

Uma breve introdução à MINERAÇÃO DE DADOS

Joaquim V. C. Assunção

Novatec

Sumário

Prefácio	5
1 Visão Geral – Subdomínios da Inteligência Artificial	9
1.1 Aprendizagem de máquina	11
1.2 Mineração de dados	11
1.3 Considerações do capítulo	15
1.4 Exercícios	16
2 Dado, Informação e Conhecimento	17
2.1 Decisão por dados	17
2.2 Tipos de dados	21
2.3 Tipos de atributos	22
2.4 Análise de dados	23
2.5 Considerações do capítulo	39
2.6 Exercícios	41
3 Pré-mineração	43
3.1 Extração/seleção	43
3.2 Pré-processamento	45
3.3 Integração e transformação	49
3.4 Exercícios	62
4 Regras de Associação	65
4.1 Algoritmo Apriori	68
4.2 FP-Growth	76
4.3 Outras medidas	81
4.4 Notas do capítulo	82
4.5 Exercícios	83
5 Análise via Classificação	85
5.1 Matriz de confusão e métricas de avaliação	87
5.2 Treino, teste e validação cruzada	88
5.3 k-vizinhos mais próximos	90
5.4 Naive Bayes	93
5.5 Árvores de decisão	96
5.6 Floresta aleatória	102
5.7 Boosting	105
5.8 Redes neurais	107

5.9	Hyperparameters	112
5.10	Resumo dos tipos de redes neurais	113
5.11	Notas do capítulo	115
5.12	Exercícios	115
6	Modelos de Regressão	117
6.1	Regressão linear	117
6.2	Correlação e regressão linear com centro fixo	120
6.3	Regressão múltipla ou polinomial	124
6.4	Avaliação de modelos	127
6.5	Exercícios	128
7	Análise de Grupos (<i>Clustering</i>)	129
7.1	Classificação dos agrupamentos	130
7.2	Classificação dos tipos de grupos	131
7.3	K-means	135
7.4	DBSCAN	138
7.5	Agrupamento hierárquico	140
7.6	Exercícios	141
8	Séries Temporais	143
8.1	Decomposição de séries temporais	149
8.2	Representação de dados	151
8.3	Medidas de distância	153
8.4	Mineração e forecasting em séries temporais	157
	Apêndice A – Introdução à linguagem R	165
A.1	Brevíssimo Histórico do R	165
A.2	Instalação do R e R Studio	166
A.3	Usando o R	166
A.4	Tipos primitivos	167
A.5	Vetores	170
A.6	Estruturas de seleção e repetição	172
A.7	Matrizes e Data Frames	172
A.8	Listas e Factors	176
A.9	Valores faltantes e nomes	178
A.10	Funções comuns (geral)	179
A.11	Pacotes	180
	Apêndice B – Métricas para avaliação de modelos	183
	Bibliografia	187
	Índice Remissivo	189