Contents

List of Symbols ix
Preface xi
Acknowledgments xv

Part I Mathematical Foundations

1 Introduction and Motivation 3
1.1 Finding Words for Intuitions 3
1.2 Two Ways to Read This Book 5
1.3 Exercises and Feedback 7

2 Linear Algebra 8
2.1 Systems of Linear Equations 10
2.2 Matrices 12
2.3 Solving Systems of Linear Equations 17
2.4 Vector Spaces 24
2.5 Linear Independence 29
2.6 Basis and Rank 33
2.7 Linear Mappings 36
2.8 Affine Spaces 48
2.9 Further Reading 50
Exercises 51

3 Analytic Geometry 57
3.1 Norms 58
3.2 Inner Products 59
3.3 Lengths and Distances 61
3.4 Angles and Orthogonality 63
3.5 Orthonormal Basis 65
3.6 Orthogonal Complement 65
3.7 Inner Product of Functions 66
3.8 Orthogonal Projections 67
3.9 Rotations 76
3.10 Further Reading 79
Exercises 80

4	**Matrix Decompositions**	82
4.1	Determinant and Trace	83
4.2	Eigenvalues and Eigenvectors	88
4.3	Cholesky Decomposition	96
4.4	Eigendecomposition and Diagonalization	98
4.5	Singular Value Decomposition	101
4.6	Matrix Approximation	111
4.7	Matrix Phylogeny	115
4.8	Further Reading	116
	Exercises	118
5	**Vector Calculus**	120
5.1	Differentiation of Univariate Functions	122
5.2	Partial Differentiation and Gradients	126
5.3	Gradients of Vector-Valued Functions	129
5.4	Gradients of Matrices	135
5.5	Useful Identities for Computing Gradients	138
5.6	Backpropagation and Automatic Differentiation	138
5.7	Higher-Order Derivatives	143
5.8	Linearization and Multivariate Taylor Series	144
5.9	Further Reading	149
	Exercises	150
6	**Probability and Distributions**	152
6.1	Construction of a Probability Space	152
6.2	Discrete and Continuous Probabilities	157
6.3	Sum Rule, Product Rule, and Bayes' Theorem	163
6.4	Summary Statistics and Independence	165
6.5	Gaussian Distribution	175
6.6	Conjugacy and the Exponential Family	182
6.7	Change of Variables/Inverse Transform	191
6.8	Further Reading	197
	Exercises	198
7	**Continuous Optimization**	201
7.1	Optimization Using Gradient Descent	203
7.2	Constrained Optimization and Lagrange Multipliers	208
7.3	Convex Optimization	211
7.4	Further Reading	220
	Exercises	221
Part II	Central Machine Learning Problems	
8	**When Models Meet Data**	225
8.1	Data, Models, and Learning	225
8.2	Empirical Risk Minimization	232

8.3	Parameter Estimation	238
8.4	Probabilistic Modeling and Inference	244
8.5	Directed Graphical Models	249
8.6	Model Selection	254
9	**Linear Regression**	260
9.1	Problem Formulation	261
9.2	Parameter Estimation	263
9.3	Bayesian Linear Regression	273
9.4	Maximum Likelihood as Orthogonal Projection	282
9.5	Further Reading	283
10	**Dimensionality Reduction with Principal Component Analysis**	286
10.1	Problem Setting	286
10.2	Maximum Variance Perspective	289
10.3	Projection Perspective	293
10.4	Eigenvector Computation and Low-Rank Approximations	300
10.5	PCA in High Dimensions	302
10.6	Key Steps of PCA in Practice	303
10.7	Latent Variable Perspective	306
10.8	Further Reading	310
11	**Density Estimation with Gaussian Mixture Models**	314
11.1	Gaussian Mixture Model	315
11.2	Parameter Learning via Maximum Likelihood	316
11.3	EM Algorithm	325
11.4	Latent-Variable Perspective	328
11.5	Further Reading	332
12	**Classification with Support Vector Machines**	335
12.1	Separating Hyperplanes	337
12.2	Primal Support Vector Machine	338
12.3	Dual Support Vector Machine	347
12.4	Kernels	351
12.5	Numerical Solution	353
12.6	Further Reading	355
References		357
Index		367