

CONTENTS

Preface	ix
Notation	xv
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	32
2.5 Estimating Risk	36
2.5.1 In-Sample Risk	36
2.5.2 Cross-Validation	39
2.6 Modeling Data	42
2.7 Multivariate Normal Models	46
2.8 Normal Linear Models	48
2.9 Bayesian Learning	49
Exercises	61
3 Monte Carlo Methods	73
3.1 Introduction	73
3.2 Monte Carlo Sampling	74
3.2.1 Generating Random Numbers	74
3.2.2 Simulating Random Variables	75
3.2.3 Simulating Random Vectors and Processes	85
3.2.4 Resampling	86
3.2.5 Markov Chain Monte Carlo	88

3.3	Monte Carlo Estimation	96
3.3.1	Crude Monte Carlo	96
3.3.2	Bootstrap Method	99
3.3.3	Variance Reduction	103
3.4	Monte Carlo for Optimization	107
3.4.1	Simulated Annealing	107
3.4.2	Cross-Entropy Method	111
3.4.3	Splitting for Optimization	115
3.4.4	Noisy Optimization	118
	Exercises	126
4	Unsupervised Learning	133
4.1	Introduction	133
4.2	Risk and Loss in Unsupervised Learning	134
4.3	Expectation–Maximization (EM) Algorithm	140
4.4	Gaussian Kernel Density Estimation	144
4.4.1	Fast Computation of the Gaussian KDE	145
4.4.2	Optimal Gaussian Kernel Bandwidth Selection	147
4.5	Theta Kernel Density Estimation	153
4.5.1	Fast Computation of the Theta KDE	155
4.5.2	Optimal Theta Kernel Bandwidth Selection	156
4.6	Clustering via Mixture Models	160
4.6.1	Mixture Models	161
4.6.2	EM Algorithm for Mixture Models	163
4.7	Clustering via Vector Quantization	167
4.7.1	K-Means	169
4.7.2	Clustering via Continuous Multiextremal Optimization	171
4.8	Hierarchical Clustering	173
4.9	Principal Component Analysis (PCA)	179
4.9.1	Motivation: Principal Axes of an Ellipsoid	179
4.9.2	PCA and Singular Value Decomposition (SVD)	181
	Exercises	185
5	Regression	197
5.1	Introduction	197
5.2	Linear Regression	199
5.3	Analysis via Linear Models	201
5.3.1	Parameter Estimation	201
5.3.2	Model Selection and Prediction	202
5.3.3	Cross-Validation	203
5.3.4	In-Sample Risk and Akaike Information Criterion	206
5.3.5	Categorical Features	208
5.3.6	Nested Models	211
5.3.7	Coefficient of Determination	212
5.4	Inference for Normal Linear Models	213
5.4.1	Comparing Two Normal Linear Models	214

5.4.2	Confidence and Prediction Intervals	217
5.5	Nonlinear Regression Models	219
5.6	Gaussian-Weighted Local Linear Regression	222
5.6.1	Computation of the Learner $g_\tau(\cdot \lambda)$	222
5.6.2	Selecting λ via Bandwidth Perturbation Matching	224
5.7	Linear Models in Python	230
5.7.1	Modeling	230
5.7.2	Analysis	232
5.7.3	Analysis of Variance (ANOVA)	234
5.7.4	Confidence and Prediction Intervals	237
5.7.5	Model Validation	237
5.7.6	Variable Selection	238
5.8	Generalized Linear Models	242
	Exercises	246
6	Feature Selection and Shrinkage	261
6.1	Introduction to the Klimax Method	261
6.1.1	Examples of the Klimax Method	263
6.1.2	Special Cases of the Klimax Estimator	266
6.1.3	Eliminating the Intercept	268
6.2	Coordinate-Descent for the Klimax Method	270
6.2.1	Klimax Estimate	271
6.2.2	(Positive-Part) James–Stein Estimate	277
6.2.3	Lasso Estimate	280
6.3	Statistical Properties of Estimators	285
6.4	Practical Selection of Regularization Parameter	292
6.4.1	Monte Carlo Estimator of σ	292
6.4.2	Plug-In Estimator of σ For Large p	296
	Exercises	300
7	Reproducing Kernel Methods	319
7.1	Introduction	319
7.2	L^2 -Regularization	320
7.3	Reproducing Kernel Hilbert Spaces	323
7.4	Construction of Reproducing Kernels	325
7.4.1	Reproducing Kernels via Feature Mapping	326
7.4.2	Kernels from Characteristic Functions	327
7.4.3	Reproducing Kernels Using Orthonormal Features	328
7.4.4	Kernels from Kernels	331
7.5	Representer Theorem	332
7.6	Smoothing Cubic Splines	337
7.7	Gaussian Process Regression	341
7.8	Kernel PCA	345
	Exercises	348

8	Classification	353
8.1	Introduction	353
8.2	Classification Metrics	355
8.3	Classification via Bayes' Rule	360
8.4	Linear and Quadratic Discriminant Analysis	362
8.5	Logistic Regression and Softmax Classification	369
8.6	<i>K</i> -Nearest Neighbors Classification	371
8.7	Support Vector Machine	373
8.8	Classification with Scikit-Learn	380
	Exercises	383
9	Decision Trees and Ensemble Methods	391
9.1	Introduction	391
9.2	Top-Down Construction of Decision Trees	393
9.2.1	Regional Prediction Functions	394
9.2.2	Splitting Rules	395
9.2.3	Termination Criterion	397
9.2.4	Basic Implementation	399
9.3	Additional Considerations	402
9.3.1	Binary Versus Non-Binary Trees	402
9.3.2	Data Preprocessing	402
9.3.3	Alternative Splitting Rules	402
9.3.4	Categorical Variables	403
9.3.5	Missing Values	403
9.4	Controlling the Tree Shape	404
9.4.1	Cost-Complexity Pruning	407
9.4.2	Advantages and Limitations of Decision Trees	408
9.5	Bootstrap Aggregation	409
9.6	Random Forests	413
9.7	Boosting	417
	Exercises	426
10	Deep Learning	429
10.1	Introduction	429
10.2	Feed-Forward Neural Networks	432
10.3	Back-Propagation	436
10.4	Methods for Training	440
10.4.1	Gradient Descent	440
10.4.2	Levenberg–Marquardt Method	442
10.4.3	Limited-Memory BFGS Method	442
10.4.4	Adaptive Gradient Methods	445
10.5	Examples in Python	447
10.5.1	Simple Polynomial Regression	447
10.5.2	Image Classification	450
	Exercises	455

11 Reinforcement Learning	461
11.1 Introduction	461
11.2 Markov Decision Processes	462
11.3 Value and Policy Iteration	466
11.3.1 Value Iteration	466
11.3.2 Policy Iteration	471
11.4 Reinforcement Learning Concepts	473
11.5 Dynamic Programming	474
11.6 Temporal Difference Learning	478
11.7 Monte Carlo Methods	484
11.7.1 Estimating the Value Function via Monte Carlo	484
11.7.2 Estimating the Q-Function via Monte Carlo	488
11.7.3 Monte Carlo Tree Search	490
11.8 Parametric Policy Improvement	499
Exercises	508
A Linear Algebra	513
A.1 Vector Spaces, Bases, and Matrices	513
A.2 Inner Product	517
A.3 Kronecker Product	518
A.4 Complex Vectors and Matrices	520
A.5 Orthogonal Projections	521
A.6 The Gram–Schmidt Procedure and QR Decomposition	522
A.7 Eigenvalues and Eigenvectors	526
A.7.1 Left- and Right-Eigenvectors	527
A.7.2 Eigen-Decomposition and Hermitian Matrices	529
A.7.3 Singular Value Decomposition	532
A.8 Solving Linear Equations	537
A.8.1 (P)LU Decompositions	537
A.8.2 Block LU Decompositions	540
A.8.3 Cholesky and LDL Decompositions	543
A.8.4 Levinson–Durbin Decomposition	549
A.8.5 Hessenberg–Arnoldi and Lanczos Decompositions	554
B Functional Analysis	563
B.1 Introduction	563
B.2 Fourier Transforms	570
B.2.1 Discrete Fourier Transform	573
B.2.2 Fast Fourier Transform	576
B.2.3 Discrete Cosine Transform	578
C Multivariate Differentiation and Optimization	581
C.1 Multivariate Differentiation	581
C.1.1 Chain Rule	584
C.1.2 Taylor Expansion	586
C.2 Optimization Theory	589

C.2.1	Convexity and Optimization	590
C.2.2	Lagrangian Method	594
C.2.3	Duality	597
C.3	Numerical Root-Finding and Minimization	599
C.3.1	Newton-Like Methods	600
C.3.2	Quasi-Newton Methods	601
C.3.3	Normal Approximation Method	603
C.3.4	Nonlinear Least Squares	605
C.3.5	Coordinate Descent Method	606
C.4	Constrained Numerical Minimization	608
C.4.1	The Majorize-Minimize Method	609
C.4.2	Alternating Direction Method of Multipliers	618
C.4.3	Barrier Function Method	621
D	Probability and Statistics	625
D.1	Random Experiments and Probability Spaces	625
D.2	Random Variables and Probability Distributions	626
D.3	Expectation	629
D.4	Joint Distributions	631
D.5	Conditioning and Independence	631
D.5.1	Conditional Probability	632
D.5.2	Independence	632
D.5.3	Expectation and Covariance	633
D.5.4	Conditional Density and Conditional Expectation	634
D.6	Functions of Random Variables	635
D.7	Multivariate Normal Distribution	638
D.8	Convergence of Random Variables	645
D.9	Law of Large Numbers and Central Limit Theorem	653
D.10	Markov Chains	660
D.11	Statistics	662
D.12	Estimation	663
D.12.1	Method of Moments	664
D.12.2	Maximum Likelihood Method	665
D.13	Confidence Intervals	666
D.14	Hypothesis Testing	667
E	Python Primer	671
E.1	Getting Started	671
E.2	Python Objects	673
E.3	Types and Operators	675
E.4	Functions and Methods	676
E.5	Modules	678
E.6	Flow Control	679
E.7	Iteration	680
E.8	Classes	682
E.9	Files	684

E.10 NumPy	686
E.10.1 Creating and Shaping Arrays	687
E.10.2 Slicing	688
E.10.3 Array Operations	689
E.10.4 Random Numbers	691
E.11 Matplotlib	691
E.11.1 Creating a Basic Plot	691
E.12 Pandas	694
E.12.1 Series and DataFrame	694
E.12.2 Manipulating Data Frames	695
E.12.3 Extracting Information	696
E.12.4 Plotting	698
E.13 Scikit-learn	699
E.13.1 Partitioning the Data	699
E.13.2 Standardization	701
E.13.3 Fitting and Prediction	701
E.13.4 Testing the Model	701
E.14 System Calls, URL Access, and Speed-Up	702

Bibliography	704
---------------------	------------

Index	717
--------------	------------