





Tu universidad de postgrado

Your university for graduate studies

Tu universidad para una formación permanente

Your lifelong learning university

Tu universidad para una enseñanza innovadora

Your innovative education university

La universidad para tu futuro

The university for your future







UNIVERSIDAD

DE MÁLAGA









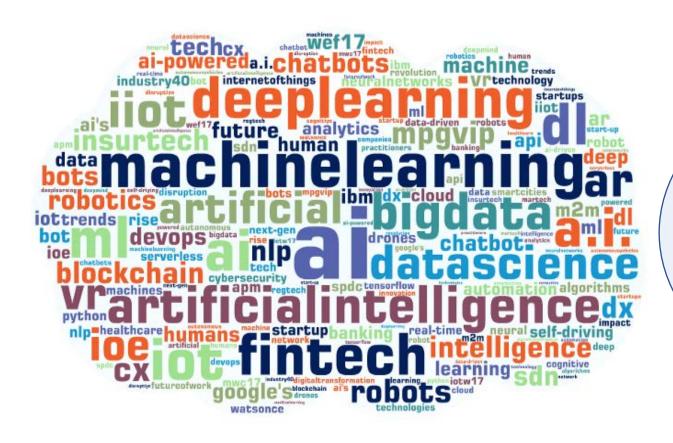
Introducción práctica a la Inteligencia Artificial y al Deep Learning

Conceptos básicos de Ciencia de Datos y Deep Learning

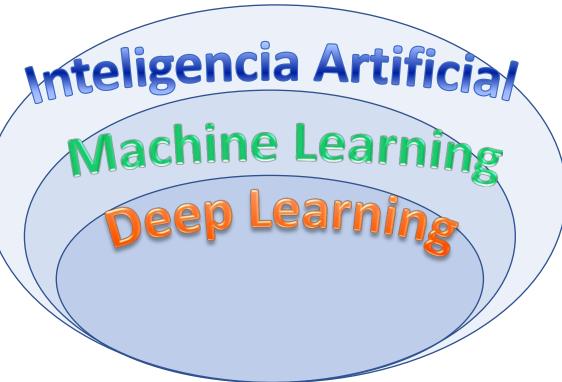


La Inteligencia Artificial y la Ciencia de Datos

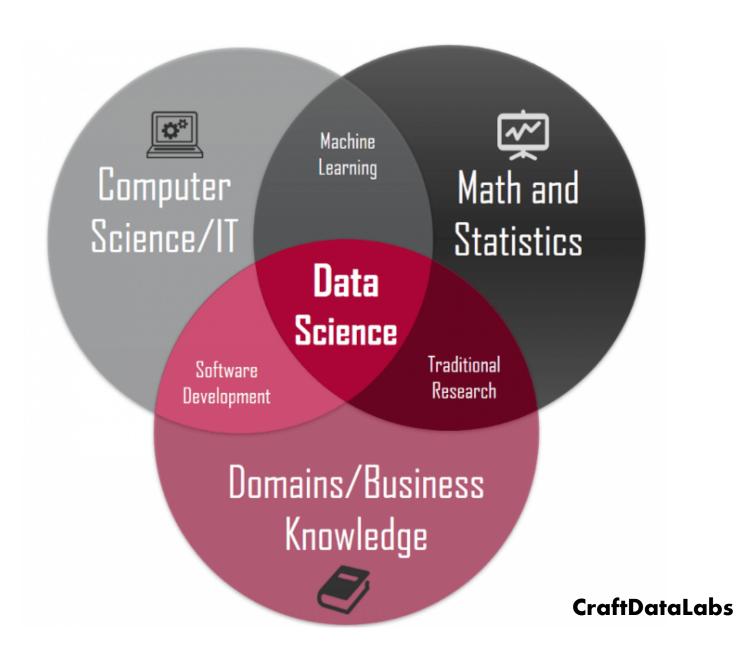




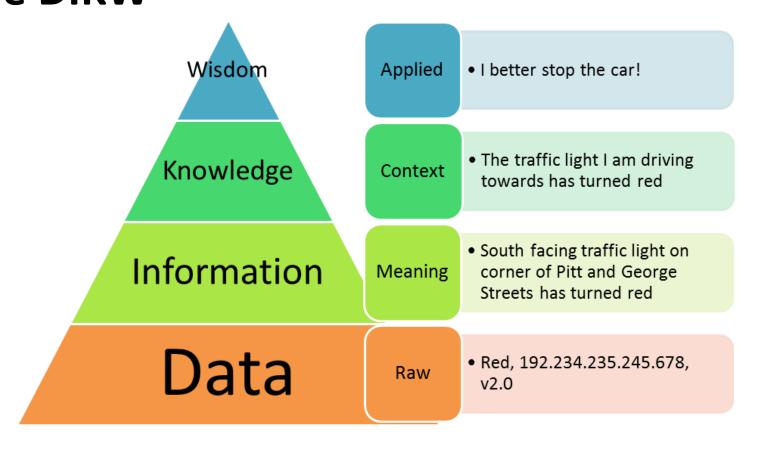
Ciencia de Datos







Objetivo: Transformar la información en conocimiento Pirámide DIKW

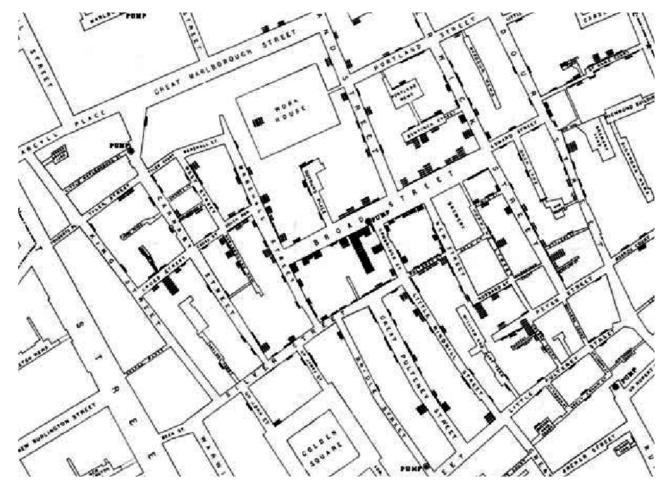




Ejemplos de aplicación de la Ciencia de Datos

El mapa de John Snow (médico británico y comandante de la guardia de la noche):

- Recopiló datos de incidencia del cólera en Londres 1854.
- Dió con la fuente de propagación
- Evidencia fundamental para aceptar que la enfermedad se propagaba por el agua y no por el aire -> Cambios en políticas de salud.





Ejemplos de aplicación de la Ciencia de Datos

Actualmente...











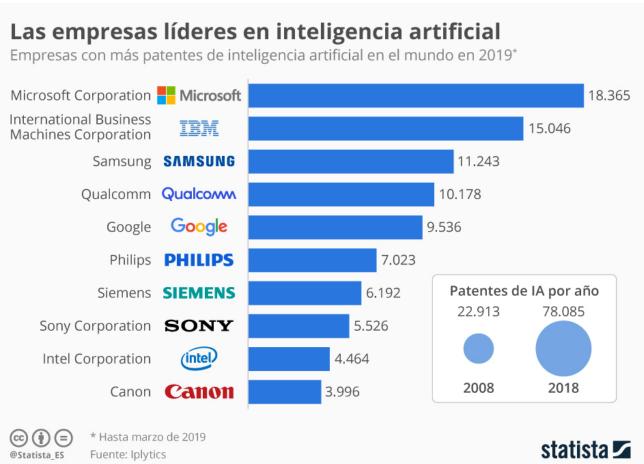












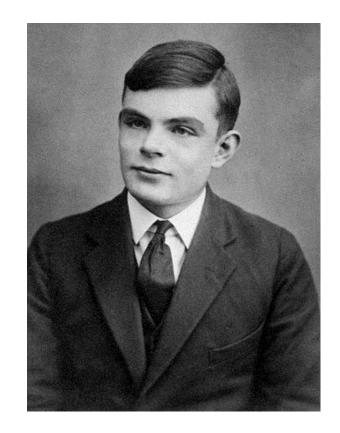




Inteligencia Artificial

Building machines that can learn from examples, experience, or even from another machines at human level are the main goal of solving AI. That goal in other words is to create a machine that pass the Turing test: when a human is interacting with it, for the human it will not possible to conclude if it he is interacting with a human or a machine.

[Turing, A.M 1950]





Evolución de la Inteligencia Artificial

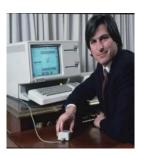
Nacimiento de la IA

- 1950. Test de Turing
- 1er programa que aprende
- 1956. Darthmouth Conference
- 1958. Perceptrón



Explosión de los 80

- Industrialización de la robótica
- Ordenadores personales
- 1981. Explanation based Learning
- 1984. Decision Trees.
- 1986. Backpropagation y MLP



La era del Machine Learning

- 2006. Hinton acuña "Deep Learning"
- 2011. IBM Watson
- 2012. GoogleBrain



1er Al Winter

- 1967. Nearest Neighbour.
- 1979. Standford Cart



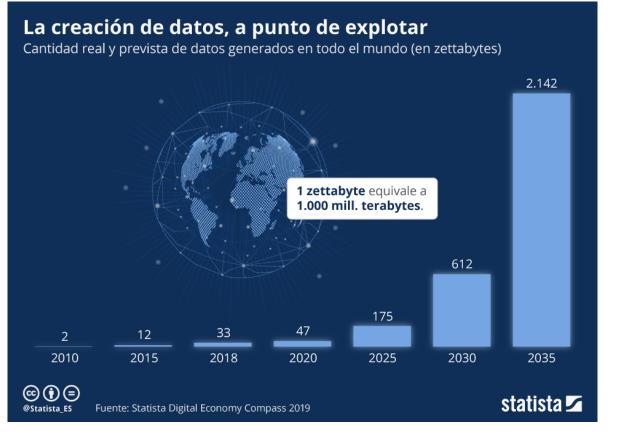
2º Al Winter

- Knowledge-driven -> Data-driven
- 1995. Q-learning (RL)
- 1997. Vapnik y SVM
- 1998. Deep Blue vs Kasparov





IA: Datos, datos, datos... Técnicas para explotarlos y oportunidades de negocio





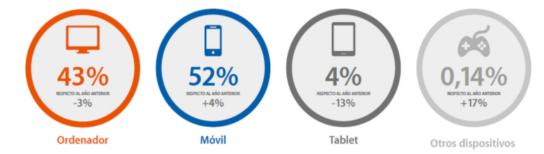


Gráfico elaborado por ditrendia a partir de datos de We are Social

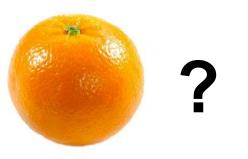




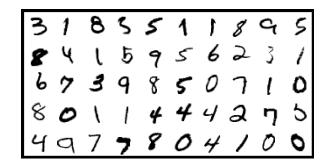


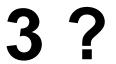
Problemas muy sencillos para los humanos pero que pueden ser complejos para un computador









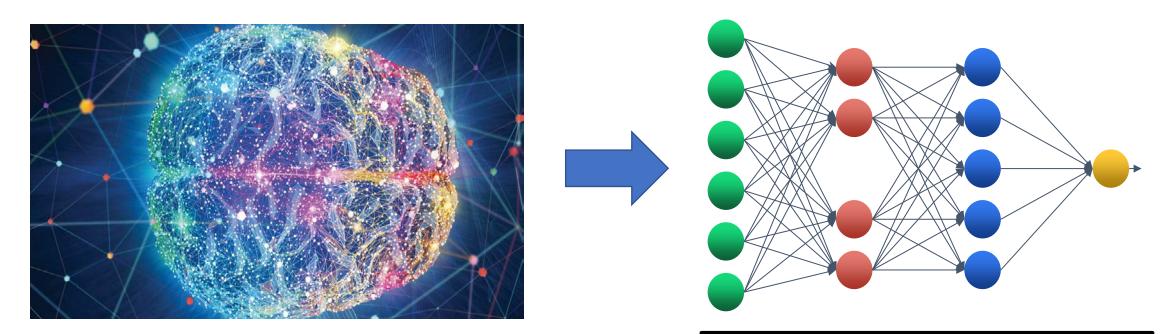


5?





Sistemas bioinspirados. Redes Neuronales Artificiales



¿Cuántas neuronas artificiales necesitaríamos para reproducir el funcionamiento del cerebro?





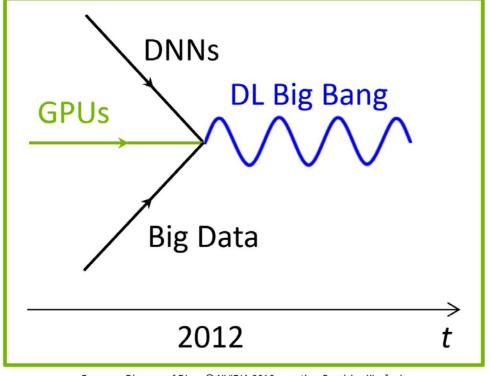
Problemas muy sencillos para los humanos pero que pueden ser complejos para un computador



Deep Learning. Un cambio de paradigma

"Deep learning is a class of machine learning that allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction".

[Lecun et al 2015]

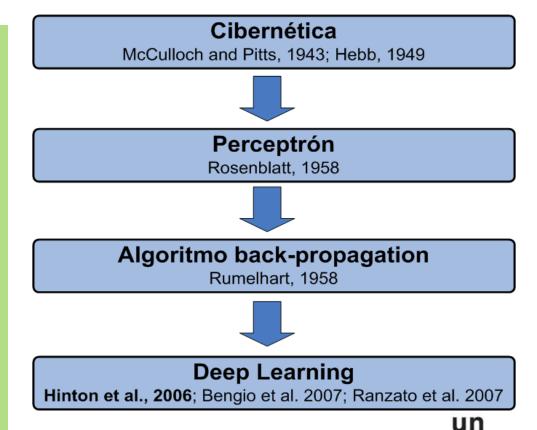






Redes neuronales y Deep Learning

- La idea de Deep Learning ha sido rebautizada a lo largo del tiempo por la influencia de diferentes investigadores.
- En 1995, se abandonaron las técnicas basadas en ANN a favor de las técnicas de aprendizaje estadístico (SVM)
- En la actualidad ha ganado popularidad por los nuevos algoritmos junto con la capacidad de cómputo de los procesadores actuales (uso de clusters y GPUs)



Aprendizaje automático

Búsqueda de relaciones estadísticas entre muestras de un conjunto de datos para reconocer patrones y realizar una acción asociada a cada uno de dichos patrones



Aprendizaje automático. Paradigmas

Aprendizaje

Supervisado

Necesaria la solución al problema para reajustar el modelo (realimentación)

No Supervisado

No es necesaria la solución al problema para reajustar el modelo. Se utilizan por ejemplo, medidas de similitud para separar clases

Reforzado

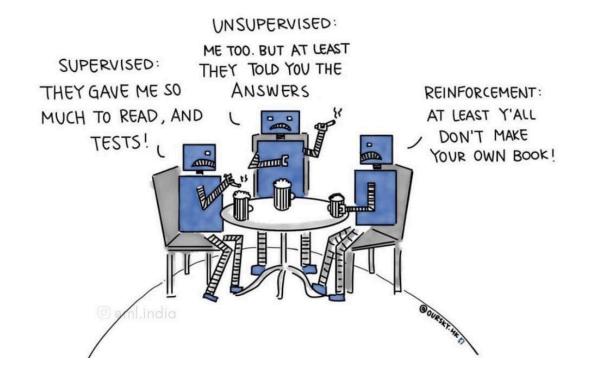
Utiliza algunas soluciones y el grado de bondad de las mismas para determinar "lo buena" que es una solución (recompensa)



Aprendizaje automático. Paradigmas



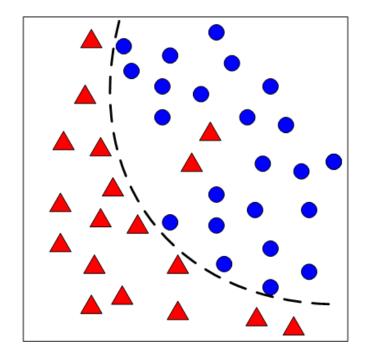
Three main types of Machine Learning Algorithms





Tipos de problemas en aprendizaje automático

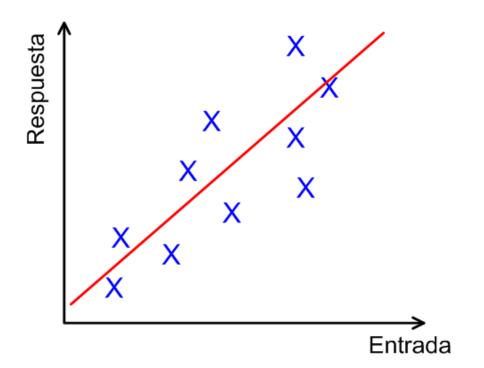
■ Clasificación: La predicción es categórica (pertenencia a una clase). Los datos de entrenamiento tienen la forma {muestra, clase}, clase $\in \{ \triangle, \circ \}$





Tipos de problemas en aprendizaje automático

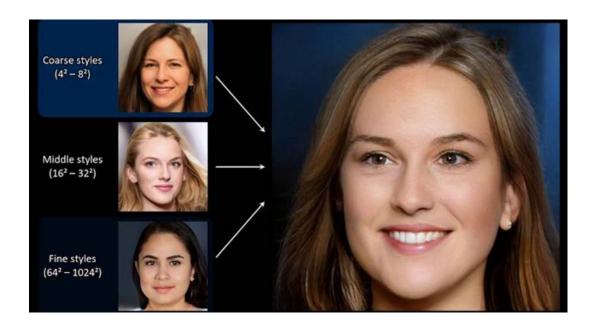
■ Regresión: La predicción es continua. Los datos de entrenamiento tienen la forma {muestra, respuesta}, respuesta $\in \square$





Tipos de problemas en aprendizaje automático

 ■ Generación: producir nuevas muestras a partir de un modelo generado mediante un proceso de aprendizaje → Modelos generativos







¿ Cómo se programan los algoritmos de Inteligencia Artificial ?

- Cualquier algoritmo, podemos programarlo cualquier lenguaje de programación (C, C++, Java, Python.... Incluso en ensamblador ②)
- No obstante, dada la complejidad de los algoritmos actuales, es MUY conveniente elegir un lenguaje que disponga de:
 - Soporte para cálculos con tensores (matrices)
 - Soporte para procesamiento paralelo
 - Soporte para uso de procesadores gráficos / tensoriales (GPU/TPU)
 - Facilidades para la optimización del código y la depuración
 - Velocidad de ejecución! (código interpretado vs. compilado...)



Herramientas más populares





Herramientas más populares

- En Inteligencia Artificial, trabajamos con Tensores
- Un tensor es una generalización de los conceptos de escalar y matriz. Un tensor puede entenderse como un array multidimensional.
- Numpy, Tensorflow y PyTorch son librerías para cálculo tensorial
- Tensorflow, además, proporciona un modelo de programación paralelo, basado en grafos, así como algoritmos de optimización
- PyTorch proporciona herramientas similares a las de TensorFlow, pero con el mismo modelo de programación que NumPy .





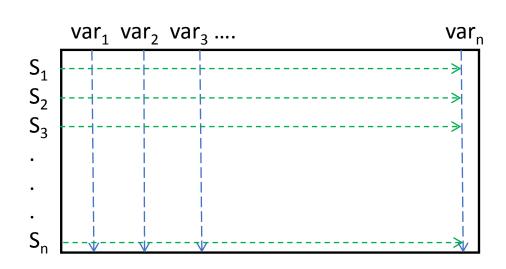
Fundamentos de Ciencia de Datos

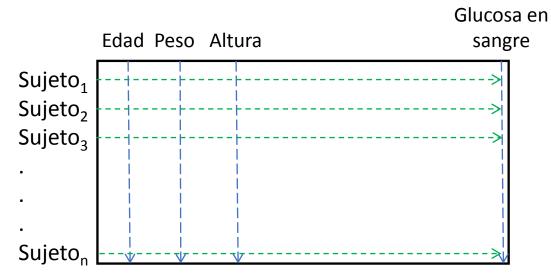


Organización de los datos

La forma habitual de organizar los datos para comenzar a trabajar con ellos, es en forma de matriz, donde cada fila es una muestra y cada columna una variable

Ejemplo









Estadística descriptiva

 Es el conjunto de técnicas numéricas y gráficas para describir un conjunto de datos sin extraer conclusiones.

Muestra vs. Población

- **Población**: universo, conjunto o totalidad de elementos sobre los que se investiga o hace un estudio.
- Muestra: subconjunto de elementos que se seleccionan previamente de una población para realizar un estudio.



Estadística descriptiva

Media

Varianza

Población

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Muestra

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}$$





Estadística descriptiva

Media

Varianza

Población

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Muestra

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}$$



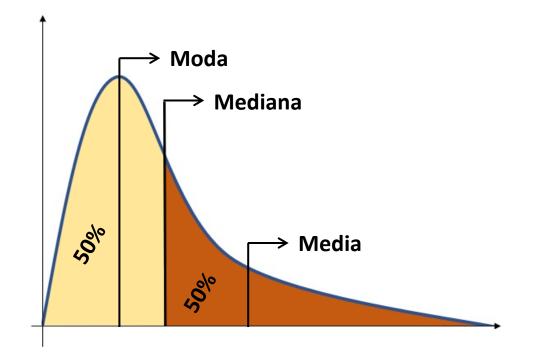






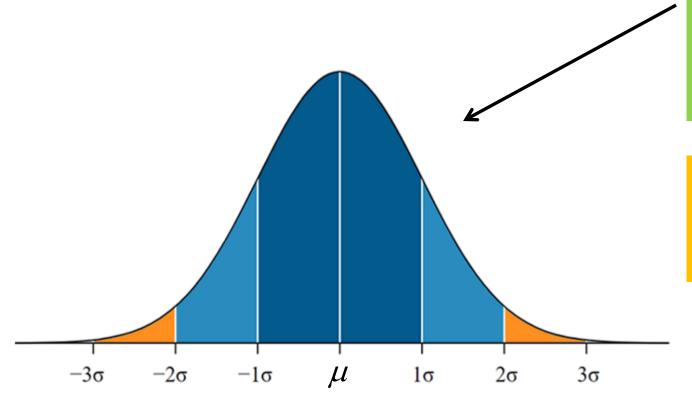
Estadística descriptiva

- Mediana: valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor.
- Moda: Valor que aparece con más frecuencia





Estadística descriptiva



 La desviación estándar y la varianza, indican la desviación con respecto a la media de una población o muestra

Distribución normal.
 Distribución con media=0 y
 varianza = 1





Estandarización

- Permite comparar puntuaciones de dos sujetos en distintas distribuciones o de un sujeto en distintas variables
- Es el número de desviaciones típicas que una medida se desvía de su media, de acuerdo a una distribución dada

Dos formas de estandarizar variables

Z-score — Suponiendo una distribución normal

t-score Suponiendo una distribución t de Student





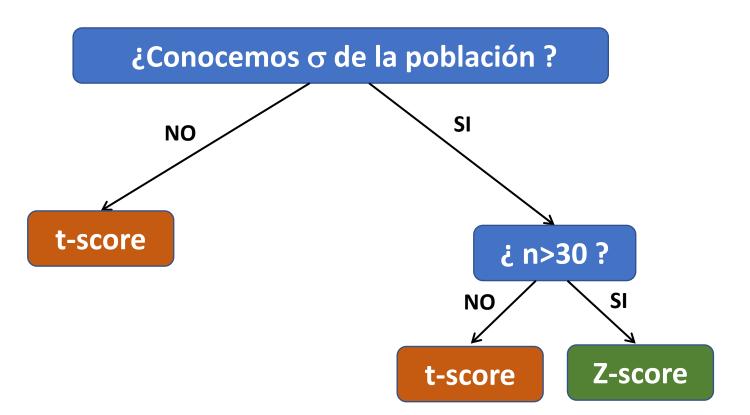
Normalización vs. Estandarización

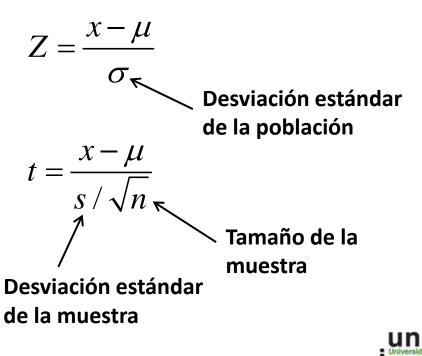
- Normalización: básicamente, consiste en modificar o adaptar la escala de los datos, con el fin de facilitar la convergencia de los algoritmos de aprendizaje
- **Estandarización:** consiste en expresar una variable como el número de desviaciones típicas que la separan de la media. De esta forma, se <u>unifica</u> la escala de todas las variables.



DE ANDALUCÍA Estandarización

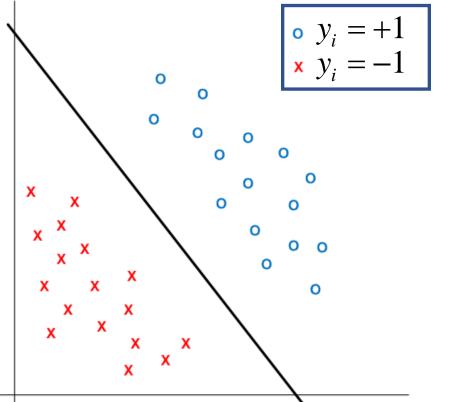
¿Cuándo usar z-score o t-score?





Clasificadores lineales

Dadas parejas
$$(x_i, y_i) \rightarrow f(x_i) = \omega^T x_i + b$$
, donde $y_i = \begin{cases} +1 \\ -1 \end{cases}$



OBJETIVO: Ajustar los pesos ω para que ante una entrada \mathcal{X}_i , la red proporcione la salida correspondiente \mathcal{Y}_i



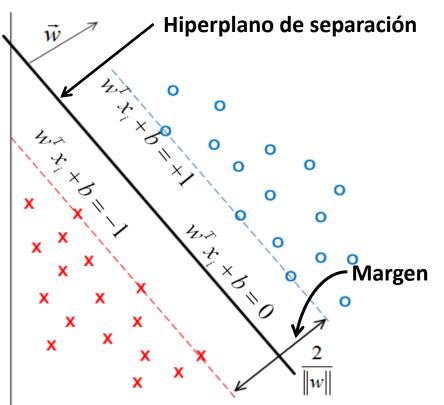




Clasificadores lineales

El cálculo del hiperplano de separación óptimo puede formularse como un problema de

optimización



$$\max \frac{2}{\|w\|} \quad \text{s.t. } w^T x_i + b = \begin{cases} \ge 1 & y_i = +1 \\ <1 & y_i = -1 \end{cases}$$

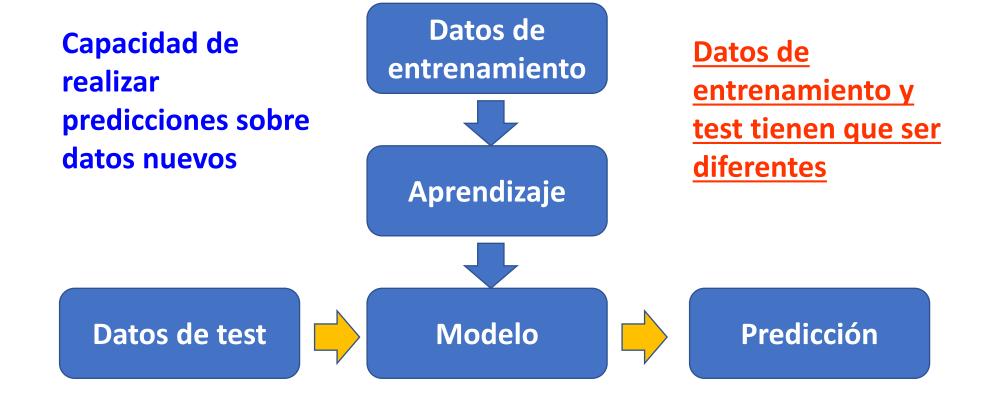
Lo que equivale a:

min
$$||w||^2$$
 s.t. $y_i(w^T x_i + b) \ge 1$

Más adelante veremos que una red neuronal con una sola capa es un clasificador lineal



Aprendizaje y capacidad de generalización ¿ Cómo entrenamos una modelo o red neuronal?

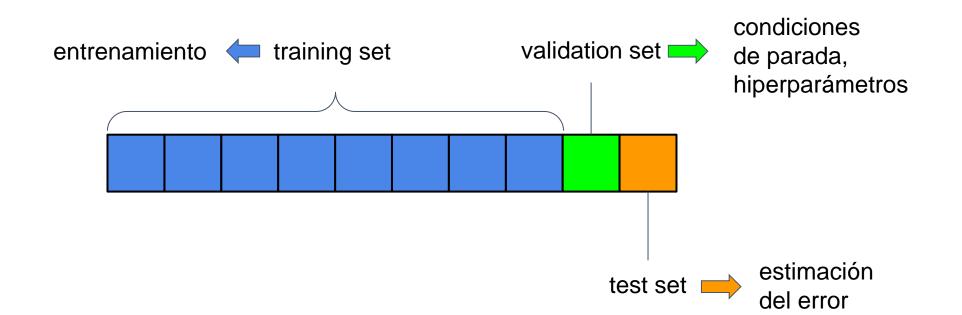






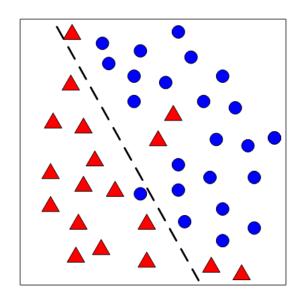
Aprendizaje y capacidad de generalización Estimación del error y ajuste de hiperparámetros

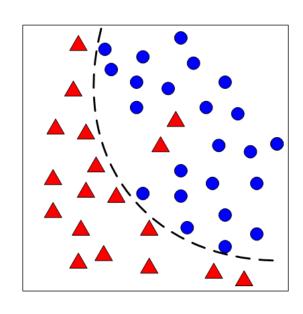
 Para evaluar la capacidad de generalización de los modelos, es necesario definir dos (tres) conjuntos de muestras: entrenamiento, validación y prueba

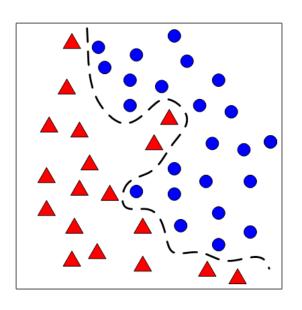




Aprendizaje y capacidad de generalización Generalización. Overfitting y Underfitting







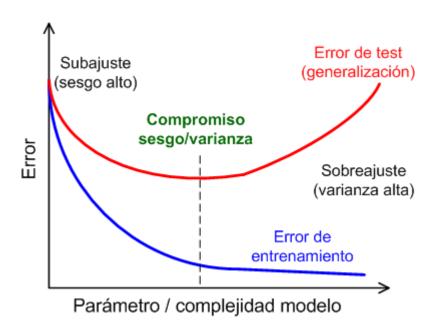
Underfitting

OK

Overfitting



Aprendizaje y capacidad de generalización Evaluación de modelos. Curva de aprendizaje

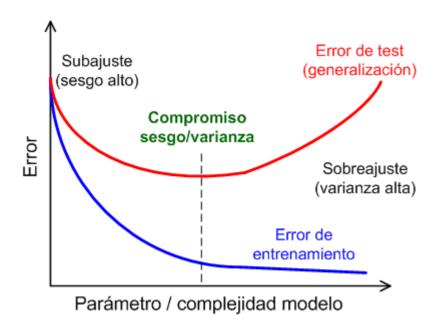


(Ej. Número de neuronas de la red)

- En aprendizaje automático, encontrar el compromiso sesgo-varianza minimiza el error de generalización.
- El sesgo (bias) es el error cometido por supuestos erróneos. Un sesgo alto puede hacer que el algoritmo pierda las relaciones entre las características y las salidas objetivo (subajuste).
- La varianza es el error producido por la sensibilidad a pequeñas fluctuaciones en el conjunto de entrenamiento. Una varianza alta puede hacer que el algoritmo modele el ruido aleatorio de los datos (overfitting).



Aprendizaje y capacidad de generalización ¿ Cómo abordar los problemas de sesgo y varianza ?



Mejora de la capacidad de generalización

- Sesgo alto.
 - Aumentar el número de muestras de entrenamiento
- Varianza alta:
 - Parada precoz del algoritmo de entrenamiento (early stopping)
 - Regularización (ej. Dropout)



Evaluación de la capacidad predictiva de un modelo Matriz de Confusión

Clase Real



$$Precisi\'on = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Especificidad = \frac{TN}{TN + FP}$$

$$Sensibilidad = \frac{TP}{TP + FN}$$



Evaluación de la capacidad predictiva de un modelo El problema del desbalanceo en la base de datos







Evaluación de la capacidad predictiva de un modelo El problema del desbalanceo en los datos

¿ Qué ocurre si nuestra base de datos no está balanceada ?



MODELO SESGADO EN FAVOR DE LA CLASE MAYORITARIA



Para evitar una mala interpretación de la métrica accuracy, en estos casos es preferible utilizar el accuracy balanceado

$$balanced _accuracy = \frac{sensitivity + specificity}{2}$$







Práctica: Conceptos básicos de Python y PyTorch



