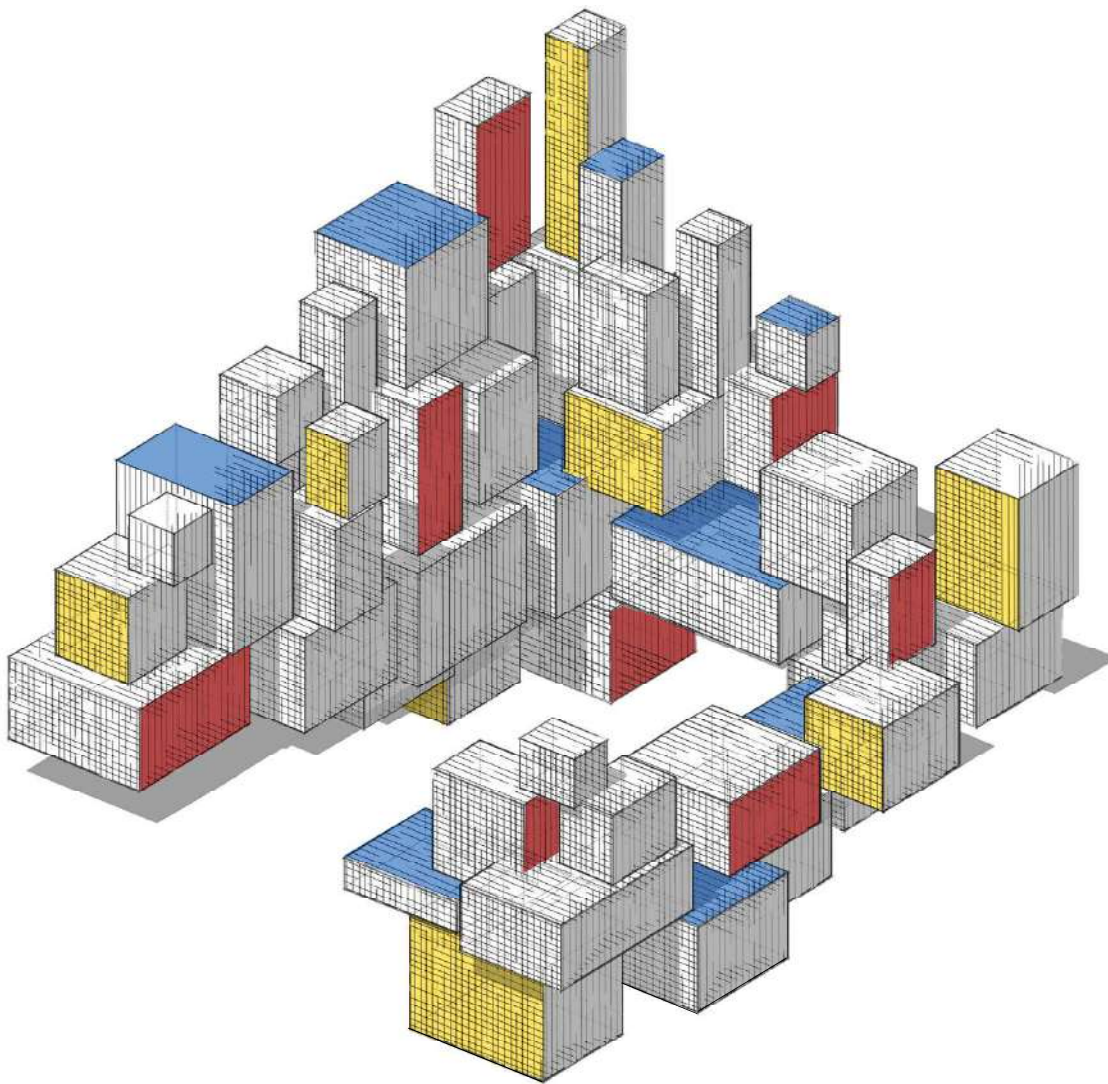


Data Science and Machine Learning

Mathematical and Statistical Methods



Dirk P. Kroese, Zdravko I. Botev, Thomas Taimre, Radislav Vaisman

8th May 2022

To my wife and daughters: Lesley, Elise, and Jessica

— DPK

To Sarah, Sofia, and my parents

— ZIB

To my grandparents: Arno, Harry, Jutta, and Maila

— TT

To Valerie

— RV

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

[Click here to download full PDF material](#)