
Brief Contents

Preface	xv
Prologue: A machine learning sampler	1
1 The ingredients of machine learning	13
2 Binary classification and related tasks	49
3 Beyond binary classification	81
4 Concept learning	104
5 Tree models	129
6 Rule models	157
7 Linear models	194
8 Distance-based models	231
9 Probabilistic models	262
10 Features	298
11 Model ensembles	330
12 Machine learning experiments	343
Epilogue: Where to go from here	360
Important points to remember	363
References	367
Index	383

Contents

Preface	xv
Prologue: A machine learning sampler	1
1 The ingredients of machine learning	13
1.1 Tasks: the problems that can be solved with machine learning	14
Looking for structure	16
Evaluating performance on a task	18
1.2 Models: the output of machine learning	20
Geometric models	21
Probabilistic models	25
Logical models	32
Grouping and grading	36
1.3 Features: the workhorses of machine learning	38
Two uses of features	40
Feature construction and transformation	41
Interaction between features	44
1.4 Summary and outlook	46
What you'll find in the rest of the book	48
2 Binary classification and related tasks	49
2.1 Classification	52

Assessing classification performance	53
Visualising classification performance	58
2.2 Scoring and ranking	61
Assessing and visualising ranking performance	63
Turning rankers into classifiers	69
2.3 Class probability estimation	72
Assessing class probability estimates	73
Turning rankers into class probability estimators	76
2.4 Binary classification and related tasks: Summary and further reading	79
3 Beyond binary classification	81
3.1 Handling more than two classes	81
Multi-class classification	82
Multi-class scores and probabilities	86
3.2 Regression	91
3.3 Unsupervised and descriptive learning	95
Predictive and descriptive clustering	96
Other descriptive models	100
3.4 Beyond binary classification: Summary and further reading	102
4 Concept learning	104
4.1 The hypothesis space	106
Least general generalisation	108
Internal disjunction	110
4.2 Paths through the hypothesis space	112
Most general consistent hypotheses	116
Closed concepts	116
4.3 Beyond conjunctive concepts	119
Using first-order logic	122
4.4 Learnability	124
4.5 Concept learning: Summary and further reading	127
5 Tree models	129
5.1 Decision trees	133
5.2 Ranking and probability estimation trees	138
Sensitivity to skewed class distributions	143
5.3 Tree learning as variance reduction	148
Regression trees	148

<i>Contents</i>	xi
Clustering trees	152
5.4 Tree models: Summary and further reading	155
6 Rule models	157
6.1 Learning ordered rule lists	158
Rule lists for ranking and probability estimation	164
6.2 Learning unordered rule sets	167
Rule sets for ranking and probability estimation	173
A closer look at rule overlap	174
6.3 Descriptive rule learning	176
Rule learning for subgroup discovery	178
Association rule mining	182
6.4 First-order rule learning	189
6.5 Rule models: Summary and further reading	192
7 Linear models	194
7.1 The least-squares method	196
Multivariate linear regression	201
Regularised regression	204
Using least-squares regression for classification	205
7.2 The perceptron	207
7.3 Support vector machines	211
Soft margin SVM	216
7.4 Obtaining probabilities from linear classifiers	219
7.5 Going beyond linearity with kernel methods	224
7.6 Linear models: Summary and further reading	228
8 Distance-based models	231
8.1 So many roads...	231
8.2 Neighbours and exemplars	237
8.3 Nearest-neighbour classification	242
8.4 Distance-based clustering	245
K-means algorithm	247
Clustering around medoids	250
Silhouettes	252
8.5 Hierarchical clustering	253
8.6 From kernels to distances	258
8.7 Distance-based models: Summary and further reading	260

9	Probabilistic models	262
9.1	The normal distribution and its geometric interpretations	266
9.2	Probabilistic models for categorical data	273
	Using a naive Bayes model for classification	275
	Training a naive Bayes model	279
9.3	Discriminative learning by optimising conditional likelihood	282
9.4	Probabilistic models with hidden variables	286
	Expectation-Maximisation	288
	Gaussian mixture models	289
9.5	Compression-based models	292
9.6	Probabilistic models: Summary and further reading	295
10	Features	298
10.1	Kinds of feature	299
	Calculations on features	299
	Categorical, ordinal and quantitative features	304
	Structured features	305
10.2	Feature transformations	307
	Thresholding and discretisation	308
	Normalisation and calibration	314
	Incomplete features	321
10.3	Feature construction and selection	322
	Matrix transformations and decompositions	324
10.4	Features: Summary and further reading	327
11	Model ensembles	330
11.1	Bagging and random forests	331
11.2	Boosting	334
	Boosted rule learning	337
11.3	Mapping the ensemble landscape	338
	Bias, variance and margins	338
	Other ensemble methods	339
	Meta-learning	340
11.4	Model ensembles: Summary and further reading	341
12	Machine learning experiments	343
12.1	What to measure	344
12.2	How to measure it	348

<i>Contents</i>	xiii
12.3 How to interpret it	351
Interpretation of results over multiple data sets	354
12.4 Machine learning experiments: Summary and further reading	357
Epilogue: Where to go from here	360
Important points to remember	363
References	367
Index	383