
Contents

Preface	xi
1 Introduction	1
1.1 Computational approach	1
1.2 Statistical learning	2
1.3 Example	3
1.4 Prerequisites	5
1.5 How to read this book	6
1.6 Supplementary materials	7
1.7 Formalisms and terminology	7
1.8 Exercises	9
2 Linear Models	11
2.1 Introduction	11
2.2 Ordinary least squares	13
2.3 The normal equations	15
2.4 Solving least squares with the singular value decomposition .	17
2.5 Directly solving the linear system	19
2.6 (★) Solving linear models using the QR decomposition	22
2.7 (★) Sensitivity analysis	24
2.8 (★) Relationship between numerical and statistical error . . .	28
2.9 Implementation and notes	31
2.10 Application: Cancer incidence rates	32
2.11 Exercises	40
3 Ridge Regression and Principal Component Analysis	43
3.1 Variance in OLS	43
3.2 Ridge regression	46
3.3 (★) A Bayesian perspective	53
3.4 Principal component analysis	56
3.5 Implementation and notes	63
3.6 Application: NYC taxicab data	65
3.7 Exercises	72

4	Linear Smoothers	75
4.1	Non-Linearity	75
4.2	Basis expansion	76
4.3	Kernel regression	81
4.4	Local regression	85
4.5	Regression splines	89
4.6	(★) Smoothing splines	95
4.7	(★) B-splines	100
4.8	Implementation and notes	104
4.9	Application: U.S. census tract data	105
4.10	Exercises	120
5	Generalized Linear Models	123
5.1	Classification with linear models	123
5.2	Exponential families	128
5.3	Iteratively reweighted GLMs	131
5.4	(★) Numerical issues	135
5.5	(★) Multi-Class regression	138
5.6	Implementation and notes	139
5.7	Application: Chicago crime prediction	140
5.8	Exercises	148
6	Additive Models	151
6.1	Multivariate linear smoothers	151
6.2	Curse of dimensionality	155
6.3	Additive models	158
6.4	(★) Additive models as linear models	163
6.5	(★) Standard errors in additive models	166
6.6	Implementation and notes	170
6.7	Application: NYC flights data	172
6.8	Exercises	178
7	Penalized Regression Models	179
7.1	Variable selection	179
7.2	Penalized regression with the ℓ_0 - and ℓ_1 -norms	180
7.3	Orthogonal data matrix	182
7.4	Convex optimization and the elastic net	186
7.5	Coordinate descent	188
7.6	(★) Active set screening using the KKT conditions	193
7.7	(★) The generalized elastic net model	198
7.8	Implementation and notes	200
7.9	Application: Amazon product reviews	201
7.10	Exercises	206

8	Neural Networks	207
8.1	Dense neural network architecture	207
8.2	Stochastic gradient descent	211
8.3	Backward propagation of errors	213
8.4	Implementing backpropagation	216
8.5	Recognizing handwritten digits	224
8.6	(★) Improving SGD and regularization	226
8.7	(★) Classification with neural networks	232
8.8	(★) Convolutional neural networks	239
8.9	Implementation and notes	249
8.10	Application: Image classification with EMNIST	249
8.11	Exercises	259
9	Dimensionality Reduction	261
9.1	Unsupervised learning	261
9.2	Kernel functions	262
9.3	Kernel principal component analysis	266
9.4	Spectral clustering	272
9.5	t-Distributed stochastic neighbor embedding (t-SNE)	277
9.6	Autoencoders	282
9.7	Implementation and notes	283
9.8	Application: Classifying and visualizing fashion MNIST	284
9.9	Exercises	295
10	Computation in Practice	297
10.1	Reference implementations	297
10.2	Sparse matrices	298
10.3	Sparse generalized linear models	304
10.4	Computation on row chunks	307
10.5	Feature hashing	311
10.6	Data quality issues	318
10.7	Implementation and notes	320
10.8	Application	321
10.9	Exercises	329
A	Linear algebra and matrices	331
A.1	Vector spaces	331
A.2	Matrices	333
B	Floating Point Arithmetic and Numerical Computation	337
B.1	Floating point arithmetic	337
B.2	Computational effort	340
	Bibliography	343
	Index	359