

CONTENTS

Preface	xiii
Notation	xvii
1 Importing, Summarizing, and Visualizing Data	1
1.1 Introduction	1
1.2 Structuring Features According to Type	3
1.3 Summary Tables	6
1.4 Summary Statistics	7
1.5 Visualizing Data	8
1.5.1 Plotting Qualitative Variables	9
1.5.2 Plotting Quantitative Variables	9
1.5.3 Data Visualization in a Bivariate Setting	12
Exercises	15
2 Statistical Learning	19
2.1 Introduction	19
2.2 Supervised and Unsupervised Learning	20
2.3 Training and Test Loss	23
2.4 Tradeoffs in Statistical Learning	31
2.5 Estimating Risk	35
2.5.1 In-Sample Risk	35
2.5.2 Cross-Validation	37
2.6 Modeling Data	40
2.7 Multivariate Normal Models	44
2.8 Normal Linear Models	46
2.9 Bayesian Learning	47
Exercises	58
3 Monte Carlo Methods	67
3.1 Introduction	67
3.2 Monte Carlo Sampling	68
3.2.1 Generating Random Numbers	68
3.2.2 Simulating Random Variables	69
3.2.3 Simulating Random Vectors and Processes	74
3.2.4 Resampling	76
3.2.5 Markov Chain Monte Carlo	78
3.3 Monte Carlo Estimation	85

3.3.1	Crude Monte Carlo	85
3.3.2	Bootstrap Method	88
3.3.3	Variance Reduction	92
3.4	Monte Carlo for Optimization	96
3.4.1	Simulated Annealing	96
3.4.2	Cross-Entropy Method	100
3.4.3	Splitting for Optimization	103
3.4.4	Noisy Optimization	105
	Exercises	113
4	Unsupervised Learning	121
4.1	Introduction	121
4.2	Risk and Loss in Unsupervised Learning	122
4.3	Expectation–Maximization (EM) Algorithm	128
4.4	Empirical Distribution and Density Estimation	131
4.5	Clustering via Mixture Models	135
4.5.1	Mixture Models	135
4.5.2	EM Algorithm for Mixture Models	137
4.6	Clustering via Vector Quantization	142
4.6.1	K-Means	144
4.6.2	Clustering via Continuous Multiextremal Optimization	146
4.7	Hierarchical Clustering	147
4.8	Principal Component Analysis (PCA)	153
4.8.1	Motivation: Principal Axes of an Ellipsoid	153
4.8.2	PCA and Singular Value Decomposition (SVD)	155
	Exercises	160
5	Regression	167
5.1	Introduction	167
5.2	Linear Regression	169
5.3	Analysis via Linear Models	171
5.3.1	Parameter Estimation	171
5.3.2	Model Selection and Prediction	172
5.3.3	Cross-Validation and Predictive Residual Sum of Squares	173
5.3.4	In-Sample Risk and Akaike Information Criterion	175
5.3.5	Categorical Features	177
5.3.6	Nested Models	180
5.3.7	Coefficient of Determination	181
5.4	Inference for Normal Linear Models	182
5.4.1	Comparing Two Normal Linear Models	183
5.4.2	Confidence and Prediction Intervals	186
5.5	Nonlinear Regression Models	188
5.6	Linear Models in Python	191
5.6.1	Modeling	191
5.6.2	Analysis	193
5.6.3	Analysis of Variance (ANOVA)	195

5.6.4	Confidence and Prediction Intervals	198
5.6.5	Model Validation	198
5.6.6	Variable Selection	199
5.7	Generalized Linear Models	204
	Exercises	207
6	Regularization and Kernel Methods	215
6.1	Introduction	215
6.2	Regularization	216
6.3	Reproducing Kernel Hilbert Spaces	222
6.4	Construction of Reproducing Kernels	225
6.4.1	Reproducing Kernels via Feature Mapping	225
6.4.2	Kernels from Characteristic Functions	225
6.4.3	Reproducing Kernels Using Orthonormal Features	227
6.4.4	Kernels from Kernels	229
6.5	Representer Theorem	231
6.6	Smoothing Cubic Splines	235
6.7	Gaussian Process Regression	239
6.8	Kernel PCA	243
	Exercises	246
7	Classification	253
7.1	Introduction	253
7.2	Classification Metrics	255
7.3	Classification via Bayes' Rule	259
7.4	Linear and Quadratic Discriminant Analysis	261
7.5	Logistic Regression and Softmax Classification	268
7.6	K -Nearest Neighbors Classification	270
7.7	Support Vector Machine	271
7.8	Classification with Scikit-Learn	279
	Exercises	281
8	Decision Trees and Ensemble Methods	289
8.1	Introduction	289
8.2	Top-Down Construction of Decision Trees	291
8.2.1	Regional Prediction Functions	292
8.2.2	Splitting Rules	293
8.2.3	Termination Criterion	294
8.2.4	Basic Implementation	296
8.3	Additional Considerations	300
8.3.1	Binary Versus Non-Binary Trees	300
8.3.2	Data Preprocessing	300
8.3.3	Alternative Splitting Rules	300
8.3.4	Categorical Variables	301
8.3.5	Missing Values	301
8.4	Controlling the Tree Shape	302
8.4.1	Cost-Complexity Pruning	305

8.4.2	Advantages and Limitations of Decision Trees	306
8.5	Bootstrap Aggregation	307
8.6	Random Forests	311
8.7	Boosting	315
	Exercises	323
9	Deep Learning	325
9.1	Introduction	325
9.2	Feed-Forward Neural Networks	328
9.3	Back-Propagation	332
9.4	Methods for Training	336
9.4.1	Steepest Descent	336
9.4.2	Levenberg–Marquardt Method	337
9.4.3	Limited-Memory BFGS Method	338
9.4.4	Adaptive Gradient Methods	340
9.5	Examples in Python	342
9.5.1	Simple Polynomial Regression	342
9.5.2	Image Classification	346
	Exercises	350
A	Linear Algebra and Functional Analysis	357
A.1	Vector Spaces, Bases, and Matrices	357
A.2	Inner Product	362
A.3	Complex Vectors and Matrices	363
A.4	Orthogonal Projections	364
A.5	Eigenvalues and Eigenvectors	365
A.5.1	Left- and Right-Eigenvectors	366
A.6	Matrix Decompositions	370
A.6.1	(P)LU Decomposition	370
A.6.2	Woodbury Identity	372
A.6.3	Cholesky Decomposition	375
A.6.4	QR Decomposition and the Gram–Schmidt Procedure	377
A.6.5	Singular Value Decomposition	378
A.6.6	Solving Structured Matrix Equations	381
A.7	Functional Analysis	386
A.8	Fourier Transforms	392
A.8.1	Discrete Fourier Transform	394
A.8.2	Fast Fourier Transform	396
B	Multivariate Differentiation and Optimization	399
B.1	Multivariate Differentiation	399
B.1.1	Taylor Expansion	402
B.1.2	Chain Rule	402
B.2	Optimization Theory	404
B.2.1	Convexity and Optimization	405
B.2.2	Lagrangian Method	408
B.2.3	Duality	409

B.3	Numerical Root-Finding and Minimization	410
B.3.1	Newton-Like Methods	411
B.3.2	Quasi-Newton Methods	413
B.3.3	Normal Approximation Method	415
B.3.4	Nonlinear Least Squares	416
B.4	Constrained Minimization via Penalty Functions	417
C	Probability and Statistics	423
C.1	Random Experiments and Probability Spaces	423
C.2	Random Variables and Probability Distributions	424
C.3	Expectation	428
C.4	Joint Distributions	429
C.5	Conditioning and Independence	430
C.5.1	Conditional Probability	430
C.5.2	Independence	430
C.5.3	Expectation and Covariance	431
C.5.4	Conditional Density and Conditional Expectation	433
C.6	Functions of Random Variables	433
C.7	Multivariate Normal Distribution	436
C.8	Convergence of Random Variables	441
C.9	Law of Large Numbers and Central Limit Theorem	447
C.10	Markov Chains	453
C.11	Statistics	455
C.12	Estimation	456
C.12.1	Method of Moments	457
C.12.2	Maximum Likelihood Method	458
C.13	Confidence Intervals	459
C.14	Hypothesis Testing	460
D	Python Primer	465
D.1	Getting Started	465
D.2	Python Objects	467
D.3	Types and Operators	468
D.4	Functions and Methods	470
D.5	Modules	471
D.6	Flow Control	473
D.7	Iteration	474
D.8	Classes	475
D.9	Files	477
D.10	NumPy	480
D.10.1	Creating and Shaping Arrays	480
D.10.2	Slicing	482
D.10.3	Array Operations	482
D.10.4	Random Numbers	484
D.11	Matplotlib	485
D.11.1	Creating a Basic Plot	485

D.12 Pandas	487
D.12.1 Series and DataFrame	487
D.12.2 Manipulating Data Frames	489
D.12.3 Extracting Information	490
D.12.4 Plotting	492
D.13 Scikit-learn	492
D.13.1 Partitioning the Data	493
D.13.2 Standardization	493
D.13.3 Fitting and Prediction	494
D.13.4 Testing the Model	494
D.14 System Calls, URL Access, and Speed-Up	495
Bibliography	497
Index	505