Contents

List of code fragments *page* viii
Preface xi

Part I Basic concepts 1

1 Pattern analysis 3
1.1 Patterns in data 4
1.2 Pattern analysis algorithms 12
1.3 Exploiting patterns 17
1.4 Summary 22
1.5 Further reading and advanced topics 23

2 Kernel methods: an overview 25
2.1 The overall picture 26
2.2 Linear regression in a feature space 27
2.3 Other examples 36
2.4 The modularity of kernel methods 42
2.5 Roadmap of the book 43
2.6 Summary 44
2.7 Further reading and advanced topics 45

3 Properties of kernels 47
3.1 Inner products and positive semi-definite matrices 48
3.2 Characterisation of kernels 60
3.3 The kernel matrix 68
3.4 Kernel construction 74
3.5 Summary 82
3.6 Further reading and advanced topics 82

4 Detecting stable patterns 85
4.1 Concentration inequalities 86
4.2 Capacity and regularisation: Rademacher theory 93

4.3 Pattern stability for kernel-based classes 97
4.4 A pragmatic approach 104
4.5 Summary 105
4.6 Further reading and advanced topics 106

Part II Pattern analysis algorithms 109

5 Elementary algorithms in feature space 111
5.1 Means and distances 112
5.2 Computing projections: Gram–Schmidt, QR and Cholesky 122
5.3 Measuring the spread of the data 128
5.4 Fisher discriminant analysis I 132
5.5 Summary 137
5.6 Further reading and advanced topics 138

6 Pattern analysis using eigen-decompositions 140
6.1 Singular value decomposition 141
6.2 Principal components analysis 143
6.3 Directions of maximum covariance 155
6.4 The generalised eigenvector problem 161
6.5 Canonical correlation analysis 164
6.6 Fisher discriminant analysis II 176
6.7 Methods for linear regression 176
6.8 Summary 192
6.9 Further reading and advanced topics 193

7 Pattern analysis using convex optimisation 195
7.1 The smallest enclosing hypersphere 196
7.2 Support vector machines for classification 211
7.3 Support vector machines for regression 230
7.4 On-line classification and regression 241
7.5 Summary 249
7.6 Further reading and advanced topics 250

8 Ranking, clustering and data visualisation 252
8.1 Discovering rank relations 253
8.2 Discovering cluster structure in a feature space 264
8.3 Data visualisation 280
8.4 Summary 286
8.5 Further reading and advanced topics 286

Part III Constructing kernels 289

9 Basic kernels and kernel types 291
9.1 Kernels in closed form 292

9.2 ANOVA kernels 297
9.3 Kernels from graphs 304
9.4 Diffusion kernels on graph nodes 310
9.5 Kernels on sets 314
9.6 Kernels on real numbers 318
9.7 Randomised kernels 320
9.8 Other kernel types 322
9.9 Summary 324
9.10 Further reading and advanced topics 325

10 Kernels for text 327
10.1 From bag of words to semantic space 328
10.2 Vector space kernels 331
10.3 Summary 341
10.4 Further reading and advanced topics 342

11 Kernels for structured data: strings, trees, etc. 344
11.1 Comparing strings and sequences 345
11.2 Spectrum kernels 347
11.3 All-subsequences kernels 351
11.4 Fixed length subsequences kernels 357
11.5 Gap-weighted subsequences kernels 360
11.6 Beyond dynamic programming: trie-based kernels 372
11.7 Kernels for structured data 382
11.8 Summary 395
11.9 Further reading and advanced topics 395

12 Kernels from generative models 397
12.1 *P*-kernels 398
12.2 Fisher kernels 421
12.3 Summary 435
12.4 Further reading and advanced topics 436

***Appendix A* Proofs omitted from the main text** 437

***Appendix B* Notational conventions** 444

***Appendix C* List of pattern analysis methods** 446

***Appendix D* List of kernels** 448

References 450

Index 460