

# Contents

Preface *page* xiii

## Part I Mathematical Foundations

<b>1</b>	<b>Solving Equations</b>	3
1.1	Linear Equation Systems	4
1.2	The Bisection and Secant Methods	6
1.2.1	The Bisection Search	6
1.2.2	The Secant Method	8
1.3	Fixed-Point Iteration	9
1.4	Newton–Raphson Method (Univariate)	20
1.5	Newton–Raphson Method (Multivariate)	24
	Problems	30
1.A	Order of Convergence of the Secant Method	31
1.B	Order of Convergence of the Newton–Raphson Method	32
1.C	Proofs of the Fixed-Point and Contraction Mapping Theorems	34
<b>2</b>	<b>Unconstrained Optimization</b>	35
2.1	Optimization by Solving Equations	36
2.2	Newton’s Method	37
2.3	Gradient Descent Method	43
2.4	Line Minimization	46
2.5	Quasi-Newton Methods	54
2.6	Conjugate Gradient Method	60
2.7	Issues of Local/Global Minimum	72
	Problems	73
<b>3</b>	<b>Constrained Optimization</b>	75
3.1	Optimization with Equality Constraints	76
3.2	Optimization with Inequality Constraints	79
3.3	Duality and KKT Conditions	84
3.4	Linear Programming (LP)	87
3.5	The Simplex Algorithm	94
3.6	Quadratic Programming (QP)	99
3.7	Interior Point Methods	101
	Problems	110

viii Contents

## Part II Regression

<b>4</b>	<b>Bias–Variance Tradeoff and Overfitting vs. Underfitting</b>	113
4.1	Supervised Learning as Optimization	113
4.2	Bias–Variance Tradeoff	117
4.3	Cross-Validation	121
4.4	Regularization and Ensemble Learning	122
<b>5</b>	<b>Linear Regression</b>	124
5.1	Linear Least Squares (LLS) Regression	124
5.2	Ridge Regression	135
5.3	Regression Based on Basis Functions	138
5.4	Bayesian Regression	148
	Problems	155
<b>6</b>	<b>Nonlinear Regression</b>	157
6.1	Nonlinear Least Squares Regression	157
6.2	Parameter Estimation by Natural Gradient Descent	164
	Problems	165
<b>7</b>	<b>Logistic and Softmax Regression</b>	166
7.1	Logistic Regression and Binary Classification	166
7.2	Softmax Regression for Multiclass Classification	174
	Problems	181
<b>8</b>	<b>Gaussian Process Regression and Classification</b>	183
8.1	Gaussian Process Regression	183
8.2	Gaussian Process Classifier – Binary	191
8.3	Gaussian Process Classifier – Multiclass	199
	Problems	207

## Part III Feature Extraction

<b>9</b>	<b>Feature Selection</b>	211
9.1	Distances and Separability Measurements	211
	Problems	215
<b>10</b>	<b>Principal Component Analysis</b>	218
10.1	Covariance and Correlation	218
10.2	Karhunen–Loève Transformation	220
10.3	Optimality of the KLT	223
10.4	Geometric Interpretation of the KLT	225
10.5	Computation of the KLT	228
10.6	Comparison with Other Orthogonal Transforms	229

	Contents	ix
10.7	Application to Image Data	232
10.8	PCA for Feature Extraction	237
	Problems	240
<b>11</b>	<b>Variations of PCA</b>	<b>243</b>
11.1	Kernel Methods	243
11.2	Kernel PCA	247
11.3	Factor Analysis and Expectation Maximization	250
11.4	Probabilistic PCA	257
11.5	Classical Multidimensional Scaling	262
11.6	t-Distributed Stochastic Neighbor Embedding	266
	Problems	274
<b>12</b>	<b>Independent Component Analysis</b>	<b>276</b>
12.1	Independence and Non-Gaussianity	276
12.2	Preprocessing and Whitening	278
12.3	Non-Gaussianity and FastICA	279
12.4	Likelihood and Independence Maximization	283
	Problems	288
<b>Part IV Classification</b>		
<b>13</b>	<b>Statistic Classification</b>	<b>293</b>
13.1	Discriminative vs. Generative Methods for Classification	293
13.2	k-nearest Neighbors and Minimum Distance Classifiers	294
13.3	Naive Bayes Classification	297
13.4	Adaptive Boosting	309
	Problems	320
<b>14</b>	<b>Support Vector Machine</b>	<b>321</b>
14.1	Maximum Margin and Support Vectors	321
14.2	Kernel SVM	330
14.3	Soft Margin SVM	331
14.4	Sequential Minimal Optimization Algorithm	333
14.5	Multiclass Classification	343
14.6	Kernelized Bayes Classifier	348
	Problems	361
<b>15</b>	<b>Clustering Analysis</b>	<b>363</b>
15.1	k-Means Clustering	363
15.2	Gaussian Mixture Model	370
15.3	Bernoulli Mixture Model	381
	Problems	385

<b>16 Hierarchical Classifiers</b>	387
16.1 Bottom-Up vs. Top-Down Methods	387
16.2 Binary Hierarchical Classification	389
16.3 Binary Hierarchical Clustering Problems	394
	401
<b>Part V Neural Networks</b>	
<b>17 Biologically Inspired Networks</b>	405
17.1 Biological Neural Network Modeling	405
17.2 Hebbian Learning	409
17.3 Hopfield Network	411
<b>18 Perceptron-Based Networks</b>	416
18.1 Perceptron	416
18.2 BackPropagation	425
18.3 Autoencoder	435
18.4 Deep Learning Problems	444
	449
<b>19 Competition-Based Networks</b>	451
19.1 Competitive Learning Network	451
19.2 Self-Organizing Map (SOM) Problems	458
	467
<b>Part VI Reinforcement Learning</b>	
<b>20 Introduction to Reinforcement Learning</b>	471
20.1 Markov Decision Process	471
20.1.1 Markov Chain	471
20.1.2 Markov Reward Process	472
20.1.3 Markov Decision Process	475
20.2 Model-Based Planning	478
20.3 Model-Free Evaluation and Control	483
20.3.1 Monte Carlo (MC) Algorithms	487
20.3.2 Temporal Difference (TD) Algorithms	490
20.3.3 TD( $\lambda$ ) Algorithm	497
20.4 Value Function Approximation	502
20.5 Control Based on Function Approximation	507
20.6 Deep Q-Learning	509
20.7 Policy Gradient Methods Problems	512
	519

## Part VII Large Language Models

<b>21 Large Language Models</b>	523
21.1 Natural Language Processing and Modeling	523
21.2 The Transformer Architecture	525
21.3 Attention Is All You Need	527
21.4 Backpropagation Learning	530
21.5 Transfer Learning in NLP	535
21.6 Reinforcement Learning with Human Feedback	536
21.7 Scaling Laws	537
21.8 Mathematical Operations in the Transformer Architecture	538
21.9 Emergence	539
21.10 Do LLMs Understand?	541
21.11 Future Directions	545
21.12 Applications of Transformers Beyond NLP	547
21.13 Generative Methods Beyond Transformers	550
Index	554