

# Índice

Prefacio .....	19
Agradecimientos.....	21
Introducción .....	23
1. Redes neuronales.....	25
1.1. Introducción .....	25
1.2. Perceptrón.....	27
1.2.1. Funciones de activación .....	33
1.2.2. Principales aplicaciones: clasificación y regresión .....	50
1.2.3. Limitaciones .....	54
1.3. Redes neuronales.....	64
1.3.1. Arquitectura.....	64
1.3.2. Aplicaciones .....	76
1.3.3. Limitaciones .....	79
2. Desarrollo de redes neuronales .....	89
2.1. Etapa de entrenamiento vs. inferencia.....	90
2.2. Paradigmas de aprendizaje.....	91
2.3. Obtención y preparación de los datos.....	95
2.4. Métricas de evaluación .....	98
2.5. Estimación del error .....	106
2.6. Aprendizaje basado en el gradiente .....	109
2.6.1. Inicialización de los pesos.....	121
2.6.2. Principales algoritmos de optimización .....	124
2.6.3. Backpropagation .....	135
2.6.4. Problemas del aprendizaje basado en gradiente.....	148

2.7. Parámetros del entrenamiento .....	156
2.8. Control y seguimiento del entrenamiento .....	163
3. Redes neuronales recurrentes .....	177
3.1. Introducción .....	177
3.2. Redes neuronales recurrentes .....	180
3.2.1. Entrenamiento de una red neuronal recurrente .....	184
3.2.2. Aplicaciones .....	188
3.2.3. Limitaciones .....	189
4. Deep learning .....	191
4.1. Introducción .....	191
4.2. Redes neuronales convolucionales .....	203
4.2.1. Arquitectura.....	212
4.2.2. Aplicaciones .....	237
4.2.3. Arquitecturas populares.....	249
Bibliografía.....	257

# Índice de figuras

1.1. Evolución del ratio entre el número de publicaciones sobre conexionismo y simbolismo .....	27
1.2. Partes de una neurona biológica .....	28
1.3. Neurona de McCulloch-Pitts.....	29
1.4. Interpretación geométrica de un perceptrón que modela la función OR .....	31
1.5. Interpretación geométrica de un perceptrón que modela la función OR con tres variables .....	34
1.6. El perceptrón.....	35
1.7. El vector normal al hiperplano definido por los pesos del perceptrón .....	37
1.8. Perceptrón (sin bias) para decidir si comprar o no una casa.....	38
1.9. Ejemplos de hiperplanos en función del valor de $w_1$ y $w_2$ (con $w_0 = 0$ ).....	39
1.10. División del espacio para diferentes hiperplanos de ejemplo .....	40
1.11. Perceptrón para decidir si comprar o no una casa a partir de 100m <sup>2</sup> .....	41
1.12. Perceptrón con dos entradas ( $x_1$ y $x_2$ ), sin sesgo ( $w_0 = 0$ ) y con función de activación identidad .....	45
1.13. Perceptrón con dos entradas ( $x_1$ y $x_2$ ), sin sesgo ( $w_0 = 0$ ) y con función de activación escalonada .....	46
1.14. Perceptrón con dos entradas ( $x_1$ y $x_2$ ), sin sesgo ( $w_0 = 0$ ) y con función de activación sigmoide.....	47
1.15. Perceptrón con dos entradas ( $x_1$ y $x_2$ ), sin sesgo ( $w_0 = 0$ ) y con función de activación tanh.....	47
1.16. Perceptrón con dos entradas ( $x_1$ y $x_2$ ), sin sesgo ( $w_0 = 0$ ) y con función de activación ReLU .....	49
1.17. Clasificación multiclase con el enfoque uno contra todos .....	52
1.18. Hiperplano (línea punteada) en un problema de clasificación (izquierda) y regresión (derecha) .....	53
1.19. Ejemplo de separación lineal mediante un hiperplano para la función OR y AND .....	55
1.20. Ejemplo de outlier en los datos.....	57
1.21. Ejemplo de datos no separables linealmente .....	57

1.22. Función XOR .....	58
1.23. Salida de perceptrón que aproxima la función OR.....	59
1.24. Salida de perceptrón que aproxima la función NOT AND.....	59
1.25. Salida de perceptrón que aproxima la función AND.....	60
1.26. Tres perceptrones aproximando la función XOR.....	60
1.27. Salida de un perceptrón que aproxima la función OR .....	61
1.28. Salida de un perceptrón que aproxima la función NOT AND.....	61
1.29. Salida construida con tres perceptrones para aproximar la función XOR .....	62
1.30. Capa de una red neuronal.....	65
1.31. Elementos de una red neuronal .....	73
1.32. Red neuronal para aproximar la función XOR.....	75
1.33. Transformación softmax y codificación one-hot.....	77
1.34. Ejemplo de la operación flattening .....	78
1.35. Efecto del cambio del valor del peso y del bias en una neurona con función de activación sigmoide.....	83
1.36. Construcción de una función «pulso» con una red neuronal.....	84
1.37. Construcción de una función «pulso» ampliado con una red neuronal .....	85
1.38. Aproximación de una función con una red neuronal .....	86
1.39. Ejemplo de función aproximada correctamente solo en algunas zonas.....	87
2.1. Etapa de entrenamiento vs. inferencia.....	91
2.2. Aprendizaje supervisado, no supervisado y por refuerzo.....	94
2.3. Efecto en la clasificación de diferentes valores de umbrales .....	100
2.4. Matriz de confusión para N clases .....	102
2.5. Matriz de confusión considerando solo las clases positivas y negativas .....	103
2.6. Métricas para un problema de clasificación de personas según están embarazadas o no.....	104
2.7. Curva ROC.....	105
2.8. Métrica IoU .....	106
2.9. Gráfica de la función ECM .....	108
2.10. Gráfica de la función CE.....	109

2.11. Superficie de la función de pérdida para un perceptrón con dos entradas.....	111
2.12. Superficie típica de la función de pérdida de una red neuronal .....	112
2.13. Evolución del vector de pesos durante el algoritmo de entrenamiento del perceptrón.....	115
2.14. Representación gráfica de una función y su derivada.....	116
2.15. Posibles puntos críticos de una función .....	117
2.16. Efecto en la definición de la superficie de error de las diferentes estrategias de optimización .....	120
2.17. Regiones de la superficie del error ideales para comenzar el entrenamiento .....	122
2.18. Modificación de la variable para acercarse al mínimo de la función.....	126
2.19. Modificación de los pesos por el algoritmo descenso por gradiente .....	127
2.20. Comparativa del algoritmo de descenso por gradiente con y sin momento .....	128
2.21. Función de dos variables y plano de sus curvas de nivel .....	129
2.22. Reducción de la oscilación por el uso del momento .....	130
2.23. Ejemplo de iteraciones del algoritmo descenso por gradiente .....	132
2.24. Ejemplo de grafo de operaciones .....	142
2.25. Ejemplo completo de red neuronal para el problema XOR .....	143
2.26. Grafo de operaciones para la red neuronal de la Figura 2.25 .....	146
2.27. Cálculo de una iteración del algoritmo descenso por gradiente sobre el grafo de operaciones de una red .....	147
2.28. Gráfica de las derivadas de las principales funciones de activación.....	149
2.29. Ejemplo de sobreajuste (izquierda), buen ajuste (centro) y infraajuste (derecha) en base a un conjunto de datos de ejemplo (color morado) .....	152
2.30. Posibles escenarios de un modelo respecto al sesgo y la varianza .....	154
2.31. Búsqueda de compromiso entre sesgo y varianza al definir la complejidad del modelo .....	155
2.32. Ciclo de vida del desarrollo de redes neuronales.....	157

2.33. Relación entre los datos de un conjunto de datos, el tamaño del lote, la iteración y la época de entrenamiento.....	158
2.34. Diferencias entre un mínimo pronunciado y aplanado de una función de pérdida definida por un lote de ejemplos y la definida por todos los datos .....	161
2.35. Gráficas habituales del entrenamiento de una red neuronal.....	165
2.36. Detección de sobreajuste e infraajuste en la gráfica de pérdida .....	166
2.37. Efecto del valor de learning rate en la gráfica de pérdida .....	167
2.38. Diferencias en el descenso por gradiente con un valor para el learning rate fijo y variable.....	169
2.39. Superficie de la función de pérdida con datos normalizados y sin normalizar .....	170
2.40. Efecto de la regularización en la complejidad del modelo .....	171
2.41. Red con y sin dropout .....	173
2.42. Efecto de la limitación del gradiente durante el entrenamiento .....	175
3.1. Sin información de los instantes anteriores, no se puede predecir la siguiente posición del vagón.....	178
3.2. En base a la información anterior, se puede predecir que el vagón se moverá hacia la derecha.....	179
3.3. Esquema de una red neuronal recurrente.....	181
3.4. Esquema de una red neuronal recurrente multicapa.....	183
3.5. División en ventanas de la secuencia de datos y uso en cada iteración de entrenamiento para entrenar (E), validar (V) y pruebas (P) .....	186
3.6. Proceso de desenrollado del grafo de operaciones para dos iteraciones de una red neuronal recurrente .....	187
4.1. Relación entre inteligencia artificial, machine learning y deep learning.....	191
4.2. Comparativa entre el desarrollo de un sistema basado en técnicas de machine learning y deep learning.....	193
4.3. Descomposición de los píxeles de una imagen en una nueva representación jerárquica de características de menor a mayor complejidad .....	195

4.4. Esquema completo de un modelo de deep learning con sus diferentes etapas.....	196
4.5. Evolución del porcentaje de error cometido por el modelo ganador de ImageNet a lo largo de los últimos años .....	197
4.6. Comparativa del rendimiento de los modelos de machine learning y deep learning en base al tamaño del conjunto de datos disponible .....	199
4.7. Coste del almacenamiento de los datos frente a su disponibilidad.....	200
4.8. De izquierda a derecha: Yann LeCun, Geoffrey Hinton y Yoshua Bengio .....	201
4.9. Panorama industrial del deep learning .....	202
4.10. Experimentos llevados a cabo por Hubel y Wiesel .....	204
4.11. Localización de la corteza visual en el cerebro .....	204
4.12. Jerarquía de características 0que obtiene el cerebro a partir de la información visual de la retina .....	208
4.13. Células S que comparten pesos para extraer la misma característica en diferentes regiones de la entrada.....	209
4.14. Estructura del Neocognitron .....	209
4.15. Características extraídas por la primera capa de una red neuronal convolucional. Arriba se muestra la imagen que produce mayor activación de la característica y, debajo, ejemplos de imágenes reales del conjunto de datos que producen una alta activación. ....	210
4.16. Características extraídas por la segunda capa de una red neuronal convolucional. A la izquierda se muestra la imagen que produce mayor activación de la característica y, a la derecha, ejemplos de imágenes reales del conjunto de datos que producen una alta activación. ....	211
4.17. Características extraídas por la tercera capa de una red neuronal convolucional. A la izquierda se muestra la imagen que produce mayor activación de la característica y, a la derecha, ejemplos de imágenes reales del conjunto de datos que producen una alta activación. ....	211
4.18. Ejemplos de diferentes datos, su representación y la denominación que tienen en base a su dimensión .....	213
4.19. Ejemplo gráfico del cálculo de la convolución de dos funciones continuas .....	221
4.20. Ejemplo gráfico del cálculo de la convolución de dos funciones discretas.....	222

4.21. Representación de la convolución de dos piezas de información de una dimensión con tamaño limitado y diferente .....	223
4.22. Representación de la convolución de dos piezas de información con relleno.....	224
4.23. Representación de la convolución de dos piezas de información con un paso de 2 .....	225
4.24. Representación de la convolución de dos piezas de información de dos dimensiones.....	226
4.25. Diferentes convoluciones sobre una imagen de tres dimensiones (canales rojo, verde y azul) .....	227
4.26. Información relativa a una capa de convolución .....	228
4.27. Comparativa entre una capa de una red neuronal y una capa de convolución .....	229
4.28. Mejora de los resultados con el aumento de la profundidad de la red Error Top-5 es el error cometido por la red al no estar la clase esperada entre las 5 primeras clases del modelo. ....	230
4.29. Ejemplo del cálculo de max-pooling y average pooling .....	232
4.30. Ejemplo de bloque inception.....	234
4.31. Ejemplo de skip connection .....	235
4.32. Ejemplo de dense block.....	236
4.33. Principales casos de uso del deep learning.....	238
4.34. De izquierda a derecha: clasificación, de- tección y segmentación .....	239
4.35. Red de conducción autónoma de Tesla: HydraNet .....	239
4.36. Red neuronal convolucional y LSTM usadas para describir una imagen .....	240
4.37. Reescalado de imágenes .....	241
4.38. Reconstrucción y coloreado de imágenes .....	241
4.39. Representación invariante de objetos .....	242
4.40. Transferencia del estilo de una imagen a otra .....	242
4.41. Estimación de la profundidad.....	243
4.42. Estimación de la pose .....	243
4.43. Modelo basado en deep learning entrenado por aprendizaje por refuerzo para jugar a los juegos de Atari .....	244



4.44. Modelo basado en deep learning entrenado por aprendizaje por refuerzo para jugar al juego del Go.....	245
4.45. Conducción autónoma con modelo de deep learning entrenado con aprendizaje por refuerzo .....	245
4.46. Manipulación de objetos con modelo de deep learning entrenado por aprendizaje por refuerzo .....	246
4.47. Procesamiento del lenguaje haciendo uso de una red neuronal convolucional .....	246
4.48. Reconocimiento del audio haciendo uso de una red neuronal convolucional combinada con una red neuronal recurrente .....	247
4.49. Ejemplos de segmentaciones por instancias realizadas con la red Mask R-CNN.....	248
4.50. Arquitectura de la red LeNet-5 .....	250
4.51. Arquitectura de la red AlexNet .....	251
4.52. Arquitectura de la red GoogleNet.....	252
4.53. Arquitectura de las redes ResNet-50, ResNet-101 y ResNet-152 .....	254
4.54. Arquitectura de la red DenseNet-121 .....	256

## Índice de tablas

1.1. Salidas de una neurona de McCulloch-Pitts para el Ejemplo 1.1 con 3 entradas $y = 3$ .....	42
1.2. Principales funciones de activación.....	44
1.3. Funciones booleanas diferentes para dos variables.....	56
1.4. Funciones booleanas diferentes, separables linealmente y ratio en función del número de variables .....	56
1.5. Entradas y salidas de cada uno de los perceptrones de la red neuronal para aproximar la función XOR .....	63

## Índice de algoritmos

1. Algoritmo de entrenamiento del perceptrón .....	114
2. Algoritmos basados en el gradiente .....	124