

Contents

Preface — V

Preface of 2nd edition — VII

List of Tables — XIII

List of Figures — XV

Notation — XIX

Introduction — XXIII

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 — 12
 - 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 — 21
 - 2.4.4 Metrics on the ordinal scale — 23
 - 2.4.5 The cosine distance — 23
 - 2.4.6 The Kullback–Leibler divergence — 24
 - 2.4.7 The t -distributed stochastic neighbor embedding — 29
 - 2.4.8 Tangential distance measure — 31

2.5	Normalization — 34
2.5.1	Alignment, elimination of physical dimension, and leveling of proportions — 35
2.5.2	Lighting adjustment of images — 35
2.5.3	Distortion adjustment of images — 39
2.5.4	Dynamic time warping — 40
2.6	Selection and construction of features — 41
2.6.1	Descriptive features — 41
2.6.2	Model-driven features — 45
2.6.3	Construction of invariant features — 50
2.7	Dimensionality reduction of the feature space — 59
2.7.1	Principal component analysis — 60
2.7.2	Kernelized principal component analysis — 73
2.7.3	Independent component analysis — 79
2.7.4	Multiple discriminant analysis — 83
2.7.5	Dimensionality reduction with t -SNE — 90
2.7.6	Autoencoder — 91
2.7.7	Dimensionality reduction by feature selection — 92
2.7.8	Bag of words — 94
2.8	Exercises — 99

3 Bayesian decision theory — 103

3.1	General considerations — 103
3.2	The maximum a posteriori classifier — 106
3.3	Bayesian classification — 109
3.3.1	The Bayesian optimal classifier — 109
3.3.2	Reference example: Optimal decision regions — 114
3.3.3	The naive Bayes classifier — 116
3.3.4	The minimax classifier — 118
3.3.5	Normally distributed features — 120
3.3.6	Arbitrarily distributed features — 125
3.4	Gaussian mixtures — 126
3.5	Exercises — 129

4 Parameter estimation — 132

4.1	Maximum likelihood estimation — 140
4.2	Bayesian estimation of the class-specific distributions — 142
4.3	Bayesian parameter estimation — 147
4.3.1	Least squared estimation error — 148
4.3.2	Constant penalty for failures — 148
4.4	Additional remarks on Bayesian classification — 149

4.5	Parameter estimation for Gaussian mixtures —	150
4.6	Exercises —	159
5	Parameter free methods —	161
5.1	The Parzen window method —	165
5.2	The k -nearest neighbor method —	169
5.3	k -nearest neighbor classification —	173
5.4	Exercises —	179
6	General considerations —	180
6.1	Dimensionality of the feature space —	180
6.2	Overfitting —	187
6.3	Exercises —	190
7	Special classifiers —	191
7.1	Linear discriminants —	191
7.1.1	More than two classes —	191
7.1.2	Nonlinear separation —	193
7.2	The perceptron —	195
7.3	Linear regression —	197
7.4	Artificial neural networks —	198
7.5	Autoencoders —	204
7.6	Deep learning —	206
7.6.1	Historical difficulties and successful approaches —	207
7.6.2	Unsupervised pre-training —	208
7.6.3	Stochastic gradient descent —	208
7.6.4	Rectified linear units —	210
7.6.5	Convolutional neural networks —	210
7.6.6	Residual networks —	215
7.6.7	Variational autoencoders —	216
7.6.8	Mixture density networks —	217
7.7	Support vector machines —	219
7.7.1	Linear separation with maximum margin —	219
7.7.2	Dual formulation —	221
7.7.3	Nonlinear mapping —	223
7.7.4	The kernel trick —	224
7.7.5	No linear separability —	228
7.7.6	Discussion —	230
7.8	Matched filters —	231
7.9	Classification of sequences —	234
7.9.1	Markov models —	235

- 7.9.2 Hidden Markov models — **236**
- 7.9.3 Recurrent neural networks — **246**
- 7.9.4 Transformers — **249**
- 7.9.5 Generative adversarial networks — **252**
- 7.10 Exercises — **255**

8 Classification with nominal features — 257

- 8.1 Decision trees — **257**
- 8.1.1 Decision tree learning — **260**
- 8.1.2 Influence of the features used — **264**
- 8.2 Random forests — **265**
- 8.3 String matching — **269**
- 8.4 Grammars — **270**
- 8.5 Exercises — **271**

9 Classifier-independent concepts — 273

- 9.1 Learning theory — **273**
- 9.1.1 The central problem of statistical learning — **274**
- 9.1.2 Vapnik–Chervonenkis learning theory — **274**
- 9.2 No-free-lunch theorem — **277**
- 9.3 Empirical evaluation of classifier performance — **278**
- 9.3.1 Receiver operating characteristic — **281**
- 9.3.2 Multi-class setting — **282**
- 9.3.3 Theoretical bounds with finite test sets — **283**
- 9.3.4 Dealing with small datasets — **284**
- 9.4 Boosting — **286**
- 9.5 Rejection — **288**
- 9.6 Exercises — **291**

A Solutions to the exercises — 293

B A primer on Lie theory — 307

C Random processes — 311

Bibliography — 315

Glossary — 319

Index — 325