

En caso de dedicatoria, se realiza con esta página. No es obligatoria, si bien es recomendable. hasta luego

*El autor*



# Acrónimos

<b>ETSIT</b>	Escuela Técnica Superior de Ingeniería de Telecomunicación
<b>UMA</b>	Universidad de Málaga
<b>PFC</b>	Proyecto Fin de Carrera
<b>TFG</b>	Trabajo Fin de grado
<b>TFM</b>	Trabajo Fin de Máster



# Índice

<b>Acrónimos</b>	<b>III</b>
<b>Resumen</b>	<b>1</b>
<b>1 Introduccion</b>	<b>3</b>
1.1 Motivacion . . . . .	4
1.2 Objetivos . . . . .	6
<b>2 Contexto</b>	<b>7</b>
2.1 La enfermedad del Alzheimer . . . . .	7
2.1.1 Fases de la AD . . . . .	8
2.1.2 Síntomas . . . . .	10
2.1.3 Histopatología . . . . .	11
2.2 Base de datos ADNI . . . . .	13
2.2.1 Imágenes médicas . . . . .	13
2.3 Estado del Arte . . . . .	16
2.3.1 Modelos Generadores . . . . .	16
2.3.2 Diagnóstico Asistido por Computador de AD . . . . .	23
2.3.3 Métodos de Aprendizaje Profundo . . . . .	24
2.4 Entorno de desarrollo . . . . .	27
<b>3 Fundamentos Teóricos</b>	<b>29</b>
3.1 Autoencoder Variacional . . . . .	29
3.1.1 Modelo de Variables Latentes . . . . .	29
3.1.2 Modelo Probabilístico . . . . .	30
3.1.3 Función Objetivo . . . . .	31
3.1.4 Optimización de la función objetivo . . . . .	32
3.1.5 El truco de Reparametrización . . . . .	34
3.1.6 Interpretación de la función objetivo . . . . .	35
3.1.7 Codificación y Decodificación . . . . .	35
3.2 Redes Neuronales . . . . .	38

3.2.1	Red Neuronal Densa . . . . .	38
3.2.2	Red Neuronal Convolucional . . . . .	43
3.3	Herramientas Complementarias . . . . .	48
3.3.1	Máquina de Vectores de Soporte . . . . .	48
3.3.2	Validación Cruzada (K-fold) . . . . .	48
3.3.3	Métricas de Validación . . . . .	48
<b>4</b>	<b>Trabajo Realizado</b>	<b>49</b>
<b>5</b>	<b>Resultados</b>	<b>51</b>
	<b>Conclusiones y líneas futuras</b>	<b>53</b>
<b>A</b>	<b>Apéndice</b>	<b>55</b>
A.1	Primera sección . . . . .	55
	<b>Índice alfabético</b>	<b>62</b>

# Índice de figuras

1.1	Esquema clásico de auto-encoder . . . . .	5
2.1	Evolución del número de paciente en millones desde 2010 hasta 2050 [26] . . . . .	7
2.2	Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [33] . . . . .	11
2.3	Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [33] . . . . .	11
2.4	(a) Cerebro con funciones cognitivas normales. (b) Cerebro con funciones cognitivas d . . . . .	14
2.5	(Izquierda) Modelo de una única Gaussiana, (Derecha) Modelo de mezcla de Gaussianas . . . . .	16
2.6	(Izquierda) Modelo de una única Gaussiana, (Derecha) Modelo de mezcla de Gaussianas . . . . .	17
2.7	Módulo básico de RBM . . . . .	18
2.8	<i>Stacked</i> RBMs . . . . .	18
2.9	Autoencoder Lineal . . . . .	19
2.10	Esquema de una Red Generativa Adversaria . . . . .	21
2.11	Visualización del espacio de dimensionalidad reducida conseguido en el trabajo [46]. Las imágenes superpuestas corresponden a las imágenes reales, no a las generadas . . . . .	22
2.12	Esquema clásico básico de un sistema de diagnóstico de AD asistido por computador . . . . .	23
2.13	Diagrama del Autoencoder Convolutivo 3D empleado en el trabajo [55] para la extracción de características . . . . .	25
2.14	Diagrama general del método desarrollado en el trabajo [56] . . . . .	26

3.1	Modelo gráfico de variables latentes para el modelo generativo del VAE. $Z$ es el espacio de variables lo más similar posible a una distribución normal ( $N(0, I)$ ). El elemento $\theta$ es el conjunto de parámetros que aplicados de manera funcional sobre las variables latentes son capaces de generar el conjunto muestral $X$ . . . . .	30
3.2	(Izquierda) Modelo de VAE sin Truco de Reparametrización. (Derecha) Modelo de VAE con Truco de Reparametrización . . . . .	34
3.3	Traducir . . . . .	36
3.4	Traducir . . . . .	36
3.5	Esquematización simple de las funciones del Codificador y el Decodificador en el VAE . . . . .	37
3.6	Sistema global de proceso de una red neuronal . . . . .	38
3.7	Principales funciones de activación. . . . .	39
3.8	(izquierda) Función de activación <i>Relu</i> . (Derecha) Función de activación <i>leakyRelu</i> . . . . .	40
3.9	Esquema de una red neuronal densa de una sola capa oculta . . . . .	41
3.10	Representación esquemática del proceso de entrenamiento de una red neuronal . . . . .	41
3.11	Red convolucional <i>LeNet5</i> . . . . .	43
3.12	Ejemplo de aplicación del operador convolución sobre una imagen. Seleccionada una región de la imagen cuyas dimensiones son las mismas que las del kernel seleccionado, se aplica el producto pixel a pixel entre dicha región y los pesos propios del kernel. La suma de estos productos se almacena en la imagen de salida, respetando la ubicación espacial de la región evaluada. . . . .	43
3.13	(Izquierda) Modelo clásico de redes neuronales. (Derecha) Modelo de Red Convolucional. Los datos son reagrupados en 3 dimensionados como se puede observar en una de las capas. Cada una de las capas tiene como entrada una imagen 3D y tiene como salida otra imagen 3D. La capa roja representa la capa de entrada por lo que la altura y la anchura son las dimensiones de la imagen y la profundidad son el número de canales . . . . .	44
3.14	Representación de la conectividad local en una red neuronal . . . . .	45
3.15	Desplazamiento del campo de recepción . . . . .	45
3.16	Representación de la extracción de varias características con varios filtros . . . . .	46
3.17	Representación del proceso de agrupamiento ( <i>pooling</i> ) . . . . .	47



# Índice de Tablas



# Resumen

Aquí debe escribirse el prólogo del proyecto fin de carrera.

La calidad en la presentación de los textos y las flexibilidad de L<sup>A</sup>T<sub>E</sub>X me llevaron a aprenderlo, a pesar de su difícil curva de aprendizaje.

Espero que esta plantilla ayude notablemente a suavizar este inconveniente.

Quiero agradecer a las personas que han colaborado en la realización de esta plantilla L<sup>A</sup>T<sub>E</sub>X. Es un sistema muy rápido y cómodo en la generación de este tipo de documentos técnicos y su lectura es francamente agradable.

Animo a todo el mundo a utilizarlo.

Desde la página de la escuela hay disponible también un manual de estilo para ayudar en la redacción y el acabado del proyecto. Puede consultarse en [?] <sup>1</sup>.

También sería interesante hacer dos manuales más:

- Uno de L<sup>A</sup>T<sub>E</sub>X, que explique con más detalle cómo utilizar este sistema. Aunque en Internet hay muchos disponibles, un manual rápido y directo suavizaría aún más la curva de aprendizaje.

Quizá, lo más importante es que integre todos los elementos que un usuario necesita, ya que normalmente es necesario acudir a varias fuentes y eso suele requerir demasiado tiempo.

El capítulo ?? contiene información orientada a un iniciado en este sistema.

- Y otro, que explique herramientas y métodos útiles que un proyectando puede necesitar en la elaboración del proyecto, tal como llevar un control de versiones de la documentación o el código fuente desarrollado utilizando Assembla<sup>TM</sup>. Este último manual es interesante también para muchos jóvenes profesionales, especialmente en el área de desarrollo de sistemas.

Me reservo el derecho de hacerlo, dado el escaso tiempo del que dispongo.

Muchas gracias.

---

<sup>1</sup> [http://www.uma.es/media/files/Manual\\_de\\_Estilo\\_TFG\\_ETSIT.pdf](http://www.uma.es/media/files/Manual_de_Estilo_TFG_ETSIT.pdf)



# Capítulo 1

## Introduccion

La demencia engloba un amplio grupo de enfermedades mentales que provocan el deterioro progresivo de las facultades mentales de la persona que la padece tales como la memoria, el aprendizaje, el lenguaje o la orientación.

La enfermedad del Alzheimer (AD) es la forma de demencia mas común en personas de la tercera edad. La AD es un desorden neurodegenerativo que afecta a la memoria en primer lugar, y progresivamente al resto de funciones cognitivas, provocando desajustes en el comportamiento de la persona que la padece. [1].

Actualmente esta enfermedad afecta a 30 millones de personas y se prevé que esta cifra alcance los 100 millones afectados en los próximos 50 años, por lo que además de ser un problema global de salud supone un reto socioeconómico para los países desarrollados y especialmente para aquellos países en vías de desarrollo.

La AD aún no tiene cura por ello es que la mayoría de las líneas de investigación relativas a esta enfermedad se centran en el diagnóstico temprano, con objeto de aplicar tratamientos que refuercen el mantenimiento de la reserva cognitiva cerebral para la prevención de su avance. Esta reserva cognitiva es la resistencia de nuestro cerebro frente a esta enfermedad [2].

Las imágenes por resonancia magnética (MRI) son ampliamente empleadas como herramienta de soporte para el diagnóstico de problemas cerebrales, formando parte de la rutina habitual para el diagnóstico del Alzheimer. No obstante los cambios estructurales no pueden ser detectados hasta una etapa avanzada de la AD, por ello es que se han desarrollado técnicas de representación estructural más avanzadas como las imágenes volumétricas. Por otro lado, las imágenes funcionales del cerebro tales como la tomografía de emisión de positrones (PET) permiten identificar cambios más sutiles en el metabolismo del cerebro en una etapa más temprana de la enfermedad en comparación con las imágenes MRI [3].

Se conoce como Diagnóstico Ayudado por Computer (CAD, del inglés Computer Aided Diagnosis) al conjunto de técnicas que usan imágenes médicas cerebrales con objeto de detectar la AD en una etapa temprana de la enfermedad. Existen multitud

de aproximaciones, tanto empleando las clásicas imágenes MRI [4] o imágenes PET [5], o incluso ambos tipos de imágenes de forma combinada [6] [7] conocidas como modelos Multimodal.

Las técnicas CAD emplean diferentes procedimientos estadísticos capaces de extraer características relevantes de las imágenes y en última instancia determinar si la imagen pertenece a una persona que padece AD en función de dichas características relevantes [9] [10].

Uno de los principales problemas asociados al análisis estadístico de imágenes médicas es la maldición de la dimensionalidad (CoD, del inglés *Curse of dimensionality*). Este término fué ya expuesto en 1961 por Richard Bellman debido a los problemas encontrados en procesos de optimización [11]. La maldición de la dimensionalidad hace referencia a la aparente intratabilidad de sistemáticamente obtener una función determinista sobre un espacio muestral de alta dimensionalidad, esto es, la inherente dificultad de integrar alta dimensionalidad en una única función [12]. Este problema ve acrecentada su repercusión debido a la escasez muestras, esto es lo que lo convierte en un ámbito de interés en el estudio de las imágenes cerebrales dado el número limitado de ejemplares.

Historicamente en el análisis estadístico de imágenes médicas cerebrales se han empleado técnicas de reducción de características capaces reducir la dimensionalidad del espacio muestral tales como Analisis de Componentes Principales (PCA), Analisis de Componentes Independientes (ICA) o *Sparse Filtering* [13].

En la actualidad un amplio colectivo científico involucrado en el diagnóstico temprano del AD ha desarrollado diferentes métodos de detección basados en Deep Learning [14] [15]. Se identifica por *Deep Learning* a una rama del aprendizaje automático que emplea redes neuronales de una o mas capas ocultas inspiradas con el propio cerebro humano. Esta técnica permite la modelación y abstracción de características complejas del espacio muestral sobre el que se aplica.

En la aproximación que se desarrollará en este trabajo se pretende aplicar técnicas de *Deep Learning* generativas tales como el Autoencoder Variacional [16]. Este método nos permite generar imágenes cerebrales no predefinidas, lo cual sería de utilidad a la hora de ampliar el espacio muestral tan limitado.

## Motivacion

En este trabajo se empleará un auto-encoder, específicamente el autoencoder variacional, con objeto de generar sintéticamente imágenes médicas.

Se conoce como auto-encoder a una modelo estadístico que pretende de generar muestras de salida lo más parecidas posibles a las muestras de entrada dadas, esto es, con la menos distorsión posible, por lo que es necesario extraer las relaciones inherentes en el espacio muestral.

Aunque conceptualmente simples, han tomado un rol muy importante en el aprendizaje automático, afrotando el paradigma clásico de los sistemas de aprendizajes auto-organizables [17], capaces de adaptar su estructura en función de los datos de manera no supervisada.

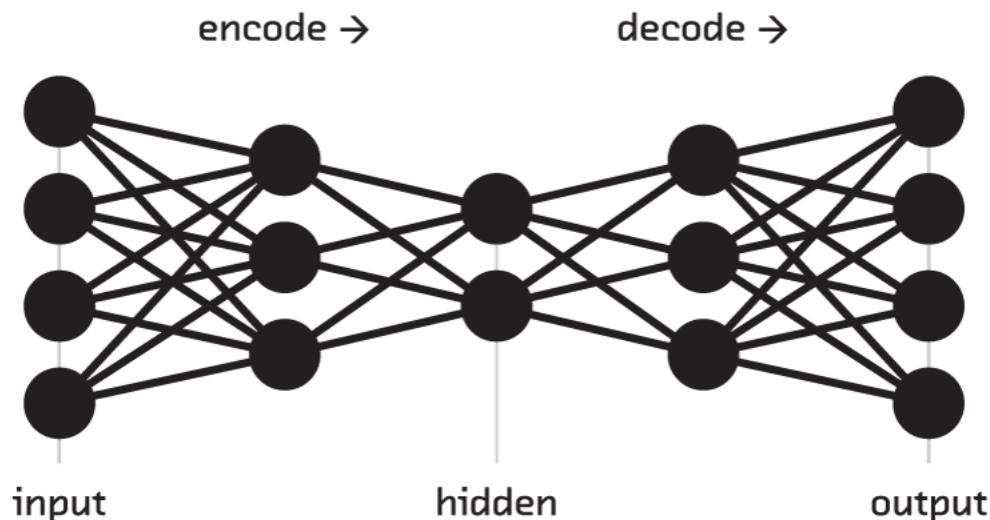


Figura 1.1: Esquema clásico de auto-encoder

Un auto-encoder está basado en el paradigma codificador-decodificador 1.1, donde el codificador se encarga de transformar la entrada en, típicamente, en una representación de baja dimensionalidad, mientras que el decodificador trata de usar esa salida de baja dimensionalidad para reconstruir la entrada original [18]. Esta capa intermedia recibe habitualmente el nombre de espacio latente.

La versión clásica de este modelo estadístico solo contenía una capa oculta, la cual era una representación de baja dimensionalidad de los datos de entrada. En la última década, el auge de las redes neuronales profundas, ámbito más conocido por *Deep Learning*, ha provocado el uso de arquitecturas más complejas en los autoencoders, con varias capas ocultas tanto en el codificador como en el decodificador. En comparación con los modelos clásicos de auto-encoders se han mejorado ampliamente los resultados, aún cuando el número de parámetros a caracterizar es el mismo en ambos sistemas.

Un auto-encoder basado en aprendizaje profundo es capaz de extraer características de manera jerárquica gracias a sus distintas capas ocultas. Existen diferentes aproximaciones entre las que cabe destacar el auto-encoder de filtrado (*Denoising auto-encoder*)[19], el auto-encoder variacional (*Variational auto-encoder*)[16] o las

redes generativas adversarias (del inglés *Generative Adversarial Networks*).

Los auto-encoders variacionales constituyen una excelente herramienta para la extracción de las características principales o de patrones de un espacio muestral, pero también pueden ser considerados un modelo generativo[21], esto es, son capaces de generar datos sintéticos. Este modelo es capaz de asociar una distribución gaussiana a cada uno de los parámetros fundamentales extraídos por el propio sistema, esto es, es capaz de caracterizar estadísticamente a lo que anteriormente denominamos como espacio latente[16]. Esta capacidad es fundamental en cualquier sistema generativo.

Los modelos generativos tienen un amplio rango de aplicaciones como son la compresión, el filtrado, el aprendizaje no supervisado de características o la síntesis de datos.

## Objetivos

- Elaborar un auto-encoder variacional basado en redes neuronales densas.
- Elaborar un auto-encoder variacional convolucional en 3d basado en redes neuronales convolucionales.
- Obtener características discriminantes por región capaces de clasificar imágenes médicas a partir de los auto-encoders elaborados.
- 
- 
- 
-



# Capítulo 2

## Contexto

### La enfermedad del Alzheimer

En 1906 Alois Alzheimer, describió las lesiones cerebrales características del trastorno que recibió su nombre: placas seniles y ovillos neurofibrilares. La AD es ahora, 100 años después, la forma más común de demencia en el mundo, estando caracterizada por un espectro de características clínicas y fallos neuropatológicos [22].

Se han desarrollado numerosos estudios de investigación relativos a la naturaleza del AD, no obstante la motivación central ha sido señalar la AD como una categoría diferente de la demencia senil [23], así como determinar si la AD es la causa principal principal de la demencia en la tercera edad. Hay teorías que defienden que la AD es una consecuencia natural del envejecimiento, mientras que hay otras que defienden lo contrario.

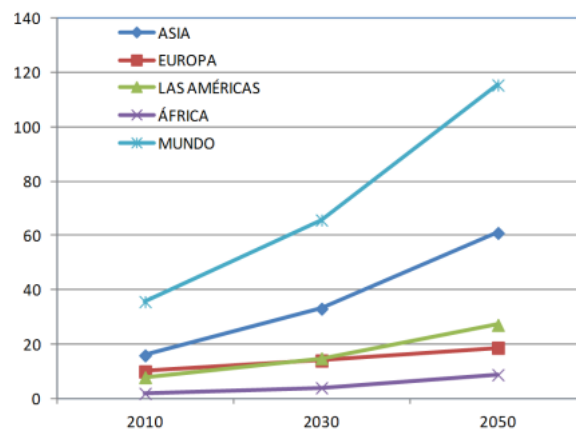


Figura 2.1: Evolución del número de paciente en millones desde 2010 hasta 2050 [26]

En la actualidad se postula desde la perspectiva de la mitocondria, orgánulo encargado de la respiración celular [24] y que toma un papel primordial en el desarrollo de la AD y del propio envejecimiento cerebral. La hipótesis en cascada de la mitocondria [25] postula que hay mecanismos comunes que conducen al envejecimiento cerebral y a la AD, así como que la producción de placas, ovillos neurofibrilares y la degeneración sináptica son consecuencias de la funcionalidad perturbada de la mitocondria.

La AD es uno de los desórdenes neurodegenerativos más serenos y frecuentes en la población de la tercera edad teniendo severas repercusiones tanto para la salud como socioeconómicas. El impacto esperado de esta enfermedad se ve incrementado debido al aumento de la esperanza de vida, se estima que durante los próximos 20 años se duplicará el número de pacientes de dicha enfermedad principalmente en los países más desarrollados. La figura 2.1 muestra la evolución del número de pacientes de AD hasta el 2050, en función de los continentes.

## **Fases de la AD**

### **Sintomas imperceptibles**

La primera de las señales tiene que ver con el descenso de los niveles de la proteína beta amiloide en el líquido cefalorraquídeo (LCR). Este proceso se puede detectar hasta 25 años antes del inicio de la pérdida de la memoria mediante una resonancia magnética. Esta proteína es la causante de la formación de las placas seniles. Durante esta fase previa a la pérdida de la memoria se hacen perceptibles las alteraciones en las estructuras tanto en las estructuras cerebrales como en el hipocampo. Es por ello que los síntomas se producen varios años de que puedan ser percibido por la propia persona o por sus familiares

### **Predemencia**

Esta fase es usualmente identificada como deterioro cognitivo o conductoal leve. Los primeros síntomas perceptibles son a menudo confundidos con la propia vejez de la persona. Una evaluación neuropsicológica detallada es capaz de determinar evidencias de AD hasta 8 años de que se cumplan los criterios de diagnóstico[27].

La deficiencia más relevante es la pérdida de memoria, ya sea como la incapacidad de adquirir nueva información o la imposibilidad de recordar hechos recientes. No obstante pueden aparecer dificultades leves en funciones ejecutivas como la atención o el razonamiento, así como trastornos en la memoria semántica [28].

### **Demencia Inicial**

El principal síntoma asociado a esta fase inicial es la pérdida de memoria puntual o incluso una pérdida de la memoria conocida a corto plazo, la cual supone dificultades para el paciente en la interacción con familiares o amigos. Una pequeña porción de los pacientes sufre de dificultades con el lenguaje, con el reconocimiento de las percepciones o con la ejecución de movimientos. [30]

La capacidad de aprender nuevos conceptos ya sean abstractos o recuerdos reales, esto es, la memoria a corto plazo, es la que se ve más afectada durante esta fase frente a otras capacidades que se ven afectadas en menor medida como es la memoria a largo plazo, la memoria semántica o la memoria implícita, la cual hace referencia al conocimiento de como realizar acciones con el propio cuerpo. [29]

### **Demencia morerada**

El síntoma diferencial de esta fase con respecto a la anterior son los cambios de conducta inesperada, incluso arranques violentos en personas que nunca han experimentado este comportamiento. Las manifestaciones neuropsiquiátricas más comunes son las distracciones, el desvarío y los episodios de confusión al final del día, así como la irritabilidad y la labilidad emocional, que incluyen llantos o risas inapropiadas. [31]

Los síntomas de fases anteriores se ven acrecentados provocando que el paciente sea incapaz de realizar tareas de cierta complejidad. Los problemas del lenguaje se hacen cada vez más evidentes, provocando parafasia. Las capacidades para leer y escribir también empeoran progresivamente. La memoria implícita y la memoria a largo que hasta entonces habían estado intactas también empiezan a verse afectadas. [30]

### **Demencia avanzada**

Esta fase última de la enfermedad trae el deterioro de la masa muscular del paciente, perdiéndose la movilidad, la capacidad de autoalimentarse y en última instancia el encamamiento del paciente.

El lenguaje se vuelve totalmente desorganizado, incluso llegando a perder completamente [30]. No obstante se conserva la capacidad de detectar y expresar señales emocionales.

La AD en sí no produce la muerte del paciente, si no que el fallo de otros sistemas que se ven afectados son los que la provocan. Los pacientes de Alzheimer pueden presentar dificultad para tragar y pueden inhalar los alimentos, lo cual puede originar neumonía por aspiración. La neumonía es la causa de la muerte en dos tercios de todas las muertes de pacientes de demencia, según la Sociedad de Alzheimer.

## Síntomas

Los síntomas asociados en la AD varía en función de las características de cada individuo, por lo que es posible que se presente en diferente grado o incluso orden en función del paciente. Estos síntomas se agrupan en tres ámbitos.

### Síntomas Cognitivos

Se ven afectadas la memoria a corto plazo en fases tempranas de la AD y seguidamente la memoria a largo plazo. La orientación espacial y temporal y la capacidad de ejecución también se ven afectadas. El síntoma principal es la incapacidad gradual de recordad a corto plazo debido a la lesiones que se producen en el hipocampo [32].

### Síntomas Psicopatológicos

Se empiezan a presentar cambios conductuales como depresión, ansiedad, agresividad o trastorno del sueño. Estos cambios en el paciente estan originados por los daños del lóbulo frontal.

### Síntomas funcionales

La interrelación entre los síntomas cognitivos, psicológicos y conductuales provocan la incapacitación del paciente para realizar las tareas cotidianas habituales así como la limitación para emprender otras nuevas.

En esta lista se indican algunos síntomas cotidianos asociados al AD:

- Cambios de memoria que dificultan la vida cotidiana.
- Dificultad para planificarse y resolver problemas.
- Dificultad para resolver tareas en la casa, en el trabajo o en el tiempo libre.
- Desorientación de tiempo o lugar.
- Dificulta para comprender imágenes visuales y cómo los objetos se relacionan uno al otro en el ambiente.
- Colocación de objetos fuera de lugar.
- Disminución o falta de buen juicio.
- Perdida de iniciativa en el trabajo o en actividades sociales.
- Cambio en el humor o la personalidad.

## Histopatología

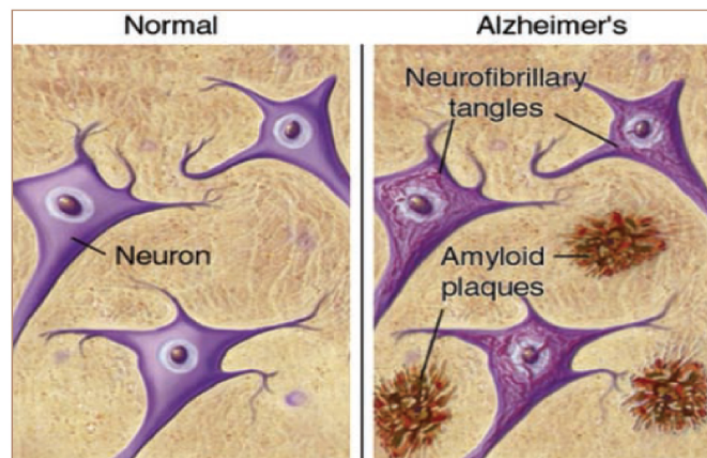


Figura 2.2: Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [33]

Las lesiones neuropatológicas comienzan a desarrollarse años antes de la completa expresión de la demencia clínica. Actualmente se desconoce cuál es el origen de este proceso de degeneración o porque los procesos normales asociados al envejecimiento se vuelven mucho más extremos en pacientes de esta enfermedad.

Desde una perspectiva patológica, los dos elementos característicos de la AD que son las placas neuríticas, la cual contiene la proteína beta amiloide ( $A\beta$ ) y los ovillos neurofibrilares sirven como línea divisoria entre la AD y otras demencias, véase la Fig. 2.4.

En la AD los ovillos neurofibrilares tienden a ser más numerosos en las estruc-

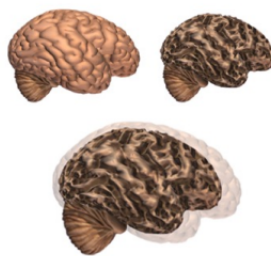


Figura 2.3: Ovillos neurofibrilares y placas seniles de los pacientes con AD respecto a los normales. Figura obtenida de [33]

turas del lóbulo temporal, incluyendo el hipocampo. Dentro del hipocampo los ovillos neurofibrilares tienden a ocupar gran parte del espacio dejado por las neuronas piramidales muertas.

Debido al depósito y acumulación de placas seniles y ovillos neurofibrilares se genera estrés resultante de la inflamación y oxidación que se añaden a la cadena patológica de las consecuencias

A medida que aumenta la enfermedad se ve reducido el número de células nerviosas y de conexiones entre ellas provocando un deterioro notable del cerebro como se aprecia en la imagen.

## Base de datos ADNI

Las imágenes empleadas en este proyecto pertenecen a la iniciativa de neuroimagen de la enfermedad de Alzheimer (ADNI, del inglés *Alzheimer Disease Neuroimaging Initiative*). Esta iniciativa fue fundada en 2004 [34] por un conjunto de instituciones de la salud norteamericanas en colaboración con diferentes compañías farmacéuticas. Una de las instituciones fundadoras más reconocidas es Instituto Nacional de la Salud (NIH, del inglés *National Institute of Health* creado en 1887, siendo actualmente referente en el ámbito de la salud en Estados Unidos.

Esta iniciativa reúne a las principales instituciones médicas tanto en Estados Unidos como en Canadá, siendo la organización que lidera las investigaciones dirigidas al entendimiento de los biomarcadores del cerebro asociados con el funcionamiento cognitivo del mismo. El investigador principal de esta iniciativa es Michael Weiner, profesor de la Universidad de California.

Hasta la fecha se han registrado hasta 1500 personas de entre 50 y 90 años en conjunto de los diferentes protocolos de la iniciativa. Se trata de una base de datos longitudinal de sujetos de la tercera edad, o cercanos a ella, que padecen AD o MCI.

Los principales objetivos de la iniciativa ADNI son los siguientes [34]:

- El desarrollo de métodos óptimos para la estandarización de la adquisición de biomarcadores, especialmente neuroimágenes, tanto MRI como PET, de una manera longitudinal sobre individuos que padecen AD o MCI.
- Uso de estos métodos optimizados de adquisición de imágenes longitudinales, tanto estructurales como metabólicas sobre un amplio conjunto de individuos sanos, de sujetos MCI o AD, acompañando dichas imágenes de una validación clínica del estado real de esos pacientes.
- Estudio de aquellos biomarcadores, medidas cognitivas o imágenes neurológicas que generan el mayor poder de diagnóstico sobre pacientes MCI y AD.
- Creación de un repositorio de datos, tanto imágenes como informes clínicos, con información longitudinal de cambios cerebrales, de metabolismo, de funcionamiento o de biomarcadores en los individuos estudiados.

## Imágenes médicas

Las imágenes neurológicas constituyen una herramienta esencial para el estudio y el diagnóstico de los trastornos psiquiátricos de desarrollo neurológico. Estas imágenes permiten el estudio longitudinal de aquellos pacientes que sufren un deterioro cognitivo o funcional, además de permitir realizar una comparación con respecto a lo que se conoce como un desarrollo neurológico normal.

En el trabajo aquí realizado nos hemos centrado en las imágenes de resonancia magnética y las imágenes tomográficas por emisión de positrones.

### Imágenes MRI

Las imágenes de resonancia magnética (MRI, del inglés *Magnetic Resonance Image*) son imágenes estructurales obtenidas en base a la aplicación de campos magnéticos sobre un cuerpo.

La técnica MRI esta basada en la resonancia magnética nuclear (NMR, del inglés *Nuclear Magnetic Resonance*). Ciertos núcleos atómicos son capaces de emitir energía a una determinada frecuencia al entrar en contacto con un campo magnético externo [37]. Generalmente, son átomos de hidrógeno los utilizados para la extracción de estas imágenes dado que este tipo de átomos existen de manera natural en las personas. Por esta razón, los escáneres evalúan la localización de la señal en el espacio, generando una imagen en función de la intensidad de la señal generada. Es posible variar el tipo de señal generada cambiando el tipo de campo magnético empleado.

Esta técnica fué inventada por Paul C. Lauterbur en Septiembre de 1971 [38], siendo aplicada por primera vez en el estudio del cerebro por Ian Robert Young y Hugh Clow en 1986 [39]. Desde entonces es considerada una técnica esencial tanto para el estudio del cuerpo humano como para la investigación biomédica, convirtiéndose en una herramienta esencial de diagnóstico.

Una de las principales limitaciones de las imágenes MRI es el ruido. El bajo nivel SNR es provocado por diversos factores, como es la alta temperatura de ruido generada por el escáner o el movimiento del cuerpo explorado. Es por ello que el procesamiento de filtrado de ruido es uno de los principales ámbitos de investigación en torno a las imágenes MRI [40]. La tecnología de los escáneres empleados ha evolucionado en aspectos como la resolución espacial o la disminución del tiempo de adquisición muestral.

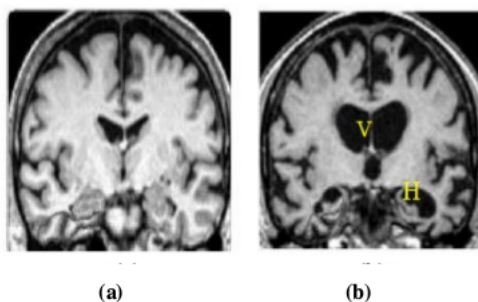


Figura 2.4: (a) Cerebro con funciones cognitivas normales. (b) Cerebro con funciones cognitivas d



## Imágenes PET

La tomografía por emisión de positrones (PET, del inglés *Positron emission tomography*) es una técnica de estracción de imágenes funcionales que permite observar los procesos metabólicos del cuerpo.

Esta modalidad de imágenes neurológicas esta basada en la detection de la radioactividad emitida por una pequeño vial inyectado en el paciente. Este vial esta compuesto por radionúclidos, isótopos radioactivos emisores de positrones. Los más utilizados en las exploraciones PET son Carbono-11, Nitrógeno-13, Oxígeno-15, Fluor-18, Cobre-62, Galio-68, Rubidio-82.

Estos isótopos son elegidos principalmente por su corto periodo de vida [13]. Los radionúclidos son incorporados en algún compuesto para expandirlo por el organismo, en el caso del Alzheimer lo más común es la glucosa, a estos compuestos se los conoce como radiofármacos. En la actualidad el radiofármaco más utilizado es el fluorodesoxiglucosa (FDG) donde el flúor de la molécula se convierte en F18. Este radiofármaco es el más utilizado debido a sus características metabólicas ya que algunos de sus compuestos están presentes en el cuerpo humano y también por su rápida expulsión del organismo sin provocar ningún efecto secundario. FDG es incorporado principalmente en las células con elevadas tasas de glucosa, como por ejemplo el cerebro, donde la fosforilación de la misma impide que sea liberada al metabolismo.

En el caso de Alzheimer, debido a su alta tasa de glucosa en la células cerebrales, la imagen PET muestra una disminución de glucosa en sus fases iniciales lo que nos permite identificar rápidamente la enfermedad. También se podrá conocer la efectividad de los tratamientos, en cuyo caso se observará un aumento del metabolismo cerebral en relación con la situación inicial.

## Estado del Arte

El presente proyecto tiene como principal objetivo la búsqueda de un modelo estadístico que nos permita generar neuroimágenes del cerebro, es por ello conveniente exponer algunos métodos clásicos de modelado generativo así como otros más novedosos en la sección que sigue a continuación.

No obstante, dado que no existen estudios de la aplicación de la técnica usada con un fin generativo sobre neuroimágenes, se expondrá algunos de los trabajos relativos a la detección y el diagnóstico temprano de AD mediante computador con objeto de mostrar el alto índice de acierto en el diagnóstico de las técnicas actuales.

## Modelos Generadores

En estadística, se define por modelo generador aquel sistema o modelo capaz de generar muestras (u observables) pertenecientes a una determinada clase o tipo respetando la función de distribución conjunta, esto es, una función de distribución multivariada asociada a dicho tipo.

### Modelos clásicos

Un modelo generador viene definido por un conjunto de distribuciones de probabilidad las cuales son capaces de aproximar de manera adecuada un conjunto de datos.

Uno de los métodos tradicionalmente más empleados es el **Modelo de Mezcla de Gaussianas**. Se trata de un modelo probabilístico que asume que todos los observables de un conjunto de datos son generados por un conjunto finito de gaussianas. La obtención del modelo se basa en estimadores de máxima verosimilitud.

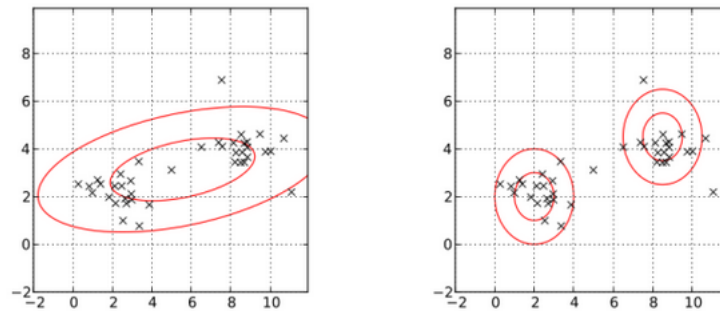


Figura 2.5: (Izquierda) Modelo de una única Gaussiana, (Derecha) Modelo de mezcla de Gaussianas

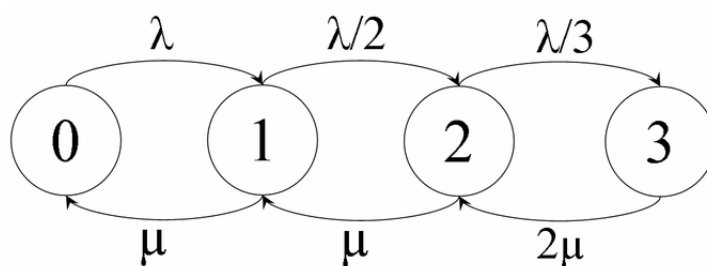


Figura 2.6: (Izquierda) Modelo de una única Gaussiana, (Derecha) Modelo de mezcla de Gaussianas

EL GMM es una de las técnicas más usadas para el modelado de datos del mundo de real. Intuitivamente podemos pensar en este método como la mezcla de varias gaussianas mono-modales.

Un **Modelo Oculto de Márkov** (HMM, del inglés *Hidden Markov Model*) es un modelo generador que asume que el proceso o sistema a modelar es un proceso de Markov. El objetivo es determinar los parámetros desconocidos de dicha cadena a partir de los parámetros.

Un HMM genera de manera explícita la distribución de probabilidad de los estados del proceso evaluado debido a la probabilidad condicional de transición entre estados, es por ello que es considerado un modelo generativo. Este tipo de modelos han sido ampliamente usado ámbitos como el reconocimiento del habla o en teoría de colas.

### La Máquina de Boltzmann

Una Máquina de Boltzman (BM del inglés *Boltzmann Machine*) es un tipo de red neuronal estocástica. Esta técnica puede emplearse para la obtención o aprendizaje de la distribución de probabilidad del conjunto de muestras en cuestión.

Dado que este proceso es costoso, se suele imponer un conjunto de restricciones en la topología de la red neuronal lo cual se conoce como máquina restrictiva de Boltzman (RBM, del inglés *Restrictive Boltzmann Machine*). [41]

Una RBM es un modelo generativo parametrizado que representa una probabilidad de distribución. Dado un conjunto de observaciones, la elaboración de una RBM implica el ajuste de los parámetros con el objetivo de que la distribución de probabilidad asociada a la BM sea lo más similar posible a la distribución real de los datos.

Tras un proceso de aprendizaje exitoso, uno RBM es capaz de extraer la distribución de probabilidad latente u oculta en un conjunto de datos. Esto puede ser empleado como referencia a la comparar con nuevas muestras, lo que se conoce como proceso de clustering clasificatorio, o puede permitirnos extraer muestras del

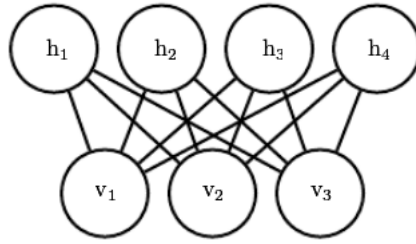
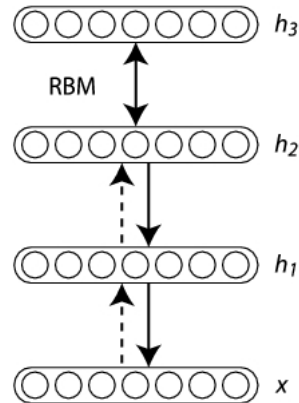


Figura 2.7: Modelo básico de RBM

Figura 2.8: *Stacked* RBMs

conjunto muestreando a partir de la distribución aprendida.

En la figura 2.7 se puede observar un modelo clásico de RBM. Las características principales de este modelo son:

- Esta constituido por dos partes bien diferenciadas; una capa de unidades ocultas y otra de unidades visibles. A menudo la unidades visibles son referenciadas como estados visibles.
- No hay conexiones entre unidades de una misma capa.
- Las unidades ocultas están probabilísticamente condicionadas a las unidades visibles, aunque son independientes entre sí.
- El uso de RBM concatenadas permite la extracción de características mas complejas, ver figura 2.8. En este modelo, el cual es comúnmente denominado *stacked RBMs*, las unidades ocultas( $h$ ) se convierten en los datos de entradas de las siguientes capas.

Las RBM son consideradas redes neuronales aplicables para el aprendizaje no supervisado, capaces de caracterizar un espacio muestral. Otro modelo similar es el Autoencoder, técnica empleada en este trabajo.

### Autoencoder

Se conoce como Autoencoder al tipo de red neuronal que tiene como objetivo la caracterización de un conjunto de datos con objeto de imitar el dato de entrada a la salida de la red. Este proceso de imitación está basado en la extracción de características, las cuales son obtenidas en lo que se denomina como capa latente. Esto permite que los autoencoders sean un método de reducción de dimensionalidad altamente empleado.

En el siguiente capítulo se profundizará en las expresiones matemáticas por lo que durante este únicamente se dará una visión general del modelo.

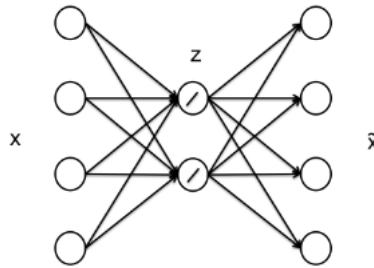


Figura 2.9: Autoencoder Lineal

En la figura 2.10 se puede observar un modelo de Autoencoder básico denominado Autoencoder Lineal. Las características principales son:

- Codificación:  $\mathbf{Z} = \mathbf{F}(\mathbf{X})$ .  $\mathbf{F}(\mathbf{X})$  es la función de codificación basada en los productos de los datos de entrada y unos pesos, obtenidos durante el entrenamiento, aplicándose una función de activación, típicamente la función sigmoide.
- Decodificación.  $\hat{\mathbf{X}} = \mathbf{G}(\mathbf{Z})$ .  $\mathbf{G}(\mathbf{X})$  realiza el proceso inverso, tomando como entrada el vector de valores del espacio latente  $\mathbf{Z}$ .

Si el espacio latente  $\mathbf{Z}$  tiene una dimensionalidad menor que el espacio  $\mathbf{X}$ , entonces la función  $\mathbf{F}$  tiene capacidad de compresión sobre las muestras de entrada  $x$ . Es por ello que un vector latente  $z$ , obtenido a partir de un vector de entrada  $x$ , puede considerarse una representación de dimensionalidad reducida de  $x$ .

Es interesante mencionar que en este caso tanto  $\mathbf{F}(\mathbf{X})$  como  $\mathbf{G}(\mathbf{Z})$  son funciones deterministas, a diferencia de las empleadas en una RBM, donde son funciones probabilísticas.

Las principales aplicaciones del Autoencoder son la extracción de características de las muestras a partir de la capa latente  $\mathbf{Z}$  y la posibilidad de generación de nuevas muestras con unas características similares a las utilizadas durante el proceso de caracterización del modelo. Esta capacidad de generación se basa en modificar los códigos latentes  $z$  de las muestras  $x$ .

No obstante, una de las principales limitaciones asociados a este modelo lineal de autoencoder es la posibilidad de que la función codificadora y decodificadora únicamente aprendan una función identidad de la muestra de entrada, lo cual imposibilita tanto el proceso de reducción de dimensionalidad sobre muestras no observadas durante el proceso de entrenamiento como el proceso de generación de nuevas muestras.

El Autoencoder de filtrado (denominado *denoising Autoencoder* en inglés) tiene como objetivo evitar dicho problema [42]. Para ello a todas las muestras de entrada  $x$  se les aplica un ruido no determinista con objeto de forzar que la capa latente tenga que aprender características robustas que extraigan pasnciones identidad de las muestras.

Un Autoencoder de filtrado realiza tres acciones principales:

- Mezclado de ruido. La muestra de entrada  $x$  es mezclada con un ruido aleatorio.
- Codificación. El código latente generado  $z$  debe preservar la información primaria de la muestra previa al mezclado del ruido.
- Decodificación. La muestra de salida  $\hat{x}$  ha de ser lo más parecida posible a la muestra de entrada  $x$ .

En el trabajo realizado por Pascal Vincent y Hugo Laroché en 2008 [43], el ruido es introducido en las muestras mediante un proceso estocástico asignando cero a algunos de los valores de las muestras. En este caso el Autoencoder de filtrado está intentando caracterizar o "predecir" los valores eliminados a partir de los valores presentes de las muestras.

Gracias al reciente de las redes profundas, se han propuesto modelos de autoencoders cada vez más complejos con varias capas intermedias, lo cual permite extraer características más complejas. El **Autoencoder Variacional** (*VAE* del inglés *Variational Autoencoder*) se ha convertido en uno de los métodos mas populares para el aprendizaje no supervisado de distribuciones complicadas.

El VAE, a diferencia de los modelos de autoencoders clásicos, busca la caracterización del espacio muestral de entrada no mediante una función determinista sino con una función de probabilidad, realizando una serie de restricciones sobre la función de distribución de los valores del espacio latente. Los conceptos matemáticos en profundidad relativos a este algoritmo serán explicado en el próximo capítulo debido a que es la principal herramienta en este trabajo.

## Red Generativa Adversaria

Otro modelo novedoso, impulsado por el desarrollo de las redes neuronales profundas, es la Red Generativa Adversaria[44]. Este modelo tiene una alta capacidad de caracterización, siendo capaz de generar imágenes realmente realistas. [45].

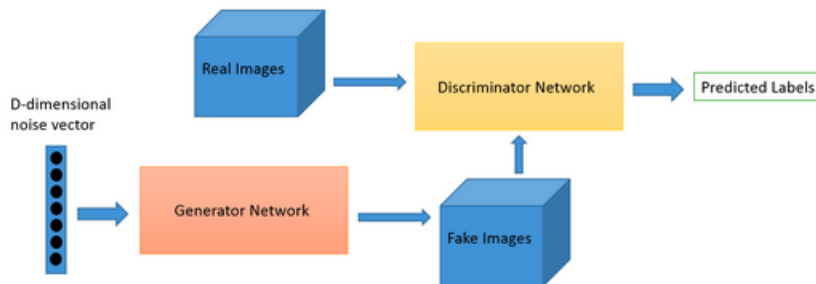


Figura 2.10: Esquema de una Red Generativa Adversaria

En este tipo de modelos tenemos dos redes bien diferenciadas la red generadora y la red discriminadora. Normalmente, la red generadora es la encargada de generar las muestras a partir del espacio latente, mientras que la red discriminadora ha de comparar las muestras generadas con las originales del espacio muestral con objeto de determinar si verdaderamente se pueden considerar muestras artificiales de dicho espacio.

El proceso de entrenamiento de la red generativa tiene como objetivo generar muestras que sean lo más parecidas a las originales, y por lo tanto, que no puedan ser detectadas por el discriminador. Mientras que el objetivo del entrenamiento de la red discriminativa es justo el contrario, esto es, ser lo más estricta posible.

Es por ello que el entrenamiento de este sistema es realmente complejo, considerándose más complicado de conseguir unos valores de entrenamiento óptimos que con respecto al VAE.

## Trabajos Previos

En la búsqueda de estudios similares al aquí desarrollado nos encontramos con el trabajo desarrollado por Eunbyung Park de la universidad de Carolina del Norte, Estados Unidos. Este trabajo parte con el objetivo de la extracción de características útiles de imágenes MRI tanto para el diagnóstico del AD como para su uso generativo [46].

En dicho trabajo se usa un autoencoder variacional convolucional 2d, lo cual difiere de los empleados en nuestro trabajo que, como se explicará más adelante, son convolucionales 3D o totalmente densos.

Se expone como dicho modelo es capaz de caracterizar componentes estructurales de las imágenes, permitiendo diferenciar la información generada en un espacio 2d, aplicando reducción de dimensionalidad mediante T-SNE. No obstante, se menciona como las imágenes regeneradas son algo borrosa, algo característico del VAE.

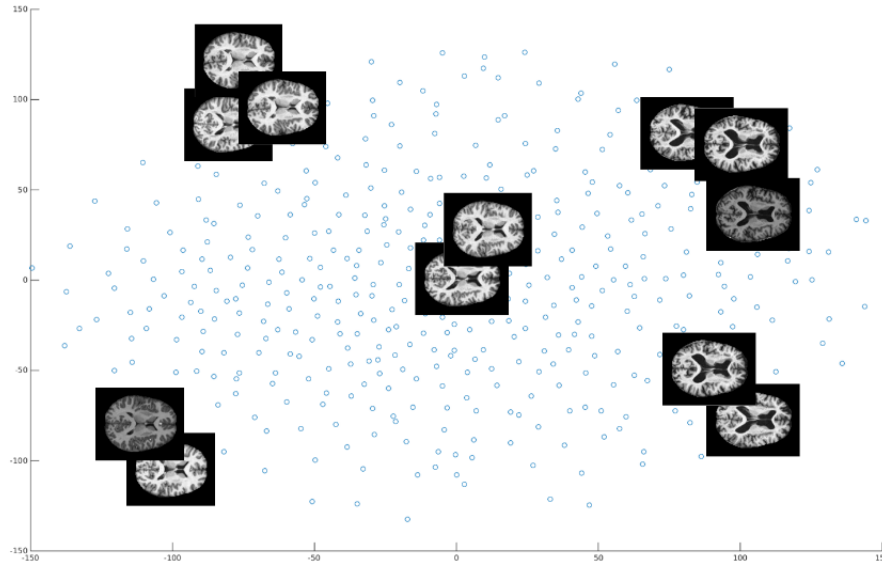


Figura 2.11: Visualización del espacio de dimensionalidad reducida conseguido en el trabajo [46]. Las imágenes superpuestas corresponden a las imágenes reales, no a las generadas

En la figura 2.11 se aprecia como es posible la diferenciación por características principales de las imágenes MRI en un espacio de dos dimensiones.



## Diagnóstico Asistido por Computador de AD

Aunque el objetivo final del trabajo es generar imágenes útiles para el diagnóstico y estudio del AD debido a los escasos trabajos que apliquen el VAE a este fin, se cree conveniente la exposición de las siguientes técnicas de diagnóstico, ya que es necesario para comprender la utilidad y el estado actual de las diferentes herramientas del ámbito tratado.

Durante la última década se han desarrollado todo tipo de aproximaciones para el diagnóstico de AD asistido por computador. Este gran desarrollo es debido, en parte, a asociaciones de gran calado como ADNI.

En este tipo de técnicas CAD, normalmente nos encontramos con dos fases bien diferenciadas, una es la extracción de características mientras que la siguiente es la clasificación realizada sobre dichas características.

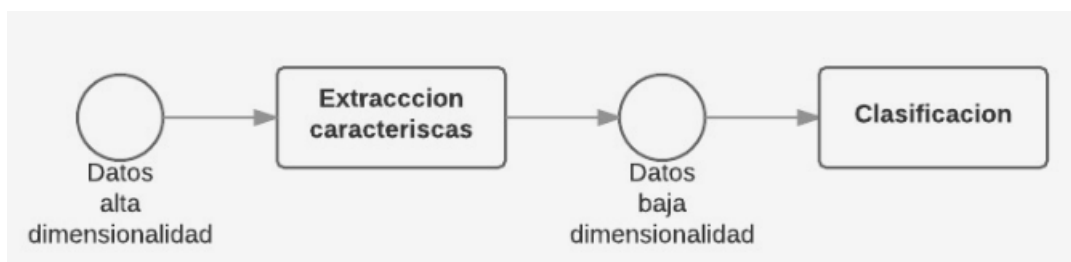


Figura 2.12: Esquema clásico básico de un sistema de diagnóstico de AD asistido por computador

### Métodos Clasicos

Algunas técnicas clásicas de extracción de características son las siguientes [47]:

- Análisis de componenets independientes (ICA, del inglés *Indepenedent Component Analysis*). Esta técnica de transformación de parámetros permita capturar información de un orden alto de dimensionalidad y transformarlo a un orden menor, utilizando componentes vectoriles estadísticiamente diferentes. [48]
- Análisis de Componentes Principales (PCA, del inglés *Principal Component Analysis*). Esta técnica es usada cuando el objetivo es reducir el número de características y convertirlas a un espacio de alta varianza y menor dimensionalidad [49].
- *Wavelets*. Se trata de un conjunto de funciones matemáticas encargadas de descomponer los datos en función de la frecuncia. La Transformada de Fourier

solo genera información frecuencial relativa al contenido, es por ello, que la Transformada *Wavelet* es una mejor herramienta para el estudio de las imágenes. [50]

- Matriz de co-ocurrencias de niveles de gris (**GLCM** del inglés *Gray Level Co-Occurrence Matrix*). Esta técnica está basada en la extracción de características estadísticas de la imagen. La matriz caracteriza la distribución de los niveles de gris de una imagen o de una region.

Algunos de los métodos que se expondrán a continuación son utilizados para la clasificación de los vectores de datos generados por los métodos de extracción de características.

- Clasificador KNN (del inglés *K-Nearest neighbor*). Este método de clasificación es uno de los más usados historicamente [51]. Este clasificador se basa en la evaluación de la distancia de una muestra en cuestión  $x_i$  con respecto al conjunto de clases o instancias posibles  $M$ , previamente predefinidas durante el entrenamiento. Se considerará que la muestra  $x_i$  pertenecerá a la clase  $m$  con respecto a la cual la distancia sea la mínima.
- Clasificador de Bayes (*Naïve bayes Classifier*). Se trata de un clasificador basado en el teorema de Bayes. En este clasificador se considera que todas las características contribuyen de manera independiente a la probabilidad de pertenecer a una clase u otra, sin tener en cuenta la presencia del resto de variables.
- Maquina de Vectores de Soporte (SVM del inglés *Support Vector Machine*). Este método de clasificación es uno de los mejores algoritmos de aprendizaje supervisado, siendo diseñado originariamente para la clasificación binaria.

Este será el método usado en este trabajo para la evaluación de las características extraídas por el sistema usado, en nuestro caso es un Autoencoder Variacional, siendo las características los valores de la capa latente. Algunos detalles de este algoritmo serán expuestos en siguientes capítulos.

## Métodos de Aprendizaje Profundo

Aunque las técnicas anteriormente expuestas constituyen los métodos clásicos relativos a las herramientas CAD, son los métodos basados en aprendizaje profundo (*deep learning*) los que han conseguido los mejores resultados de clasificación [52].

En estas técnicas de aprendizaje se han conseguido más de un 95 % de precisión en el diagnóstico del Alzheimer [53][54].

Estas técnicas hacen uso de redes neuronales capaces de extraer características complejas de las imágenes. En algunos casos, se emplean un primer tipo de red,

como por ejemplo de un Autoencoder, como método de extracción de características y seguidamente un método de clasificación ya se clásico o basado también en aprendizaje profundo.

Generalmente se usa un tipo de red neuronal denominadas redes neuronales convolucionales (CNN del inglés *Convolutional Neural Networks*), ideales para la extracción de información de las imágenes. Este tipo de redes estan basadas en el comportamiento de la vista humana, ya que son capaces de extraer información espacial de los datos, lo cual es ideal para las imágenes

En el trabajo realizado por Ehsan Hossini-Asl y Robert Keynton se consiguen resultados de exactitud de 97.6 % en la diferenciación de pacientes AD de los NC, mientras que consigue hasta un 90 % en la diferenciación de los MCI de los AD [55]. Este trabajo emplea un modelo convolucional 3D, capaz de aprender características genericas empleando para ello un Autoencoder, el cual es entrenado para capturar las variaciones estructurales en las imágenes MRI.

El planteamiento realizado en este trabajo es diferente al nuestro dado que se emplea un tipo de autoencoder en el que se capturan variables locales debido al uso de conexión mediante nodos locales en lugar de conexión global que es nuestro caso. A priori, el empleo de conexiones no globales puede dificultar la extracción de características, pero permite reducir de forma notable el tiempo necesario para el entrenamiento del sistema. Por otro lado, en nuestro trabajo se evalúan imágenes seccionadas por regiones cerebrales, en lugar de emplear imágenes completas.

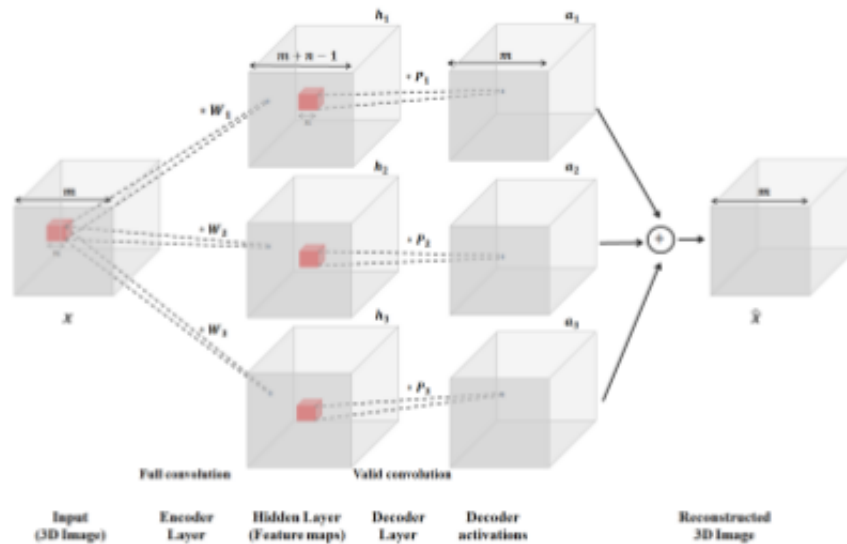


Figura 2.13: Diagrama del Autoencoder Convolucional 3D empleado en el trabajo [55] para la extracción de características

Como método de clasificación se empleó una red neuronal densa, esto es, totalmente conectada entre las unidades de capas adyacentes, la cual recibía los datos generados por el Autoencoder.

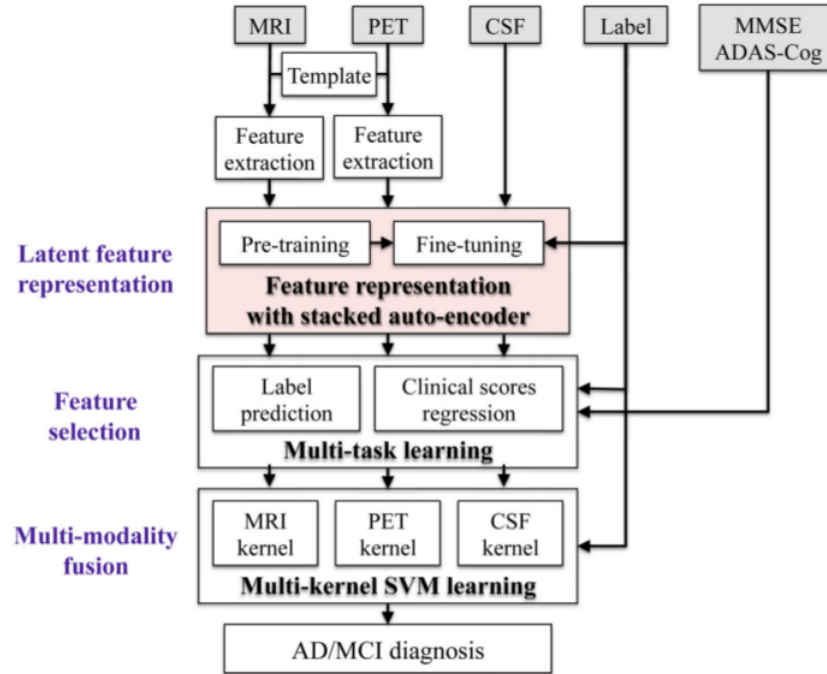


Figura 2.14: Diagrama general del método desarrollado en el trabajo [56]

Uno de los primeros estudios enfocados a la aplicación de aprendizaje profundo para el diagnóstico de AD fue el desarrollado por Heung-Il y Dinggang Shen en 2013 [56]. Es interesante notar como un Autoencoder es usado como método de extracción de características. El modelo propuesto en este estudio es el de la imagen 2.14.

La principal innovación de este método es la capacidad de extracción de correlaciones no lineales entre los datos gracias al uso de un autoencoder, lo cual teóricamente permite la extracción de mejores características globales.

Se trata de un modelo de diagnóstico multimodal ya que emplea diferentes tipos de neuroimágenes como son MRI, PET y CSF. Se emplea un Autoencoder en cascada, que permite la extracción de características, pero a diferencia del modelo Variacional, usado en este trabajo, este modelo de Autoencoder emplea una función determinista en lugar de probabilística. Como herramienta de clasificación se emplea un SVM multikernel, aplicado cada kernel a las diferentes fuentes de datos.

Los resultados indicados son de un 95 % de exactitud en la diferenciación AD sobre NC.

## Entorno de desarrollo

Los diferentes elementos software generados en este proyecto han sido desarrollados sobre Python. Este lenguaje de programación cuenta con una comunidad científica en auge. Los principales motivos por lo que se ha seleccionado son:

- Se trata de un lenguaje de código libre, lo cual evita cualquier tipo de coste asociado a la licencia de lenguaje.
- Cuenta con reconocidas librerías de métodos numéricos y estadísticos que agilizan el desarrollo de los modelos. Algunas de estas librerías son *SciPy*, *Numpy* o *Sklearn*. *Numpy* es una librería que trabaja con vectores, lo cual resulta ideal para científicos provenientes del entorno Matlab.
- Debido a su amplia comunidad hay una gran cantidad de código en repositorios públicos los cuales son útiles como referencia.

Para el desarrollo de los algoritmos basados en aprendizaje profundo hemos hecho uso de *Tensorflow* [57]. Esta biblioteca desarrollada por *Google* constituye una interfaz para el desarrollo de algoritmos de aprendizaje que facilita la implementación de dichos algoritmos.

El código de esta librería fue hecho público por *Google* el 9 de Noviembre de 2015. *Tensorflow* fue originalmente desarrollado por el equipo *Google Brain* desde 2011, denominándose *DisBelief*.

Basado en la unión de grafos de las diferentes unidades del algoritmo implementado, esta librería está principalmente orientada al desarrollo de redes neuronales, proveyendo métodos para generar redes neuronales densas y convolucionales.

Otro punto que fomenta el uso *Tensorflow* es la posibilidad de ejecutar los algoritmos sobre tarjetas gráficas (GPU, del inglés, *Graphical Processing Unit*) en lugar de sobre el procesador central (CPU, del inglés *Central Processing Unit*).



# Capítulo 3

## Fundamentos Teóricos

El trabajo realizado en este proyecto es englobado dentro de la temática denominada visión por computador, dado que los métodos empleados se basan en la detección de patrones sobre las imágenes dadas, en nuestro caso neuroimágenes.

En el ámbito de la visión por computador cada imagen en sí misma es una muestra de miles dimensiones, cada uno de los píxeles. Un modelo generativo trata de capturar la relación entre las múltiples dimensiones de los datos. En nuestro caso el modelo empleado para capturar dichas relaciones es el Autoencoder Variacional.

Es por ello que este capítulo se centrará en la exposición de este método en primer lugar. Dado el VAE esta fundamentado en el aprendizaje profundo, se dedicará la siguiente sección a las redes neuronales, haciendo especial hincapié a aquellas empleadas en este trabajo. Finalmente se expondrán brevemente los métodos estadísticos usados de manera auxiliar a lo largo de este proyecto.

### Autoencoder Variacional

Este apartado está dedicado a la exposición del Autoencoder Variacional desde una perspectiva meramente teórica con objeto de mostrar los fundamentos y, en última instancia, la capacidad de convergencia del método, basada en una función objetivo sobre la cual se puede aplicar descenso en gradiente estocástico.

### Modelo de Variables Latentes

A lo largo del entrenamiento o la caracterización de un modelo generativo, la parte más complicada es la extracción de las dependencias entre las múltiples dimensiones. Son estas relaciones multidimensionales las que permiten generar muestras artificiales pertenecientes a clases distintas. Se denomina variable latente, a las

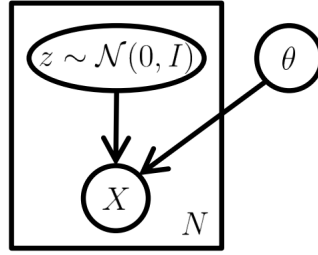


Figura 3.1: Modelo gráfico de variables latentes para el modelo generativo del VAE.  $Z$  es el espacio de variables lo más similar posible a un distribución normal ( $N(0, I)$ ). El elemento  $\theta$  es el conjunto de parámetros que aplicados de manera funcional sobre las variables latentes son capaces de generar el conjunto muestral  $X$

unidades del modelo generativo capaces de discernir entre las distintas clases, esto es, capacitan al modelo para generar elementos diferenciados.

Un modelo generativo es representativo de un espacio muestral ( $X$ ) si para cada una de las muestras de dicho espacio ( $x$ ) hay al menos alguna configuración de las variables latentes ( $z$ ) que genera un variable ( $\hat{x}$ ) muy similar a la original. Formalmente, dada una función  $f(z, \theta)$  parametrizada por un vector  $\theta$  en un espacio  $\Theta$  tal que:

$$f : Z \times \Theta \rightarrow X \quad (3.1)$$

## Modelo Probabilístico

El objetivo es maximizar la probabilidad de cada  $x$  de el espacio muestral de acuerdo con:

$$P(X) = \int P(X|z; \theta) \quad (3.2)$$

En la ecuación 3.2,  $f(z; \theta)$  es reemplazada por la distribución  $P(X|z; \theta)$ , la cual nos permite hacer explícita la dependencia de  $X$  sobre  $z$ , debido a la probabilidad condicionada. La idea de detrás de dicha expresión es principio de máxima verosimilitud (ML, del inglés *Maximum Likelihood*), el cual indica que si el modelo es capaz de generar muestras del espacio  $X$ , entonces será probable que le modelo generativo construya muestras similares.

En el VAE, la función de probabilidad  $P(X|z; \theta)$  es las siguiente:



$$P(X|z; \theta) = N(X|f(z; \theta), \sigma^2 * I) \quad (3.3)$$

El uso de una distribución gaussiana nos permite emplear descenso en gradiente durante la optimización, con objeto de caracterizar el modelo. Esta caracterización permite incrementar  $P(X)$ , entendida como la probabilidad global de generar algún tipo de muestra de dicho espacio. Esto no sería posible si esta función de probabilidad fuera una delta de Dirac. Es importante notar que es fundamental disponer de una función  $P(X|z)$  que sea computable y continua sobre  $\theta$ .

Teóricamente, para la mayoría de los valores  $z$ ,  $P(X|z)$  será aproximadamente cero, y por lo tanto su contribución para la estimación de  $P(X)$  será prácticamente nula.

## Función Objetivo

La principal idea en la que se fundamenta el VAE es en muestrear los valores de  $z$  a partir de  $X$ , esto es, necesitamos una nueva función  $Q(z|X)$  que nos permita generar el conjunto de valores del espacio  $Z$  a partir de  $X$ . Esto nos reduce el espacio de  $Z$  ya que, teóricamente, este se verá limitado en  $Q(z|X)$ . En última instancia, esto nos permitirá estimar  $E[P(X|z)]$ , siendo esta el valor esperado de la distribución de probabilidad de los valores de  $X$  generados.

La relación entre  $E(P(X|z))$  y  $P(X)$  es uno de los fundamentos de los métodos variacionales Bayesianos. Comencemos con la definición de la divergencia de Kullback-Leibler ( $KL$  o  $D$ ) entre una distribución  $P(z|X)$  y  $Q(z)$ :

$$D[Q(z)||P(z|X)] = E[\log(Q(x)) - \log(P(z|X))] \quad (3.4)$$

La expresión anterior, ecuación 3.4, es una medida no simétrica de la similitud o diferencia entre las dos funciones de probabilidad  $P(X|z)$  y  $Q(z)$ . Dicha expresión mide diferencia (o el extra de información) entre un código  $P(x)$  y uno  $Q(z)$ . Aplicando la regla de Bayes sobre la expresión anterior conseguimos dejarlo en función de  $P(X)$  y  $P(X|z)$ :

$$D[Q(z)||P(z|X)] = E_z[\log(Q(x)) - \log(P(X|z)) - \log(P(z)) + \log(p(X))] \quad (3.5)$$

Ordenando la expresión anterior, y teniendo en cuenta que  $\log(p(X))$  no depende de  $z$  por lo que puede salir del valor esperado:

$$\log(p(X)) - D[Q(z)||P(z|X)] = E_z[\log(P(X|z))] - D[Q(z)||P(z)]. \quad (3.6)$$

Llegados a este punto es importante notar que el espacio  $X$  es fijo y por lo tanto también lo es su función de probabilidad  $P(X)$ . No obstante  $Q(z)$  puede ser cualquier distribución, siempre que nos permita generar  $Z$  a partir de  $X$ .

Dado que en nuestro caso estamos interesados en inferir  $P(X)$ , es necesario generar una función  $Q$  dependiente sobre  $X$  que permita que la divergencia  $D[Q(z)||P(z|X)]$  sea pequeña, esto es, haya la menor pérdida de información entre ambas distribuciones.

$$\log(p(X)) - D[Q(z|X)||P(z|X)] = E_z[\log(P(X|z))] - D[Q(z|X)||P(z)]. \quad (3.7)$$

La expresión anterior, ecuación 3.7, es la principal del VAE, por lo que es necesario examinarla detenidamente. Analizando cada término por separado:

- La expresión de la izquierda representa la cantidad que se pretende maximizar:  $\log(P(x))$ , mas un término de error representado por  $D[Q(z)||P(z|X)]$  que es la capacidad de generar  $z$  a partir de  $X$ . Este término de error será disminuido si  $Q$  es de alta capacidad.

Se trata de maximizar  $\log(P(X))$  mientras simultaneamente  $D[Q(z|X)||P(z)]$  se minimiza . El término de probabilidad  $P(z|X)$  no es computable analíticamente, describe la distribución de valores de  $z$  que son capaces de generar  $X$ .

- La expresión de la derecha es lo que se pretende optimar mediante el descenso en gradiente, dada una correcta seleccion de  $Q(x)$ .

Este segundo término fuerza la similitud entre  $(Q(z|X))$  y  $P(X|z)$ . Asumiendo que el término  $Q(z|X)$  es de alta capacidad, tendremos que el término de divergencia KL será cercano a cero. En última instancia, conseguiremos manejar de forma analítica  $P(z|X)$  gracias a su similitud con  $Q(z|X)$

## Optimización de la función objetivo

Con objeto de poder realizar el descenso en gradiente sobre la expresión de la derecha de la ecuación 3.7, necesitamos definir de manera más exacta la forma de  $Q(z|X)$ . La elección habitual es la siguiente:

$$Q(z|X) = N(z|\mu(X; \vartheta), \Sigma(X; \vartheta)) \quad (3.8)$$

donde  $\mu$  y  $\Sigma$  son funciones deterministas con una serie de parámetros  $\vartheta$  (en las siguientes expresiones se omitirá  $\vartheta$ ). Normalmente tanto  $\mu$  como  $\Sigma$  son implementados mediante redes neuronales y  $\Sigma$  esta limitada a una función diagonal, que permite facilitar los cálculos.

El segundo término de la expresión 3.7,  $D[Q(z|X)||P(z)]$ , al ser una divergencia KL entre dos funciones de gaussianas multivariadas queda definida por:

$$D(N(\mu_0(X), \Sigma_0(X))||N(\mu_1(X), \Sigma_1(X))) = \frac{1}{2} \left( \text{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T \Sigma_1^{-1}(\mu_1 - \mu_0) - k + \log\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right) \right)$$

donde  $k$  es la dimensionalidad de la distribución, la expresión queda de la siguiente manera:

$$D[N(\mu(X), \Sigma(X))||N(0, I)] = \frac{1}{2} \left( \text{tr}(\Sigma(X)) + (\mu(X))^T (\mu(X) - k - \log(\det(\Sigma(X)))) \right) \quad (3.9)$$

El primer término de la expresión 3.7,  $E_z[\log(P(X|z))]$ , es algo más complicado de determinar, aunque a priori se podría estimar usando un número suficiente de  $z$  y aplicando la función  $f$  asociada a  $P(X|z)$ , aunque esto sería tremendamente costoso computacionalmente.

En su lugar, se aplica un procedimiento denominado Descenso en Gradiente Estocástico (SGD, del inglés *Stochastic Gradient Descent*), que se basa en tomar únicamente un valor de  $z$  aplicarlo sobre  $P(X|z)$ , por lo que se obtendría una aproximación de  $E_z[\log(P(X|z))]$ . Durante este proceso, estamos tomando como referencia cada una de las muestras  $X$  de un conjunto de datos  $D$  a la hora de estimar el error. Teniendo en cuenta esto, la ecuación completa que se pretende optimizar es:

$$E_X[\log(P(X)) - D[Q(z|X)||P(z|X)]] = E_X[E_z[\log(P(X|z))] - D[Q(z|X)||P(z)]] \quad (3.10)$$

Tomando el gradiente de la expresión anterior, reducimos la expresión a los valores internos de las esperanzas. Además, podemos tomar un único valor de  $X$  y un único valor de  $z$  de la distribución  $Q(z|X)$ , lo que no nos permite hacer computable el gradiente de la siguiente forma:

$$\log(P(X|z)) - D[Q(z|X)||P(z)]. \quad (3.11)$$

No obstante hay un problema significativo en la ecuación 3.1.4 ya que  $E_z[\log(P(X|z))]$  depende de los parámetros de  $P$  y también de los valores de  $Q$ . Esto es problemático a la hora de realizar el descenso en gradiente, quedando resuelto con lo que se conoce como "Truco de Reparametrización".

## El truco de Reparametrización

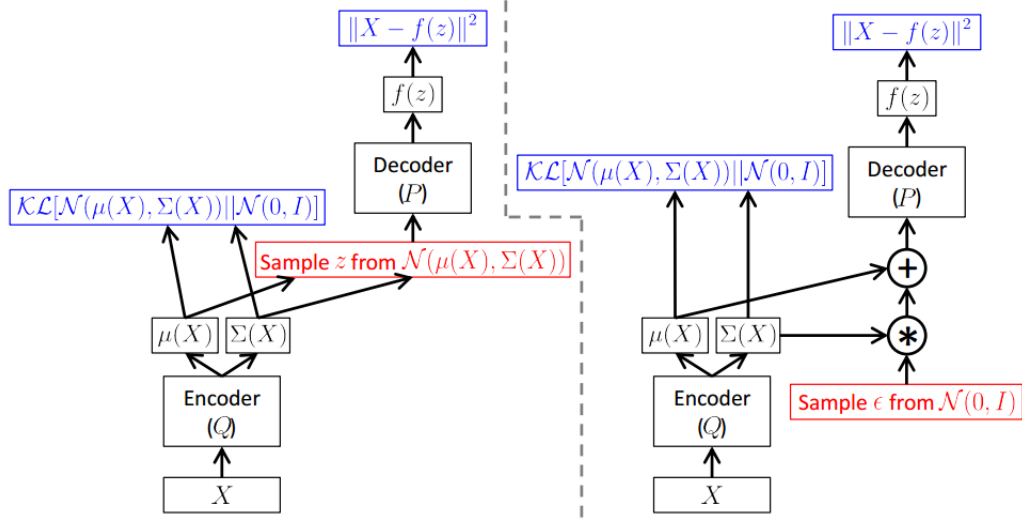


Figura 3.2: (Izquierda) Modelo de VAE sin Truco de Reparametrización. (Derecha) Modelo de VAE con Truco de Reparametrización

Para garantizar el correcto funcionamiento del VAE es necesario que la función codificadora ( $f$ ) asociada a  $Q(z|X)$  genere un conjunto  $Z$  capaz de ser decodificado por la función generadora ( $g$ ) asociada a  $P(X|z)$ .

Analizando el problema desde otra perspectiva, tomando como referencia el diagrama de izquierda de la figura 3.2. El paso hacia delante<sup>1</sup> funciona de manera correcta y es de esperar (si los parámetros están correctamente entrenados) que la salida produzca una salida acertada de manera general.

No obstante, es necesario realizar el paso hacia atrás<sup>2</sup> teniendo que determinar el gradiente sobre la función  $Q(z|X)$  encargada de generar  $z$ , pero este modelo de

<sup>1</sup>En el ámbito de las redes neuronales se denomina paso hacia delante (del inglés *forward pass*) al proceso inicial de evaluar la salida generada a partir de una determinada entrada. En nuestro caso la entrada es  $X$  y la salida  $f(z)$ , siendo la evaluación realizada  $\|X - f(z)\|$

<sup>2</sup>En el ámbito de las redes neuronales, el paso hacia atrás (del inglés *backpropagation*) hace referencia al proceso de evaluar el gradiente en cada uno de los elementos del sistema, tomando como referencia que el error se va al determinado del paso hacia delante

generación esta basado en el mapeo sobre una distribución gaussiana, lo cual es una función no continua.

La solución a este problema se denomina truco de reparametrización (del inglés *reparameterization trick*) el cual se basa en trasladar el mapeo sobre la distribución gaussiana a una capa de entrada.

Dados  $\mu_X$  y  $\Sigma_X$ , media y covarianza respectivamente de  $Q(z|X)$ , podemos mapear  $N(\mu_X, \sigma(X))$  tomando un valor de la función Normal ( $\epsilon \sim N(0, I)$ ) y aplicando la siguiente expresión:

$$z = \mu(X) + \Sigma(X)^{1/2} * \epsilon. \quad (3.12)$$

Por lo tanto la función final, la cual queda representada en el diagrama de la derecha de la figura 3.2, sobre la que se aplica el gradiente es la siguiente:

$$E_{X \sim Z} [E_{\epsilon \sim N(0, I)} [\log(P(X|z = \mu(X) + \Sigma^{1/2} * \epsilon))] - D[Q(z|X)||P(z)]] . \quad (3.13)$$

Cabe notar que ninguna de las esperanzas son con respecto a las distribuciones características del sistema (ni  $P(X|z)$  ni  $Q(z|X)$ ) lo que nos permite realizar el gradiente sin ningún problema sobre los elementos contenidos dentro de los valores esperados, ya que el gradiente es la derivada sobre los parámetros funcionales de estas distribuciones.

Por lo tanto dado un valor de  $X$  y  $\epsilon$  la función 3.1.5 será continua y determinista sobre los parámetros de  $P$  y  $Q$ , lo cual nos permite realzar el paso hacia atrás de manera eficaz.

## Interpretación de la función objetivo

### Codificación y Decodificación

La eficacia y tratabilidad del método reside en la asunción de que  $Q(z|X)$ , la función codificadora, puede ser modelada como una gaussiana con una media determinada  $\mu(X)$  y varianza  $\Sigma(X)$ , por otro lado es necesario que  $P(X)$  converja de manera eficaz a la distribución real de los datos del espacio  $D$ . Estas condiciones solo son superadas si y solo si  $D[Q(z|X)||P(z|X)]$  es cercana a cero.

Es por ello necesario una función  $Q$  de alta capacidad, lo cual puede llevarnos a modelos complejos. Los modelos basados en funciones usados en los VAE son las redes neuronales

El codificador es una red neuronal. Su entrada es el dato  $X$  y su salida es la representación latente  $z$ . Representa la distribución de probabilidad  $Q(z|X)$ , y esta

The *loss function* of the variational autoencoder is the negative log-likelihood with a regularizer. Because there are no global representations that are shared by all datapoints, we can decompose the loss function into only terms that depend on a single datapoint  $l_i$ . The total loss is then  $\sum_{i=1}^N l_i$  for  $N$  total datapoints. The loss function  $l_i$  for datapoint  $x_i$  is:

$$l_i(\theta, \phi) = -E_{z \sim q_\theta(z|x_i)}[\log p_\phi(x_i|z)] + KL(q_\theta(z|x_i)||p(z))$$

The first term is the reconstruction loss, or expected negative log-likelihood of the  $i$ -th datapoint. The expectation is taken with respect to the encoder's distribution over the representations. This term encourages the decoder to learn to reconstruct the data. If the decoder's output does not reconstruct the data well, it will incur a large cost in this loss function.

The second term is a regularizer that we throw in (we'll see how it's derived later). This is the Kullback-Leibler divergence between the encoder's distribution  $q_\theta(z|x)$  and  $p(z)$ . This divergence measures how much information is lost (in units of nats) when using  $q$  to represent  $p$ . It is one measure of how close  $q$  is to  $p$ .

Figura 3.3: Traducir

The second term is a regularizer that we throw in (we'll see how it's derived later). This is the Kullback-Leibler divergence between the encoder's distribution  $q_\theta(z|x)$  and  $p(z)$ . This divergence measures how much information is lost (in units of nats) when using  $q$  to represent  $p$ . It is one measure of how close  $q$  is to  $p$ .

In the variational autoencoder,  $p$  is specified as a standard Normal distribution with mean zero and variance one, or  $p(z) = \text{Normal}(0, 1)$ . If the encoder outputs representations  $z$  that are different than those from a standard normal distribution, it will receive a penalty in the loss. This regularizer term means 'keep the representations  $z$  of each digit sufficiently diverse'. If we didn't include the regularizer, the encoder could learn to cheat and give each datapoint a representation in a different region of Euclidean space. This is bad, because then two images of the same number (say a 2 written by different people,  $z_{alice}$  and  $z_{bob}$ ) could end up with very different representations  $z_{alice}$ ,  $z_{bob}$ . We want the representation space of  $z$  to be meaningful, so we penalize this behavior. This has the effect of keeping similar numbers' representations close together (e.g. so the representations of the digit two  $z_{alice}$ ,  $z_{bob}$ ,  $z_{ali}$  remain sufficiently close).

We train the variational autoencoder using gradient descent to optimize the loss with respect to the parameters of the encoder and decoder  $\theta$  and  $\phi$ . For stochastic gradient descent with step size  $\rho$ , the encoder parameters are updated using  $\theta \leftarrow \theta - \rho \frac{\partial l}{\partial \theta}$  and the decoder is updated similarly.

Figura 3.4: Traducir

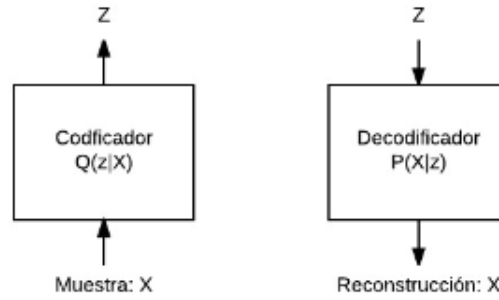


Figura 3.5: Esquematzación simple de las funciones del Codificador y el Decodificador en el VAE

determinada por el conjunto de parámetros y pesos de la red neuronal asociada. Denominaremos a la función encargada de la codificación  $q_{\theta}(z|x)$

El codificador se identifica a menudo con el proceso de reducción de la dimensionalidad de  $x$  a  $z$ . Cabe notar que el codificador tiene asociadas dos funciones, una encargada de obtener la media  $q_{\mu}(X)$  y otro la varianza  $q_{\Sigma}(X)$  del espacio latente. Para la obtención final de  $z$  se ha de aplicar el truco de reparametrización, ver sección 3.1.5, con respecto a los valores  $\Sigma$  y  $\mu$  obtenidos anteriormente.

El decodificador es otra red neuronal. Su entrada es la variable del espacio latente  $z$  y su salida es la reconstrucción del dato inicial  $X$ . Denominaremos a la función encargada de la decodificación  $p_{\phi}(x|z)$ , donde  $\phi$  son el conjunto de parámetros y pesos que definen la red neuronal.

El hecho de que ambas funciones estén basadas en redes neuronales hace el aprendizaje profundo sea una parte primordial del VAE. Típicamente los formatos de redes neuronales aplicados en este sistema son dos; redes neuronales densas (DNN) o redes neuronales convolucionales (CNN).

## Redes Neuronales

Las Redes Neuronales permiten generar funciones complejas no lineales gracias a su capacidad inherente de aprendizaje con el proceso denominado propagación hacia atrás, que permiten ajustar los pesos de las distintas unidades o neuronas del sistema.

Dada la complejidad del ámbito del aprendizaje profundo, en las siguientes secciones se pretenden exponer las ideas fundamentales para comprender el comportamiento de las funciones de codificación y decodificación del VAE, sin entrar en explicaciones excesivamente teóricas sobre los fundamentos de las redes neuronales.

Es por ello que en primer lugar se expondrá el modelo de redes neuronales densas, aprovechando para exponer de manera somera algunos conceptos de redes neuronales, como son el concepto de funciones de activación o el proceso de propagación hacia atrás.

Seguidamente se expondrá el otro modelo de aprendizaje profundo utilizado en este trabajo que son las redes neuronales convolucionales, explicando por qué son ideales para la captura de patrones sobre imágenes.

### Red Neuronal Densa

Este modelo constituye el paradigma básico de redes neuronales. Fundamentado en el estándar de neurona artificial según los principios descritos por Rumelhart y McClelland en 1986 [58]. Siguiendo dichos principios, la  $i$ -ésima neurona artificial consiste en:

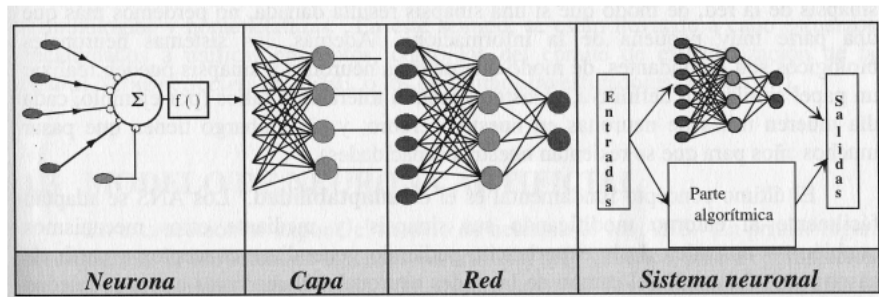


Figura 3.6: Sistema global de proceso de una red neuronal

- Un conjunto de entradas  $x_j$  con un conjunto de pesos sinápticos asociados  $w_{ij}$ , con  $j = 1, 2, \dots, n$



- Una regla de propagación  $h_i$  a definida partir del conjunto de entradas y de los pesos sinápticos. Normalmente la regla de propagación utilizada el producto lineal entre los pesos sinápticas y las entradas. Esto es:

$$h_i(x_1, \dots, x_m, w_{i1}, \dots, w_{in}) = \sum_{j=1}^n w_{ij} * x_j \quad (3.14)$$

- Una función de activación, la cual representa simultáneamente la salida de la neurona y su estado de activación. Denotando por  $y_i$  dicha función de activación:

$$y_i = f_i(h_i) = f_i\left(\sum_{j=0}^n w_{ij}x_j\right) \quad (3.15)$$

### Función de Activación

La elección de la función de activación constituye una parte determinante en el diseño de redes neuronales, dado que afectará en gran medida a la capacidad de decisión de la red y la rapidez con que la red sea capaz de converger durante el entrenamiento [59].

En general el principal requerimiento sobre estas funciones es que sean capaces de respetar el proceso de propagación hacia atrás, no provocando que el gradiente se haga cero lo cual repercutiría negativamente en el proceso de descenso en gradiente. Este es uno de los problemas asociados a la clásica función sigmoide, dado que para

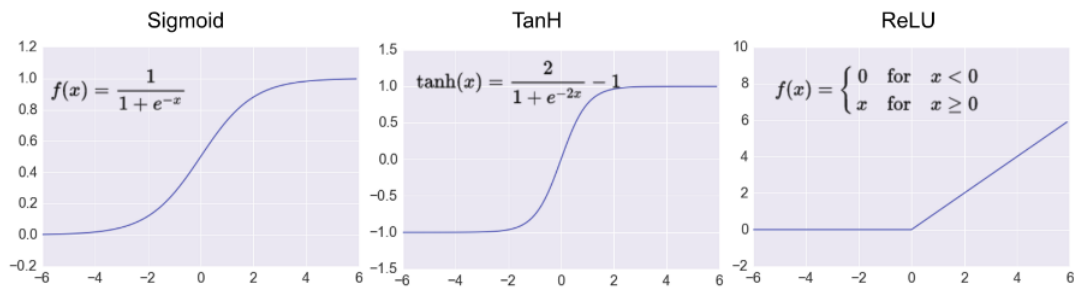


Figura 3.7: Principales funciones de activación.

valores de  $x$  ampliamente negativos o positivos, provoca que el gradiente sea cero<sup>3</sup>, interrumpiendo el descenso en gradiente para la neurona en cuestión y, por tanto, la optimización de sus pesos sinápticos.

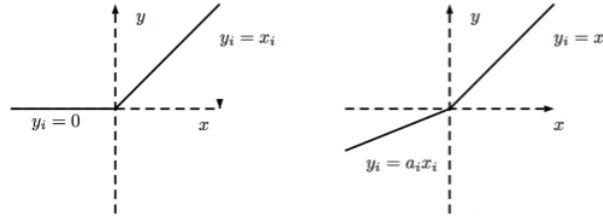


Figura 3.8: (izquierda) Función de activación *Relu*. (Derecha) Función de activación *leakyRelu*.

Actualmente la función de activación más utilizada es la unidad lineal de rectificación [60] (ReLU del inglés *Rectifier Linear Unit*), representada en la figura 3.7. No obstante, otro tipo de función de activación basada en la anteriormente expuesta denominada unidad lineal de rectificación con pérdidas (leakyRelu) ha ganado peso en el ámbito. La única diferencia entre ambas funciones es la capacidad de la *leakyRelu* de no hacer nulo el gradiente para valores negativos, ver figura 3.8 para apreciar esta diferencia. En este proyecto han sido utilizadas tanto la función Sigmoide como la función *leakyRelu*.

### Topología de Conexionado

Otro concepto determinante en el comportamiento de las redes neuronales es la topología empleada, esto es, el patrón de conexionado de una red neuronal. En una red neuronal artificial los nodos se conectan entre sí, siendo este conjunto de conexiones internas junto con los pesos sinápticos lo que determina el comportamiento de la red y, en última instancia, la función asociada a la red.

Las unidades neuronales suelen agruparse en lo que se denominan capas. La unión de dos o más capas constituyen una red neuronal. Se distinguen tres tipos de capas: de entrada, de salida y ocultas. Una capa de entrada esta compuesta por las neuronas que reciben las señales. Una capa de salida está constituida por el conjunto de neuronas que proporcionan la respuesta de la red. Las capas ocultas no tienen conexionado con el exterior. A más capas ocultas más capacidad de aprendizaje tendrá el sistema, aunque el tiempo necesario para su optimización aumentará considerablemente.

<sup>3</sup>Este efecto es comúnmente denominado como saturación

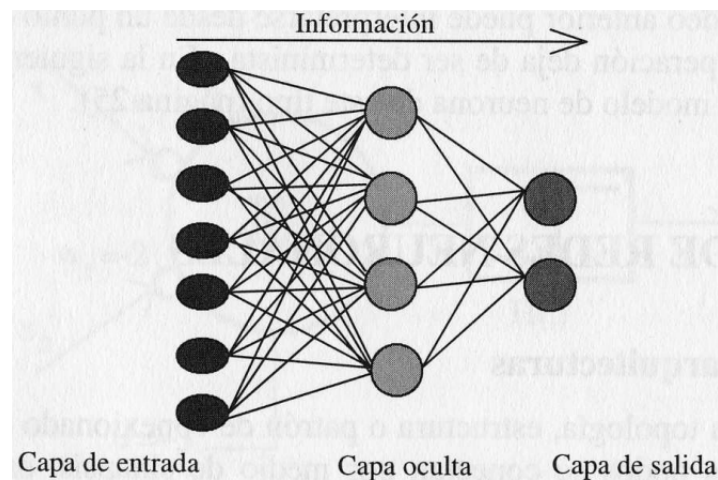


Figura 3.9: Esquema de una red neuronal densa de una sola capa oculta

### Propagación Hacia Atrás

Se denomina propagación hacia atrás al proceso empleado para el entrenamiento de las redes neuronales. Este entrenamiento tiene como objetivo el ajuste de los pesos sinápticos de la red. Se considera un buen ajuste de pesos aquel que minimiza el error a la salida de una red [61]. De manera breve los principales pasos de este proceso de entrenamiento son:

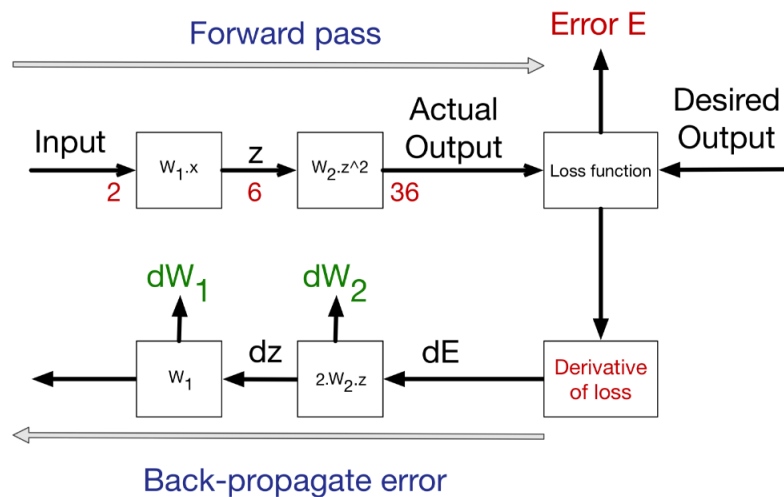
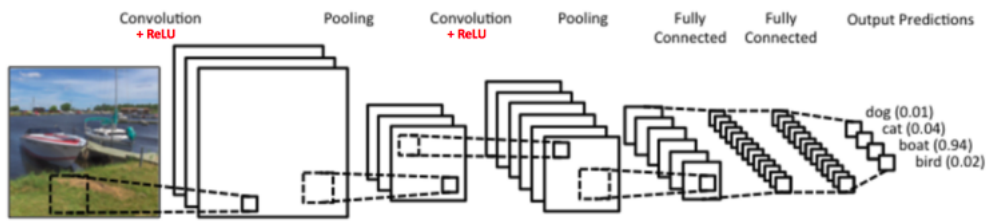


Figura 3.10: Representación esquemática del proceso de entrenamiento de una red neuronal

- Inicialización. Se asigna un valor por defecto a los distintos pesos. Se considera un paso determinante, puesto que una mala inicialización puede implicar la saturación de los gradientes en los nodos.

Los siguientes pasos constituyen un proceso iterativo, durante el cual se irá minimizando progresivamente el error asociado a la salida de la red.

- Paso hacia delante (*Fast Forward*). Se comprueba el comportamiento de la red, se calcula la salida de la red para un conjunto de muestras de entrada.
- Estimación del error de salida. Dada una salida, se evalúa la diferencia con respecto a la salida esperada según las muestras de entrada.
- Se realiza la propagación hacia atrás. Dado el error a la salida se realizan el conjunto de derivadas necesarias recorriendo desde la salida hacia la entrada la red, identificando el comportamiento del gradiente del error con respecto a los diferentes pesos de la red.
- Se modifican los pesos en función del gradiente previamente calculado.

Figura 3.11: Red convolucional *LeNet5*

## Red Neuronal Convolucional

Las redes convolucionales (CNN, del inglés *Convolutional Neural Networks*) son una categoría de redes neuronales que se consideran un método altamente eficaz en áreas como el reconocimiento de imágenes [63][64].

Este modelo fué introducido en 1989 [62] por Yann leCunn, la red de este trabajo fue denominada *LeNet5*. Dicha red se puede observar en la imagen 3.11

Las redes convoluciones suelen ser aplicadas a las imágenes. Cada imagen puede ser representada por una matriz de números sí se trata de una imagen en escala de grises, o por tres matrices sí es una imagen a color. Es esta propiedad de las imágenes donde cada dimensión, es decir cada pixel, queda definida espacialmente con respecto al resto de dimensiones, lo que convierte a las imágenes en las muestras ideales para este tipo de red.

Se asume que los conjuntos de pixeles vecinos formarán unas características más significativas que sí tomáramos grupos sin tener en cuenta su disposición espacial a capa anterior lo que se emplea es el operador de convolución.

Este tipo de redes derivan su nombre del operador de red convolución cuyo

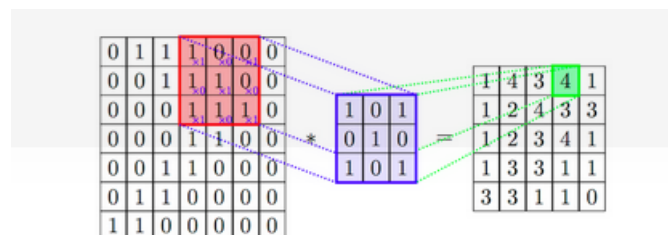


Figura 3.12: Ejemplo de aplicación del operador convolución sobre una imagen. Seleccionada una región de la imagen cuyas dimensiones son las mismas que las del kernel seleccionado, se aplica el producto pixel a pixel entre dicha región y los pesos propios del kernel. La suma de estos productos se almacena en la imagen de salida, respetando la ubicación espacial de la región evaluada.

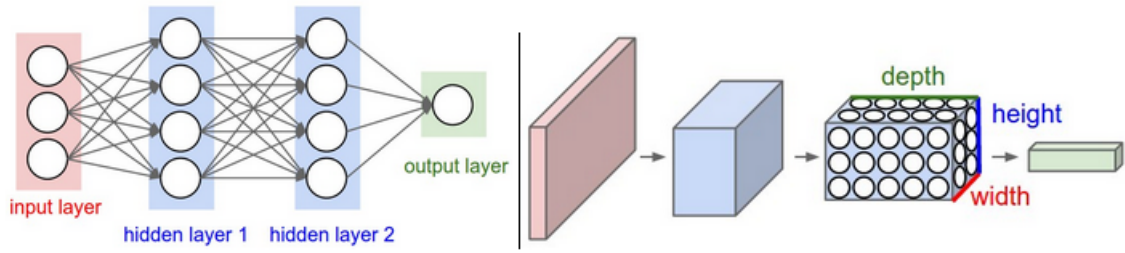


Figura 3.13: (Izquierda) Modelo clásico de redes neuronales. (Derecha) Modelo de Red Convolucional. Los datos son reagrupados en 3 dimensionados como se puede observar en una de las capas. Cada una de las capas tiene como entrada una imagen 3D y tiene como salida otra imagen 3D. La capa roja representa la capa de entrada por lo que la altura y la anchura son las dimensiones de la imagen y la profundidad son el número de canales

objetivo es extraer características de las imágenes preservando la relación espacial entre píxeles.

Dada una imagen bidimensional  $I$  y una matriz  $K$  de dimensiones  $h \times w$  (denominada kernel de convolución) la cual es capaz de extraer algún tipo de característica relevante. La operación de convolución se puede representar como:

Formalmente, se puede expresar como:

$$(I * K) : xy = \sum_{i=1}^h \sum_{j=1}^w K_{ij} I_{x+i-1, y+j-1} \quad (3.16)$$

A diferencia de las redes neuronales convencionales en las redes convolucionales los datos a la entrada y entre el conexionado de capas se agrupan en 3 dimensiones: ancho, alto y profundidad. En este caso nos referimos a "profundidad" por capa no a la profundidad de la red, lo cual se refiere al número de capas de la red en cuestión. Por ejemplo, dada una imagen de entrada de tres canales (los tres canales de color) de  $32 \times 32$  píxeles, la agrupación de los datos en la capa de entrada será  $32 \times 32 \times 3$ . Ver figura 3.13

Otra diferencia con respecto a la redes neuronales clásicas es que las unidades de una capa solo están conectadas a un espacio reducido de unidades de la capa inmediatamente anterior.

Las redes neuronales convolucionales se fundamentan en tres principios básicos que son los campos receptivos locales, los pesos compartidos y el empleo de agrupaciones o *pooling*

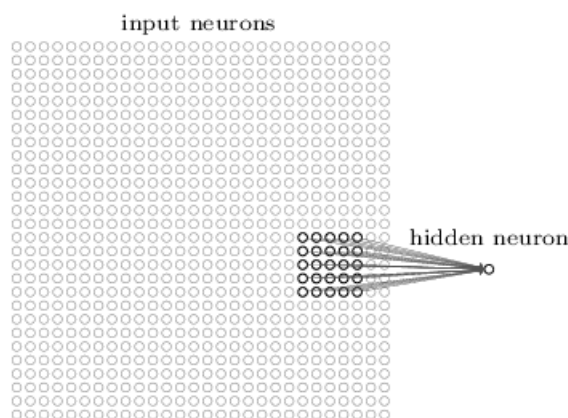


Figura 3.14: Representación de la conectividad local en una red neuronal

### Filtros Locales

En una red neuronal densa, esto es, una red totalmente conectada como la de imagen 3.13 las entradas se interpretan como un conjunto "vertical" de unidades. Sin embargo, en una red convolucional es preferible organizar las unidades de entrada en forma bidimensional.

Las capas consecutivas estarán conectadas entre sí, pero cada unidad de una capa oculta estará conectada solo a un conjunto de unidades de la capa inmediatamente anterior.

Se denomina filtro local a la ventana que se aplica a las diferentes regiones seleccionables de la imagen, cada una de estas regiones seleccionables están conectadas a una única unidad de la siguiente capa oculta. A este término a menudo nos referimos como kernel. Este filtro se desplazará por toda la imagen, realizando el proceso de

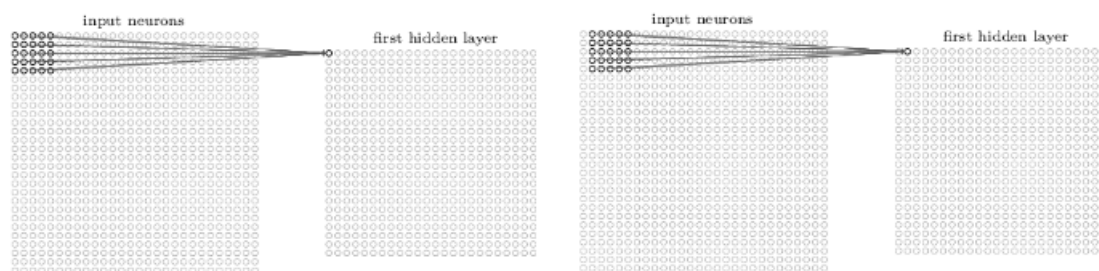


Figura 3.15: Desplazamiento del campo de recepción

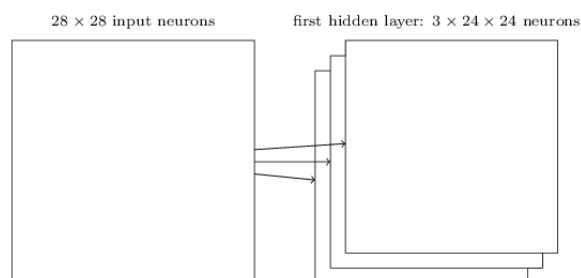


Figura 3.16: Representación de la extracción de varias características con varios filtros

convolución por toda ella, ver imagen 3.15. Es esto lo que permite extraer características de manera local por toda la imagen

Normalmente el desplazamiento se hace pixel a pixel aunque es posible aumentar el número de pixeles por desplazamiento. Este hiperparámetro se denomina generalmente *stride*. En este trabajo se ha utilizado un valor de dos. Otro concepto a tener en cuenta es que por lo general hay varios tipos de filtros para la extracción de características en las diferentes capas.

### Pesos Compartidos

Cada uno de los filtros de recepción serán aplicados a toda la imagen con el mismo peso para todas las diferentes regiones. Esto significa que el patrón de selección de características será el mismo, por lo que las neuronas de la siguiente capa detectarán el mismo tipo de característica.

El punto anterior se fundamenta en que generalmente un patrón de una parte de la imagen es probable que se repita en otra parte de la imagen dada la propia naturaleza de las imágenes.

Con objeto de no limitar cada capa a la extracción de un tipo de característica se aplican numeros filtros en cada una de las capas de convolución. Gracias a esto se consiguen extraer distintos tipos de patrones.

Una de la ventajas del uso de pesos compartidos es que permite reducir el número de parámetros de la red.

### Agrupamiento

Otro tipo de capa característica de las redes convolucionales son las capas de agrupamiento o *pooling*. Esta capa tiene como objetivo reducir el número de datos generado, realizando una estimación del valor más importante de una determinada region. Esto permite reducir progresivamente el tamaño de la imagen. Este proceso



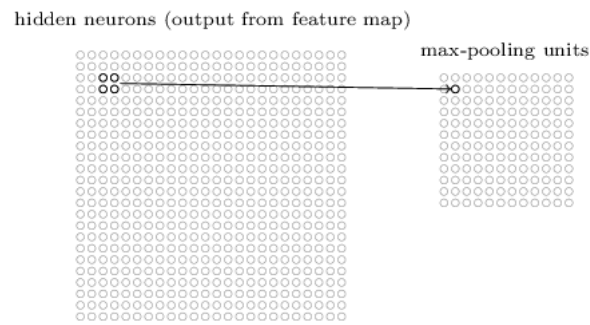


Figura 3.17: Representación del proceso de agrupamiento (*pooling*)

de agrupamiento se aplica individualmente a cada una de las imagenes generadas por cada filtro.

No obstante, esta funcionalidad no ha sido utilizada en el modelo generado en este trabajo dado que actualmente la librería empleada (*TensorFlow*) no tiene implementada esta operación para imágenes 3D.

## Herramientas Complementarias

Máquina de Vectores de Soporte

Validación Cruzada (K-fold)

Métricas de Validación

## Capítulo 4

### Trabajo Realizado



# Capítulo 5

## Resultados



# Conclusiones y líneas futuras

Después de todo el desarrollo del proyecto, es pertinente hacer una valoración final del mismo, respecto a los resultados obtenidos, las expectativas o el resultado de la experiencia acumulada.

En esta sección se exponen todos esos conceptos y enuncian unas conclusiones finales.

Además, considerando también el estado de la técnica, se pueden deducir líneas futuras de trabajo, proponer otros puntos de vista o cualquier otra sugerencia como postámbulo del presente trabajo, para ser considerada por el lector o el tribunal evaluador.

Nombre del autor  
6 de octubre de 2017





# Apéndice A

## Apéndice

### Contenido

A.1 Primera sección . . . . .	55
-------------------------------	----

### Primera sección



# Bibliografía

- [1] Henley, David B., Sundell, Karen L., Sethuraman, Gopalan, Siemers, Eric R., 2011, *Safety profile of Alzheimer's disease populations in Alzheimer's Disease Neuroimaging Initiative and other 18-month studies*, 407-416
- [2] Rodríguez Álvarez M., Sánchez J. L., 2004, *Reserva cognitiva y demencia*, 2004, vol. 20, nº 2 (diciembre), 175-186
- [3] Petrella J.R, Coleman R.E., Doraiswamy P.M., Neuroimaging and Early Diagnosis of Alzheimer Disease: A Look to the Future. *Radiology* 2003; 226:315-336.
- [4] Rémi Cuingnet et al, *Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database*, Neuroimage, vol. 56, no. 2, pp. 766-781, 2011.
- [5] W. Cai, D. Feng, R. Fulton, *Content-based retrieval of dynamic PET functional images* IEEE Trans. Inf. Technol. Biomed., vol. 4, no. 2, pp. 152-158, 2000.
- [6] Ortiz, A.; Fajardo, D.; Górriz, J.M.; Ramírez, J.; Martínez-Murcia, F.J., *Multimodal image data fusion for Alzheimer's Disease diagnosis by sparse representation*, International Conference on Innovation in Medicine and Healthcare (InMed), 2014
- [7] Daoqiang Zhanga, Yaping Wanga, Luping Zhoua, Hong Yuana, Dinggang Shena, *Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment* Neuroimage. 2011 April 1; 55(3): 856-867
- [8] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. Lopez, I Alvarez, *Compute-aided diagnosis of Alzheimer's disease using support vector machines and classification trees* Phys. Med. Biol. 55 (2010) 2807-2817
- [9] Ruaa Adeeb Abdulmunem Al-falluji *MRI based Techniques for Detection of Alzheimer: A Survey* International Journal of Computer Applications (0975 - 8887) Volume 159 - No 5, February 2017

- [10] S.Mareeswari<sup>1</sup>, Dr.G.Wiselin, *A survey Early Detection of Alzheimer's Disease using different techniques* International Journal on Computational Science and Applications (IJCSA) Vol.5, No.1,February 2015
- [11] Bellman RE, 1961, *Adaptive control processes: a guided tour*, Princeton University Press.
- [12] David L. Donoho Department of Statistics *High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality* August 8, 2000
- [13] Benson Mwangi, Tian Siva Tian, Jair C. Soares *A review of feature reduction techniques in neuroimaging* Neuroinformatics. 2014 April ; 12(2): 229-244. doi:10.1007/s12021-013-9204-3.
- [14] Siqi Liu, Sidong Liu, *EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE WITH DEEP LEARNING*
- [15] Saman S., Ghassem T., *Classification of Alzheimer's Disease Structural MRI Data by Deep Learning Convolutional Neural Networks* 22 Jul 2016
- [16] Diederik P. Kingma, Max Welling, *Auto-Encoding Variational Bayes* 1 May 2014
- [17] D. E. Rumelhart, G. E. Hinto, and R. J. Williams. Learning Internal Representations by Error Propagation 9 October 1986
- [18] : I. Guyon, G. Dror, V. Lemaire, G. Taylor and D. Silver Autoencoders, Unsupervised Learning, and Deep Architectures 2012
- [19] P. Vincent H. Larochelle I. Lajoie Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion 2010
- [20] ADVERSARIAL EXAMPLES FOR GENERATIVE MODELS
- [21] Danilo J. Rezende, Shakir Mohamed, Daan Wierstra Stochastic Backpropagation and Approximate Inference in Deep Generative Models
- [22] Bennett, D. A., Evans, D. A., 1922. Alzheimer's Disease. Disease-a-Month 38(1), 7-64.
- [23] Nakako, S., Kato, T., Nakamura., 1996. Acetylcholinesterase activity in cerebrospinal fluid of patients with alzheimer's disease and senile dementia. Journal of the Neurological Sciences 75(2).

- [24] Rodriguez-Violante M., Cervantes A., Vargas S., 2010. Papel de la función mitocondrial en las enfermedades neurodegenerativas. Arch Neurocién (Mex) Vol. 15, N°1: 39-46
- [25] Swerdlow, R., 2011 Brain aging, Alzheimer's diseases, and mitochondria. Biochim Biophys Acta 1812(12), 1630-1639
- [26] Nations U., 2008 Department of economic and social affairs, world population prospects.
- [27] Arnáiz E., Almkvist O., 2003,. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease.
- [28] Palmer K., Berger A. K., Monastero R., Winblad B., Bäckman L., Fratiglioni L. 2007. Predictors of progression from mild cognitive impairment to Alzheimer disease. Neurology 68 (19): 1596-1602. PMID 17485646. doi:10.1212/01.wnl.0000260968.92345.3f.
- [29] Carlesimo GA, Oscar-Berman M (junio de 1992). «Memory deficits in Alzheimer's patients: a comprehensive review». Neuropsychol Rev 3 (2): 119-69. PMID 1300219.
- [30] Frank EM (septiembre de 1994). «Effect of Alzheimer's disease on communication function». J S C Med Assoc 90 (9): 417-23. PMID 7967534.
- [31] Volicser L, Harper DG, Manning BC, Goldstein R, Satlin A., Mayo de 2001. Sundowning and circadian rhythms in Alzheimer's disease». Am J Psychiatry 158 (5): 704-711. PMID 11329390.
- [32] Mu Y., Gage FH, 2011 Dec, Mol Neurodegener. 2011 Dec 22;6:85. doi: 10.1186/1750-1326-6-8 Adult hippocampal neurogenesis and its role in Alzheimer's disease.
- [33] Feldman H. H., Atlas of Alzheimer's Disease. Informa Healthcare
- [34] Susanne G. Mueller, Michael W. Weiner, Neuroimaging Clin N Am. 2005 November ; 15(4): 869-xii. The Alzheimer's Disease Neuroimaging Initiative
- [35] Gorji, H. T.; Haddadnia, J. (2015-10-01). «A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI». Neuroscience. 305: 361-371. ISSN 1873-7544. PMID 26265552. doi:10.1016/j.neuroscience.2015.08.013.

- [36] Zhu, Xiaofeng; Suk, Heung-Il; Shen, Dinggang (2014-10-15). "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis". *NeuroImage*. 100: 91-105. ISSN 1095-9572. PMC 4138265. Freely accessible. PMID 24911377. doi:10.1016/j.neuroimage.2014.05.078.
- [37] Wright G., *Magnetic Resonance Imaging* *IEEE Signal Process. Mag.* 14 (1997)56-66.
- [38] Lauterbur P. C., 1973. "Image Formation by Induced Local Interactions: Examples of Employing Nuclear Magnetic Resonance". *Nature*. 242 (5394):
- [39] "Britain's brains produce first NMR scans". *New Scientist*: 588. 1978.
- [40] Snehal More, V.V.Hanchate, "A Survey on Magnetic Resonance Image Denoising Methods" *International Research Journal of Engineering and Technology (IRJET)*, Volume: 03 Issue: 05 | May-2016
- [41] Asja Fischer, Christian Igel, *An Introduction to Restricted Boltzmann Machines*
- [42] Vincent H., Larochelle H., Lajoie. I., *Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion*
- [43] Vincent, Pascal Larochelle, Hugo Bengio, Yoshua *Extracting and Composing Robust Features with Denoising Autoencoders*, (ICML'08), pages 1096 - 1103, 2008
- [44] Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014), "Generative Adversarial Networks". arXiv:1406.2661
- [45] Salimans T., Goodfellow I., Zaremba., Cheung, Vicki; Radford, Alec; Chen, Xi (2016). "Improved Techniques for Training GANs". arXiv:1606.03498
- [46] Eunbyung Park University of North Carolina at Chapel Hill
- [47] *MRI based Techniques for Detection of Alzheimer: A Survey* Ruaa Adeeb Abdulmunem Al-falluji University of Babylon, *International Journal of Computer Applications* (0975 - 8887), Volume 159 - No 5, February 2017
- [48] Yang, Wenlu, et al. "Independent component analysis- based classification of Alzheimer's disease MRI data." *Journal of Alzheimer's disease* 24.4 (2011): 775-783.

- [49] 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis? I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M.M. López, F. Segovia, R. Chaves, M. Gómez-Río, C.G. Puntonet, *Information Sciences* 181 903-916, 2011.
- [50] Herrera, Luis Javier, et al. Classification of MRI Images for Alzheimer's Disease Detection. "Social Computing (SocialCom), 2013 International Conference on. IEEE, 2013.
- [51] Zhang, H., Berg, A.C., Maire, M., Svm-knn, J.M.: Discriminative nearest neighbor classification for visual category recognition. In: CVPR '06, pp. 2126-2136. IEEE Computer Society, Los Alamitos, CA, USA (2006)
- [52] Weiner, Michael (2017). Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials". *Alzheimer's and dementia*.
- [53] Suk, Heung-Il; Lee, Seong-Whan; Shen, Dinggang; Alzheimer's Disease Neuroimaging Initiative (2016-06-01). "Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis".
- [54] Zu, Chen; Jie, Biao; Liu, Mingxia; Chen, Songcan; Shen, Dinggang; Zhang, Daoqiang; Alzheimer's Disease Neuroimaging Initiative (2016-12-01). "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment". *Brain Imaging and Behavior*.
- [55] Hosseini-Asl E., Keynton R., El-Baz A., *Alzheimer's disease diagnostics by adaptation of 3d convolutional network* (Julio 2016), arXiv:1607
- [56] Suk H-I., Shen D., *Deep Learning-Based Feature Representation for AD/MCI Classification*, Medical image computing and computer-assisted intervention (MICCAI), International Conference on Medical Image Computing and Computer-Assisted Intervention. 2013;16(0 2):583-590.
- [57] Abadi M, Agarwal A., Barham P., B. Eugene., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, November 9, 2015
- [58] D.E. Rumelhart, J.L. MacClelland (eds.) (1986). *Parallel Distributed Processing*. Vol 1. Foundations, MIT Press
- [59] Glorot, X., Bengio, Y., (2010). *Understanding the difficulty of training deep feedforward neural networks* *Journal of Machine Learning Research - Proceedings Track*. 9. 249-256.

- [60] LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015). "Deep learning". *Nature*. 521 (7553): 436–444. Bibcode:2015Natur.521..436L. PMID 26017442. doi:10.1038/nature14539.
- [61] Xavier Glorot and Yoshua Bengio, *Understanding the difficulty of training deep feedforward neural networks*, (2010) In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics
- [62] LeCun, Y., Boser, B., Denker, J. S., Backpropagation applied to handwritten zip code recognition. *Neural Comput.* , 1(4):541–551, 1989
- [63] Visualizing and Understanding Convolutional Networks Ciresan, D. C., Meier, J., and Schmidhuber, J. Multi-column deep neural networks for image classification. In *CVPR* , 2012.
- [64] Krizhevsky, A., Sutskever, I., and Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *NIPS* , 2012.