Contents

Acknowledgements		xi
1	**Introduction**	1
1.1	A (very) brief history of statistics	1
1.2	Twentieth-century versus twenty-first-century statistics	2
1.3	Modern statistical theory	4
1.4	A perspective on data, models, methods and theory	14
1.5	Overview of the book	15
1.6	Notation	18
1.7	Bibliographic details and background	19
2	**Linear models and ordinary least squares**	21
2.1	Basic properties of OLS	23
2.2	Gaussian errors	24
2.3	Inference	26
2.4	Model misspecification	34
2.5	Multiple testing	38
2.6	Exercises	42
2.7	Bibliographic details and background	47
3	**High-dimensional linear regression**	50
3.1	Ridge regression	51
3.2	Standardisation of predictors	54
3.3	Cross-validation	56
3.4	Model selection	58
3.5	The Lasso estimator	60
3.6	Residual prediction tests with the Lasso	82
3.7	Extensions of the Lasso	84
3.8	Inference	92
3.9	Exercises	99
3.10	Bibliographic details and background	103
4	**Kernel density estimation**	106
4.1	Kernel density estimators	107
4.2	Mean squared error of kernel estimators	108
4.3	Bounds on the pointwise variance and bias	110
4.4	Bounds on the integrated variance and squared bias	114

Published online by Cambridge University Press

4.5 Bandwidth selection 116
4.6 Choice of kernel 122
4.7 Exercises 123
4.8 Bibliographic details and background 127

5 Nonparametric regression 129
5.1 Fixed and random design 130
5.2 Local polynomial estimators 130
5.3 Splines 139
5.4 Exercises 143
5.5 Bibliographic details and background 147

6 Reproducing kernel Hilbert spaces and kernel machines 148
6.1 The kernel trick 148
6.2 Kernels 150
6.3 Reproducing kernel Hilbert spaces 156
6.4 The representer theorem 162
6.5 Support vector machines 163
6.6 Mercer's theorem 166
6.7 Prediction error of kernel ridge regression 169
6.8 Additive models 176
6.9 Large-scale kernel machines 178
6.10 Exercises 179
6.11 Bibliographic details and background 182

7 Conditional independence, graphical models and causality 184
7.1 Conditional independence 184
7.2 Graphs 188
7.3 Undirected graphical models 190
7.4 Directed graphical models and causality 191
7.5 Causal structure learning 199
7.6 Gaussian graphical models 207
7.7 The generalised covariance measure 211
7.8 Exercises 216
7.9 Bibliographic details and background 219

8 Minimax lower bounds 221
8.1 Reduction to testing 221
8.2 f-divergences 224
8.3 Le Cam's two-point lemma 225
8.4 Assouad's lemma 229
8.5 The data processing inequality 233
8.6 Fano's lemma 234
8.7 High-probability minimax lower bounds 240
8.8 Exercises 246
8.9 Bibliographic details and background 252

Published online by Cambridge University Press

9 Shape-constrained estimation 255
9.1 Decreasing densities and the Grenander projection 255
9.2 Log-concave density estimation 262
9.3 Linear regression via optimal convex M-estimation 287
9.4 Shape-constrained regression problems 291
9.5 Exercises 306
9.6 Bibliographic details and background 310

Appendix: Mathematical background 315
A.1 Metric and topological spaces 315
A.2 Measure theory 323
A.3 Convex analysis 334
A.4 Characteristic functions 343
A.5 The multivariate normal distribution 348
A.6 Concentration inequalities 351
A.7 Taylor's theorem and absolute continuity of functions 362
A.8 Functional analysis 364
A.9 Conditional expectation and martingales 380
A.10 Matrix theory 390
A.11 Stochastic convergence 400
A.12 Bibliographic details and background 408

References 412

Index 435

Published online by Cambridge University Press

Published online by Cambridge University Press