

Índice general

I	Introducción General	1
1.	Introducción general	5
1.1.	Organización del libro	6
2.	Paradigmas de Aprendizaje Automático	9
2.1.	Introducción	9
2.2.	Validación	11
2.3.	Búsqueda en el espacio de modelos	14
	Bibliografía: libros relacionados	16
II	Técnicas de Clasificación Supervidada	19
3.	Algoritmos de clasificación por vecindad	23
3.1.	Introducción	23
3.1.1.	Regla del vecino más próximo	25
3.1.2.	Regla de los K vecinos más próximos (K-NN)	25
3.2.	Variantes del algoritmo K-NN	26
3.2.1.	Regla K-NN con rechazo	27
3.2.2.	Regla K-NN por distancia media	27
3.2.3.	Clasificador de la distancia mínima	27
3.3.	Pesado de casos y de atributos	30
3.3.1.	Pesado de casos	30
3.3.2.	Pesado de atributos	31
3.4.	Reducción del conjunto de entrenamiento	32
3.4.1.	Selección de prototipos	33
3.4.2.	Técnicas de edición	34
3.4.3.	Técnicas de condensado	35
	Bibliografía sobre clasificación basada en la distancia	38
4.	Árboles de clasificación	41
4.1.	Introducción	41
4.2.	Un poco de historia	41

4.3.	Descripción general de los árboles de clasificación	42
4.4.	Profundizando un poco	45
4.5.	Selección de la división	46
4.6.	Árboles parsimoniosos	51
4.7.	Información incompleta: casos <i>missing</i>	53
4.8.	Los más conocidos	53
	Bibliografía sobre Árboles de Clasificación	56
5.	Aprendizaje de reglas de decisión	59
5.1.	Introducción	59
5.2.	El algoritmo AQ	61
5.3.	El algoritmo CN2	64
5.4.	Algoritmos más recientes, otros enfoques	65
5.5.	Aplicación práctica	68
5.5.1.	El dominio de aplicación	68
5.5.2.	Aprendizaje supervisado para modelar al usuario	69
5.6.	Conclusiones	74
	Bibliografía sobre Inducción de Reglas	74
6.	Redes Bayesianas	77
6.1.	Introducción	77
6.2.	Redes bayesianas	77
6.2.1.	Inferencia	80
6.3.	Aprendizaje de clasificadores bayesianos	84
6.3.1.	Clasificador bayesiano simple	84
6.3.2.	Extensiones al clasificador bayesiano	85
6.3.3.	Mejora estructural de un clasificador bayesiano	86
6.4.	Aprendizaje de redes bayesianas	88
6.4.1.	Aprendizaje paramétrico	88
6.4.2.	Aprendizaje estructural	91
6.4.3.	Aprendizaje de árboles	91
6.4.4.	Aprendizaje de poliárboles	92
6.4.5.	Aprendizaje de redes	93
6.5.	Aprendizaje de redes bayesianas dinámicas	96
6.5.1.	Redes bayesianas dinámicas	96
6.5.2.	Aprendizaje	97
6.6.	Lecturas adicionales	98
	Bibliografía sobre Redes Bayesianas	98
7.	Introducción a las Redes Neuronales	101
7.1.	Introducción	101
7.2.	Generalidades	101
7.3.	El perceptrón simple	103
7.3.1.	Regla de adaptación del perceptrón	104
7.4.	Neuronas lineales y el algoritmo LMS	105
7.4.1.	Ejemplo: cancelación de ruido en registros de electrocardiograma	109
7.5.	Funciones de activación no lineales, Perceptrones multicapa y el algoritmo de Retropropagación	110

7.5.1.	Ajuste de pesos en la capa de salida	112
7.5.2.	Ajuste de pesos en la capa oculta	113
7.5.3.	Algunas mejoras al algoritmo de retropropagación.	114
7.5.4.	Ejemplo: aproximación de funciones.	115
7.6.	Memoria Asociativa y Redes de Hopfield	116
7.6.1.	Almacenamiento de patrones en los pesos de la red	118
7.6.2.	Capacidad de almacenamiento	119
7.6.3.	Atractores espúreos	120
7.6.4.	Funciones de energía y optimización	121
7.6.5.	Ejemplo: almacenamiento de patrones bidimensionales.	121
7.7.	Redes Auto-organizadas	122
7.7.1.	Aprendizaje Hebbiano Generalizado	122
7.7.2.	Ejemplo: direcciones principales en un conjunto de datos	125
7.7.3.	Aprendizaje Competitivo	125
7.7.4.	Algoritmo para aprendizaje competitivo	127
7.7.5.	Ejemplo: clasificación de datos bidimensionales	128
7.7.6.	Comentarios finales y síntesis histórica	129
	Bibliografía sobre Redes Neuronales	130
8.	Modelos Ocultos de Markov	133
8.1.	Descripción del modelo	134
8.2.	Arquitectura de los HMM	135
8.2.1.	Modelos de izquierda a derecha	135
8.2.2.	Modelos ergódicos	135
8.3.	Tipos de HMM	136
8.3.1.	HMM discretos	136
8.3.2.	HMM continuos	136
8.3.3.	HMM semicontinuos	137
8.4.	Problemas básicos	137
8.4.1.	Primer problema básico: problema de la evaluación.	138
8.4.2.	Segundo problema básico: problema de la decodificación.	139
8.4.3.	Tercer problema básico: problema del entrenamiento.	139
8.5.	Algunas consideraciones previas	139
8.5.1.	Definiciones	139
8.5.2.	Notación	140
8.5.3.	Proposiciones	140
8.6.	El problema de la evaluación	141
8.6.1.	Algoritmo <i>Forward-Backward</i>	141
8.7.	El problema del alineamiento	145
8.7.1.	Principio de optimalidad de Bellman	145
8.7.2.	Algoritmo de Viterbi	147
8.8.	El problema del entrenamiento	148
8.8.1.	Estimación de los parámetros	149
8.8.2.	Algoritmo EM	150
8.8.3.	Algoritmo de reestimación de Baum-Welch	151
8.8.4.	Algoritmo de Viterbi	157

8.8.5. Algunas consideraciones	158
Bibliografía sobre Modelos Ocultos de Markov	160
9. Métodos Kernel y Máquinas de Vectores Soporte	163
9.1. Introducción	163
9.2. Kernels	164
9.2.1. La función kernel	164
9.2.2. Ejemplos de funciones kernel	166
9.2.3. Combinación de funciones kernel	169
9.3. Métodos Kernel	169
9.3.1. La ' <i>kernelización</i> ' de algoritmos	169
9.3.2. Perceptrón: versión con kernels	171
9.4. Máquinas de Vectores Soporte	174
9.4.1. Introducción	174
9.4.2. Hiperplano de margen máximo	176
9.4.3. Optimización de funciones	179
9.4.4. Hiperplano de margen máximo (continuación)	183
9.4.5. Propiedades de la solución	185
9.4.6. Margen blando	186
9.4.7. Regularización	189
9.4.8. Vectores soporte para regresión	191
9.4.9. Algunas implementaciones	195
9.5. Lecturas adicionales	196
Bibliografía sobre SVM	198
10.Programación Lógica Inductiva (ILP)	201
10.1. Introducción	201
10.2. Nociones de lógica y notación utilizada	203
10.3. Programación Lógica Inductiva (ILP)	204
10.3.1. Búsqueda de hipótesis	205
10.3.2. Sistemas de específico a general	206
10.3.3. Sistemas de general a específico	211
10.3.4. Restricciones y técnicas adicionales	214
10.3.5. Algunas extensiones recientes	215
10.3.6. Aplicaciones	216
Bibliografía sobre ILP	216
III Otros paradigmas dentro del Aprendizaje Automático	223
11.Aprendizaje por refuerzo	227
11.1. Introducción	227
11.2. El aprendizaje por refuerzo y los procesos de decisión de Markov	228
11.2.1. Procesos de decisión de Markov	229
11.2.2. Funciones de valor	231
11.3. Métodos de resolución tradicionales	233
11.3.1. Programación dinámica	233
11.3.2. Métodos libres de modelo	239

11.3.3. Métodos basados en el modelo	244
11.3.4. Exploración y explotación	245
11.4. Generalización en aprendizaje por refuerzo	246
11.4.1. Generalización y aproximación de funciones	247
11.4.2. Métodos no lineales de estimación	249
11.4.3. Modelos de representación del espacio de estados	252
11.4.4. Aprendizaje de la representación	254
11.4.5. Generalización de Acciones	256
Bibliografía sobre Aprendizaje por Refuerzo	256
12. Clasificación: Análisis de clusters (Clustering)	261
12.1. Introducción	261
12.2. Distancias y similitudes	262
12.2.1. Distancias para variables continuas	264
12.2.2. Similitudes para variables binarias	266
12.2.3. Similitud para variables mixtas	267
12.3. Métodos jerárquicos	268
12.3.1. Dendrograma, jerarquía indexada y ultramétrica	269
12.3.2. Algoritmo básico de clasificación (ABC)	271
12.3.3. Algoritmo de clasificación (AC)	273
12.3.4. Método del mínimo	274
12.3.5. Método del máximo	276
12.3.6. Método de Ward	278
12.3.7. Fórmula de Lance-Williams (1967)	281
12.3.8. Cluster basado en distancias (DB-Cluster)	282
12.4. Aplicaciones de los métodos jerárquicos	284
12.5. Los coches y sus diferentes gamas	284
12.6. Clasificación de especies del género <i>Dianthus L.</i>	286
12.7. Métodos Particionales	286
12.7.1. Criterios	287
12.7.2. Optimización del criterio	289
12.7.3. Algoritmo de k-medias	290
12.7.4. Ejemplo real	291
Bibliografía sobre Clasificación no Supervisada	292
13. Algoritmos Evolutivos	295
13.1. Introducción	295
13.2. Un Algoritmo Genético en acción	296
13.3. Componentes de Algoritmos Evolutivos	297
13.3.1. Función Objetivo	298
13.3.2. Selección	298
13.3.3. Representaciones	300
13.3.4. Operadores de Variación: Mutación	301
13.3.5. Operadores de Variación: Cruzamiento	303
13.4. Otros Algoritmos Evolutivos	304
13.5. Conclusiones	305
Bibliografía sobre Algoritmos Evolutivos	306

14. Análisis Discriminante y Regresión Logística	309
14.1. Introducción	309
14.2. Análisis Discriminante Descriptivo	310
14.2.1. Planteamiento del problema	311
14.2.2. Criterio discriminante de Fisher y Análisis Discriminante Canónico	311
14.2.3. Determinación de las dimensiones de discriminación	315
14.2.4. Correlaciones canónicas	316
14.2.5. Interpretación de los resultados	316
14.3. Análisis Discriminante Predictivo	328
14.3.1. Planteamiento del problema: la regla Bayes	330
14.3.2. Discriminación bajo hipótesis de normalidad	330
14.3.3. Estimación de la regla Bayes	331
14.3.4. Evaluación del procedimiento de clasificación	332
14.3.5. Ejemplos	333
14.4. Regresión Logística	335
14.4.1. Discriminación de dos grupos: Regresión Logística Binomial	336
14.4.2. Discriminación de tres o más grupos: Regresión Logística Multinomial	337
14.4.3. Estimación y contraste de los parámetros del modelo	337
14.4.4. Procedimiento de clasificación	338
14.4.5. Ejemplos	338
14.5. Selección de variables	340
14.5.1. Criterios de selección de modelos	342
14.5.2. Algoritmos de selección de variables	343
14.5.3. Ejemplos	344
Bibliografía sobre Análisis Discriminante	350

IV Conceptos relacionados **353**

15. Discretización de atributos continuos	357
Bibliografía sobre métodos de discretización	360
16. Descomposición en Valores Singulares	363
16.1. Introducción.	363
16.2. Fundamentos de LSI	364
16.2.1. Construcción de la matriz	364
16.2.2. Descomposición en valores singulares de la matriz	364
16.2.3. Reducción de la dimensión del espacio	365
16.2.4. Búsqueda semántica	365
16.3. Ejemplo	366
16.4. Aplicaciones: Procesamiento del Lenguaje Natural	367
16.4.1. Aplicación a la categorización de documentos.	368
16.5. Conclusiones	371
Bibliografía relacionada con LSI y Lenguaje Natural	372

V Aspectos avanzados y de investigación	375
17. Combinación de clasificadores	379
17.1. Introducción	379
17.2. Fundamentos teóricos	380
17.3. Métodos generales de construcción de conjuntos de clasificadores	382
17.3.1. Manipular el conjunto de entrenamiento	383
17.3.2. Manipular los elementos de entrada	383
17.3.3. Manipular los elementos de salida	384
17.3.4. Introducir aleatoriedad	384
17.4. Combinación de predicciones	384
17.4.1. Voto mayoritario	385
17.4.2. Voto mayoritario ponderado	385
17.4.3. Voto bayesiano	385
17.5. Medidas de diversidad	386
17.5.1. Medida E de Entropía	386
17.5.2. Diversidad de Kohavi-Wolpert	387
17.6. Bagging	387
17.7. Boosting	389
17.8. Stacking	391
17.9. Meta decision trees	393
17.10 Autores recomendados	393
Bibliografía sobre Combinación de Clasificadores	394
18. Clasificadores Híbridos	397
18.1. Multclasificador híbrido para el comportamiento de cruzar puertas: Bayes-nearest	397
18.1.1. Planteamiento del problema	398
18.1.2. Recopilación de datos para el aprendizaje	399
18.1.3. Estructura de la red bayesiana	399
18.1.4. Propagación de la evidencia	401
18.1.5. Resultados preliminares	401
18.1.6. Descripción de la implementación en el robot y evaluación del comporta- miento	402
18.1.7. Validación del multclasificador híbrido Bayes-nearest	406
Bibliografía sobre Clasificadores Híbridos	406
19. Aspectos avanzados en Árboles de clasificación	409
19.1. Introducción	409
19.2. Aumentar la capacidad discriminante	409
19.3. Nuevas funciones o criterios de división	410
19.4. Introducción de aleatoriedad en la construcción	410
19.5. Combinación de árboles	411
19.6. Eliminación de limitaciones representacionales	412
19.7. Manejar un gran volumen de datos <i>large databases</i>	413
19.8. Estabilidad estructural y capacidades explicativas	413
19.9. Árboles consolidados: árboles basados en múltiples submuestras sin renunciar a la explicación	414
19.10 Algoritmo de construcción de árboles consolidados	414

19.10.1 Resultados experimentales	417
Bibliografía sobre aspectos avanzados de Árboles de Clasificación	422
20. Una aplicación de las SVM	427
20.1. Introducción	427
20.2. Aprender a ordenar es útil	427
20.3. Cómo se puede aprender a ordenar	428
20.4. La regresión no se debe utilizar	429
20.5. Cómo aprender funciones de ranking, preferencias o utilidad	430
20.5.1. Funciones lineales a partir de un conjunto de juicios de preferencias	431
20.5.2. Funciones no lineales	434
20.5.3. Implementación directa como clasificación con una clase	437
20.6. Regresión ordinal	437
20.7. Aplicaciones	438
20.7.1. Selección de ganado vacuno de carne	439
20.7.2. Análisis de la calidad sensorial de los alimentos	439
20.7.3. Otras aplicaciones	443
Bibliografía avanzada sobre SVM	444
21. Medición del ritmo cardíaco fetal usando Análisis de Componentes Independientes (ICA)	447
21.1. Introducción	447
21.2. Naturaleza del problema. El caso del 'cocktail party'	447
21.3. Consideraciones preliminares. El modelo ICA lineal	448
21.4. Independencia estadística y el teorema del límite central	449
21.5. Medidas de distancia a la densidad normal	450
21.5.1. Kurtosis	450
21.5.2. Entropía	451
21.6. Información mutua, Redes Neuronales e ICA	451
21.7. Separación del electrocardiograma fetal	454
21.7.1. Técnicas avanzadas de ICA	457
Bibliografía Relacionada con la aplicación de Redes Neuronales	458
22. Algoritmo Genético para el pesado de atributos	461
22.1. Reconocimiento de paneles de salidas de emergencia	461
22.2. Sistema de percepción activa	466
22.2.1. Resultados experimentales del sistema de visión activa	469
Bibliografía relacionada	470
VI Aplicaciones Software para Aprendizaje Automático: WEKA	473
23. Una aproximación al software WEKA	477
23.1. Generalidades acerca de WEKA	477
23.2. Los distintos modos de uso	478
23.3. <i>Explorer</i> : seis tareas básicas para la Minería de Datos	479
23.4. Proyectos de software relacionados	481
Bibliografía: software disponible	483