

Índice

Prefacio	17
1 Introducción	21
1.1 Extracción del conocimiento y minería de datos	21
1.2 Análisis de redes. Recorrido histórico	26
1.3 Aplicaciones	30
1.4 Software libre para el data science	32
I Inteligencia Artificial y Análisis Masivo de Datos	37
2 Bases algebraicas y estadísticas	39
2.1 Introducción	39
2.2 Álgebra lineal	41
2.2.1 Espacios vectoriales reales	42
2.2.2 Matrices y transformaciones lineales	45

2.2.3	Matrices simétricas. Descomposición espectral . . .	49
2.2.4	Proyección ortogonal y método de los mínimos cuadrados	52
2.3	Teoría de la probabilidad y variable unidimensional . . .	54
2.3.1	Variable aleatoria	59
2.3.2	Principales parámetros unidimensionales	64
2.4	Datos multivariantes	66
2.4.1	Representación de datos multivariantes. Principales parámetros	72
2.4.2	Medidas de variabilidad, asociación y dependencia	73
2.5	Distancia y similitud	77
3	Minería de datos I. Reducción y Visualización	83
3.1	Introducción	83
3.2	Análisis por Componentes Principales	84
3.2.1	Reducción de la dimensión	88
3.3	Análisis Factorial	95
3.3.1	Método del factor principal	98
3.3.2	Reducción de la dimensión	100
3.4	Escalado Multidimensional (MDS)	106
3.4.1	Escalado Multidimensional No Métrico	113
3.5	Notas finales	118

4	Minería de datos II. Clustering	121
4.1	Introducción	121
4.2	Algoritmos jerárquicos. Dendrogramas	124
4.3	Algoritmos particionales. <i>K</i> -medias	129
4.4	Algoritmos basados en densidad. DBSCAN	133
4.4.1	Elección de parámetros	138
4.5	Algoritmos borrosos. <i>c</i> -medias borroso	140
4.6	Evaluación y validación de los algoritmos de clustering	147
4.6.1	Índices externos	149
4.6.2	Índices internos	155
4.6.3	Validación de algoritmos de clustering jerárquicos	160
5	Minería de datos III. Clasificación Supervisada	165
5.1	Introducción	165
5.1.1	Metodología general de la clasificación supervisada	166
5.2	Indicadores de discriminación. Análisis ROC	169
5.3	Análisis Discriminante	183
5.3.1	Métodos bayesianos de clasificación. Discriminador Lineal de Fisher y Método Naïve Bayes .	183
5.3.2	Clasificador lineal en dos poblaciones normales	185
5.3.3	Clasificador lineal en poblaciones con distribución desconocida	187

5.3.4	Variables canónicas discriminantes. Reducción de la dimensión	188
5.4	Regresión logística	195
5.4.1	Regresión lineal	196
5.4.2	Regresión logística	199
5.4.3	Interpretación de los coeficientes	204
5.5	Máquinas de vectores soporte	207
5.5.1	Datos no linealmente separables. Primera aproximación	211
5.5.2	Funciones kernel. Segunda aproximación	212
5.6	Otros métodos de clasificación	217
5.6.1	Árboles de discriminación	217
5.6.2	k Vecinos más cercanos	228
5.7	Notas finales	231

II Teoría de Grafos y Análisis de Redes 233

6 Teoría de Grafos. Algoritmos instrumentales 235

6.1	Introducción y preliminares	235
6.1.1	La Clase Grafo	237
6.2	Algoritmos de exploración de grafos	241
6.2.1	Recorrido de árboles	241

6.2.2	Recorrido de un grafo cualquiera	244
6.2.3	Otros algoritmos instrumentales para el análisis de redes	247
6.2.4	Curvas geodésicas	251
6.3	Cadenas de Markov	259
6.3.1	Cadenas de Markov en tiempo discreto	260
6.3.2	Comportamiento de transición	263
6.3.3	Comportamiento estacionario	267
6.3.4	Clasificación de estados y teoremas límite	270
7	Redes y Sistemas Complejos	281
7.1	Introducción	281
7.2	Librerías Python para el análisis de redes sociales	282
7.2.1	Redes para la experimentación	287
7.2.2	Formato estándar para grafos	288
7.3	Visualización de redes. Algoritmos de fuerzas dirigidas	290
7.4	Propiedades emergentes en redes complejas	297
7.4.1	Redes de Erdős-Renyi	298
7.4.2	Redes de Watts-Strogatz	303
7.4.3	Generación y crecimiento de redes. Redes expo- nenciales	311
7.4.4	Redes de Barabasi-Albert	314

7.5	Notas finales	318
8	Centralidad en redes	319
8.1	Introducción	319
8.2	Protagonistas de la red. Centralidad de cercanía	324
8.3	Intermediarios. Centralidad de Intermediación	325
8.4	Actores en la sombra. Centralidad de autovalor	330
8.5	Emisores y receptores. Centralidad de hubs y authorities	331
8.6	Centralidad de <i>PageRank</i>	336
8.6.1	Algoritmo de Brin y Page	336
8.7	Centralidad de Katz-Bonacich	340
8.8	Notas finales	340
9	Detección de comunidades	343
9.1	Introducción	343
9.2	Métodos clásicos de detección de comunidades	346
9.2.1	Particionamiento de grafos	347
9.2.2	Métodos de clustering jerárquicos	350
9.2.3	Métodos de clustering particional	352
9.2.4	Métodos de clustering espectral	352
9.3	Algoritmos divisivos. Método de Girvan - Newman	355
9.4	Funciones de calidad. Algoritmo de Modularidad	358

9.4.1	División en más de dos comunidades	362
-------	--	-----

III Aplicaciones 365

10 Análisis de riesgos en modelos de interdependencia 367

10.1	Introducción	367
10.2	El método DEMATEL en análisis de riesgos	368
10.3	Cálculo alternativo del riesgo repercutido	372
10.4	Riesgo repercutido de impago en relaciones comerciales	374
10.5	Riesgo repercutido de incendio en un bosque	378
10.6	Notas finales	380

11 Sistemas de recomendación 383

11.1	Introducción	383
11.2	Recomendadores basados en filtrado colaborativo . . .	385
11.3	Recomendadores basados en contenidos	389
11.4	Notas finales	393

12 Procesamiento del lenguaje natural 397

12.1	Introducción	397
12.2	Representación gráfica de textos. Nube y red de palabras	398
12.3	Clasificación de textos	403

12.4	Notas finales	410
13	Minería de Sonidos e Imágenes	413
13.1	Introducción	413
13.2	Separación ciega de fuentes	414
13.3	Clasificación de melodías en géneros musicales	422
13.3.1	Transformada de Fourier	429
13.3.2	Extracción de características y clasificación	432
13.4	Tratamiento y clasificación de imágenes	437
13.4.1	Sistemas de reconocimiento facial	438
13.4.2	Notas finales	445
	Bibliografía	447