. Reeves. Modern Heuristic Techniques for Combinatorial Problems. Blackwell Scientific Publishing, Oxford, UK, 1993.

-
- [95] T. Crainic and M. Toulouse. Handbook of Metaheuristics, chapter Parallel Strategies for Metaheuristics, pages 475-513. Kluwer Academic Publishers, 2003.
 - [96] Kennedy, J. & Eberhart, R. (1995), 'Particle swarm optimization', Neural Networks, 1995. Proceedings., IEEE International Conference
 - [97] J. Kennedy, R. Eberhart, and Y. Shi. Swarm Intelligence. San Francisco: Morgan Kaufmann Publishers, 2001.
 - [98] A Modified Particle Swarm Optimizer. Yuhui Shi and Russell Eberhart Department of Electrical Engineering Indiana, University Purdue, University Indianapolis, IN 46202-5160, 1998.
 - [99] CLERC M., KENNEDY J., The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space. IEEE Transactions on Evolutionary Computation, Vol. 6, No. 1, pp. 58-73, 2002.
 - [100] García Nieto, P. J. (2001). Parametric study of selective removal of atmospheric aerosol by coagulation, condensation and gravitational settling. *International Journal of Environmental Health Research*, 11, 151–162.
 - [101] García Nieto, P. J. (2006). Study of the evolution of aerosol emissions from coal-fired power plants due to coagulation, condensation, and gravitational settling and health impact. *Journal of Environmental Management*, 79(4), 372–382.
 - [102] Lutgens, F. K., & Tarbuck, E. J. (2001). *The atmosphere: an introduction to meteorology*. Prentice Hall, New York.
 - [103] Wark, K., Warner, C. F., & Davis, W. T. (1997). *Air pollution: its origin and control*. Prentice Hall, New Jersey.
 - [104] Wang, L. K., Pereira, N. C., & Hung, Y. T. (2004). *Air pollution control engineering*. Humana Press, New York.
 - [105] Karaca, F., Alagha, O., & Ertürk, F. (2005). Statistical characterization of atmospheric PM₁₀ and PM_{2.5} concentrations at a non-impacted suburban site of Istanbul, Turkey. *Chemosphere*, 59(8), 1183–1190.
 - [106] Comrie, A. C., & Diem, J. E. (1999). Climatology and forecast modeling of ambient carbon monoxide in Phoenix. *Atmospheric Environment*, 33, 5023–5036.
 - [107] Elbir, T., & Muezzinoglu, A. (2000). Evaluation of some air pollution indicators in Turkey. *Environment International*, 26(1-2):5–10.
 - [108] Godish, T., & Davis, W. T., & Fu, J. S. (2014). *Air quality*. CRC Press, Boca Raton (FL), USA.

Referencias bibliográficas

- [109] Akkoyunku, A., & Ertürk, F.A. (2003). Evaluation of air pollution trends in Istanbul. *International Journal of Environment and Pollution*, 18, 388–398.
- [110] Suárez Sánchez, A., García Nieto, P. J., Riesgo Fernández, P., del Coz Díaz, J. J., & Iglesias-Rodríguez, F. J. (2011). Application of a SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling*, 54(5–6), 1453–1466.
- [111] Cooper, C. D., & Alley, F. C. (2002). *Air pollution control*. Waveland Press, New York.
- [112] Vapnik, V. (1998). *Statistical learning theory*. Wiley-Interscience, New York.
- [113] Hastie, T., Tibshirani, R., & Friedman, J. (2003). *The elements of statistical learning*. Springer-Verlag, New York.
- [114] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, New York.
- [115] Schölkopf, B., & Smola, A. J., Williamson, R., & Bartlett, P. (2000). New support vector algorithms. *Neural Computing and Applications*, 12(5), 1207–1245.
- [116] Hansen, T., & Wang, C. J. (2005). Support vector based battery state of charge estimator. *Journal of Power Sources*, 141, 351–358.
- [117] Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis & Prevention*, 40, 1611–1618.
- [118] Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, New York.
- [119] Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer, New York.
- [120] Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In Proceedings of the Fourth IEEE International Conference on Neural Networks (vol. 4, pp. 1942–1948). Perth, Australia: IEEE Service Center.
- [121] Eberhart, R. C., Shi, Y., & Kennedy, J. (2001). *Swarm intelligence*. Morgan Kaufmann, San Francisco, USA.
- [122] Clerc, M. (2006). *Particle swarm optimization*. Wiley-ISTE, London (UK).
- [123] Olsson, A. E. (2011). *Particle swarm optimization: theory, techniques and applications*. Nova Science Publishers, New York.

-
- [124] Boznar, M., Lesjack, M., & Mlakar, P. (1993). A neural network based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex Terrain. *Atmospheric Environment*, 270, 221–230.
- [125] Haykin, S. (1999). *Neural networks: comprehensive foundation*. Prentice Hall, New Jersey.
- [126] Hooyberghs, J., Mensink, C., Dumont, D., Fierens, F., & Brasseur, O. (2005). A neural network forecast for daily average PM₁₀ concentrations in Belgium. *Atmospheric Environment*, 39(18), 3279–3289.
- [127] Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., & Cawley, G. (2003). Extensive evaluation of neural networks models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment*, 37, 4539–4550.
- [128] Gardner, M. W., & Dorling, S. R. (1999). Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmospheric Environment*, 33(5), 709–719.
- [129] Chaloulakou, A., Saisana, M., & Spyrellis, N. (2003). Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. *Science of the Total Environment*, 313, 1–13.
- [130] Karaca, F., Nikov, A., & Alagha, O. (2006). NN-AirPol: a neural-network-based method for air pollution evaluation and control. *International Journal of Environment and Pollution*, 28(3–4), 310–325.
- [131] Quinlan, J. R. (1992). Learning with continuous classes. In Proceedings of Australian Joint Conference on Artificial Intelligence (pp. 343–348). Singapore: World Scientific Press.
- [132] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth and Brooks, Monterrey (CA), USA.
- [133] Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, 528, 312–320.
- [134] Colbeck, I. (2008). *Environmental chemistry of aerosol*. Wiley-Blackwell, New York.
- [135] Hewitt, C. N., & Jackson, A. V. (2009). *Atmospheric science for environmental scientists*. Wiley-Blackwell, New York.

Referencias bibliográficas

-
- [136] Schnelle, K. B., Dunn, R. F., & Ternes, M. E. (2015). *Air pollution control technology handbook*. CRC Press, Boca Ratón (FL), USA.
 - [137] Simon, D. (2013). *Evolutionary optimization algorithms*. Wiley, New York.
 - [138] Yang, X.-S., Cui, Z., Xiao, R., Gandomi, A. H., & Karamanoglu, M. (2013). *Swarm intelligence and bio-inspired computation: theory and applications*. Elsevier, London.
 - [139] Monteiro, A., Lopes, M., Miranda, A. I., Borrego, C., & Vautard, R. (2005). Air pollution forecast in Portugal: a demand from the new air quality framework directive. *International Journal of Environment and Pollution*, 5, 1–9.
 - [141] Friedlander, S. K. (2000). *Smoke, dust and haze: fundamentals of aerosol dynamics*. Oxford University Press, New York.
 - [142] Vincent, J. H. (2007). *Aerosol sampling: science, standards, instrumentation and applications*. John Wiley & Sons, Chichester, UK.
 - [143] de Cos Juez, F. J., García Nieto, P. J., Martínez Torres, J., & Taboada Castro, J. (2010). Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model. *Mathematical and Computer Modelling*, 52, 1177–1184.
 - [144] Shawe-Taylor, J., & Cristianini, N. (2004). Kernel methods for pattern analysis. Cambridge University Press, New York.
 - [144] Clerc, M. (2012). Standard particle swarm optimisation: from 2006 to 2011. Technical report.
http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf. Accessed 23 September 2012.
 - [145] Solomatine, D. P., & Xue, Y. P. (2004). M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Hual River in China. *Journal of Hydrologic Engineering*, 9(6), 491–501.
 - [146] Rahimikhoob, A., Asadi, M., & Mashal, M. (2013). A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region. *Water Resources Management*, 27(14), 4815–4826.
 - [147] Pal, M., & Deswal, S. (2009). M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes*, 23(10), 1437–1443.
 - [148] Pal, M. (2006). M5 model tree for land cover classification. *International Journal of Remote Sensing*, 27(4), 825–831.
 - [149] Wasserman, L. (2003). *All of statistics: a concise course in statistical inference*. Springer, New York.

- [149] Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics*. W.W. Norton & Company, New York.
- [150] Picard, R., & Cook, D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387), 575–583.
- [151] Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: the .632 + bootstrap method. *Journal of the American Statistical Association*, 92(438), 548–560.
- [152] Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 1–27.
- [153] Zambrano-Bigiarini, M., & Rojas, R. (2013). A model-independent Particle Swarm Optimisation software for model calibration. *Environmental Modelling & Software*, 43, 5–25.
- [154] Zambrano-Bigiarini, M., & Rojas, R. (2014). HydroPSO: A flexible and model-independent Particle Swarm Optimisation (PSO) package for calibration/optimisation of environmental models. In R Package, version 0.3–4. Vienna, Austria: R Foundation for Statistical Computing.