

Contents

Preface — V

List of Tables — XI

List of Figures — XIII

Notation — XVII

Introduction — XIX

1 Fundamentals and definitions — 1

1.1 Goals of pattern recognition — 1

1.2 Structure of a pattern recognition system — 2

1.3 Abstract view of pattern recognition — 4

1.4 Design of a pattern recognition system — 5

1.5 Exercises — 9

2 Features — 10

2.1 Types of features and their traits — 10

2.1.1 Nominal scale — 10

2.1.2 Ordinal scale — 12

2.1.3 Interval scale — 13

2.1.4 Ratio scale and absolute scale — 13

2.2 Feature space inspection — 13

2.2.1 Projections — 14

2.2.2 Intersections and slices — 15

2.3 Transformations of the feature space — 17

2.4 Measurement of distances in the feature space — 17

2.4.1 Basic definitions — 19

2.4.2 Elementary norms and metrics — 20

2.4.3 A metric for sets — 22

2.4.4 Metrics on the ordinal scale — 23

2.4.5 The Kullback–Leibler divergence — 23

2.4.6 Tangential distance measure — 28

2.5 Normalization — 32

2.5.1 Alignment, elimination of physical dimension, and leveling of proportions — 32

2.5.2 Lighting adjustment of images — 33

2.5.3 Distortion adjustment of images — 37

2.5.4	Dynamic time warping —	38
2.6	Selection and construction of features —	39
2.6.1	Descriptive features —	39
2.6.2	Model-driven features —	43
2.6.3	Construction of invariant features —	48
2.7	Dimensionality reduction of the feature space —	56
2.7.1	Principal component analysis —	57
2.7.2	Kernelized principal component analysis —	69
2.7.3	Independent component analysis —	76
2.7.4	Multiple discriminant analysis —	80
2.7.5	Dimensionality reduction by feature selection —	87
2.7.6	Bag of words —	88
2.8	Exercises —	94
3	Bayesian decision theory —	98
3.1	General considerations —	98
3.2	The maximum a posteriori classifier —	101
3.3	Bayesian classification —	104
3.3.1	The Bayesian optimal classifier —	104
3.3.2	Reference example: Optimal decision regions —	109
3.3.3	The minimax classifier —	111
3.3.4	Normally distributed features —	113
3.3.5	Arbitrarily distributed features —	118
3.4	Exercises —	119
4	Parameter estimation —	122
4.1	Maximum likelihood estimation —	130
4.2	Bayesian estimation of the class-specific distributions —	132
4.3	Bayesian parameter estimation —	137
4.3.1	Least squared estimation error —	138
4.3.2	Constant penalty for failures —	138
4.4	Additional remarks on Bayesian classification —	139
4.5	Exercises —	140
5	Parameter free methods —	142
5.1	The Parzen window method —	146
5.2	The k -nearest neighbor method —	150
5.3	k -nearest neighbor classification —	154
5.4	Exercises —	160
6	General considerations —	162
6.1	Dimensionality of the feature space —	162

6.2	Overfitting — 169
6.3	Exercises — 171
7	Special classifiers — 173
7.1	Linear discriminants — 173
7.1.1	More than two classes — 173
7.1.2	Nonlinear separation — 175
7.2	The perceptron — 177
7.3	Linear regression — 179
7.4	Artificial neural networks — 180
7.5	Autoencoders — 184
7.6	Deep learning — 184
7.6.1	Historical difficulties and successful approaches — 185
7.6.2	Unsupervised pre-training — 186
7.6.3	Stochastic gradient descent — 187
7.6.4	Rectified linear units — 188
7.6.5	Convolutional neural networks — 189
7.7	Support vector machines — 194
7.7.1	Linear separation with maximum margin — 194
7.7.2	Dual formulation — 196
7.7.3	Nonlinear mapping — 198
7.7.4	The kernel trick — 199
7.7.5	No linear separability — 203
7.7.6	Discussion — 205
7.8	Matched filters — 206
7.9	Classification of sequences — 210
7.9.1	Markov models — 210
7.9.2	Hidden states — 212
7.10	Exercises — 214
8	Classification with nominal features — 215
8.1	Decision trees — 215
8.1.1	Decision tree learning — 218
8.1.2	Influence of the features used — 222
8.2	Random forests — 222
8.3	String matching — 227
8.4	Grammars — 228
8.5	Exercises — 229
9	Classifier-independent concepts — 231
9.1	Learning theory — 231
9.1.1	The central problem of statistical learning — 232

9.1.2	Vapnik–Chervonenkis learning theory —	232
9.2	Empirical evaluation of classifier performance —	235
9.2.1	Receiver operating characteristic —	238
9.2.2	Multi-class setting —	238
9.2.3	Theoretical bounds with finite test sets —	240
9.2.4	Dealing with small datasets —	241
9.3	Boosting —	242
9.4	Rejection —	244
9.5	Exercises —	247

A Solutions to the exercises — 249

A.1	Chapter 1 —	249
A.2	Chapter 2 —	250
A.3	Chapter 3 —	252
A.4	Chapter 4 —	255
A.5	Chapter 5 —	257
A.6	Chapter 6 —	259
A.7	Chapter 7 —	260
A.8	Chapter 8 —	261
A.9	Chapter 9 —	262

B A primer on Lie theory — 263

C Random processes — 268

Bibliography — 271

Glossary — 275

Index — 281