Contents

1 Introduction 1
References 11

2 Linear Regression 13
2.1 Least Squares Linear Regression 13
2.2 Principal Component Analysis and Principal Component Regression 24
2.3 Least Absolute Shrinkage and Selection Operator (L_1) 34
2.4 Ridge Regression (L_2) 38
2.5 Elastic Net Regression 43
2.6 Multiply Task LASSO (MultiTaskLASSO) 47
References 50

3 Linear Classification 53
3.1 Perceptron 55
3.2 Logistic Regression 59
3.3 Linear Discriminant Analysis 73
References 83

4 Support Vector Machine 85
4.1 SVC 85
4.2 Kernel Functions 91
4.3 Soft Margin 100
4.4 SVR 107
References 115

5 Decision Tree and *K*-Nearest-Neighbors (KNN) 117
5.1 Introduction 117
5.2 Classification Trees 117
5.3 Regression Tree 126
5.4 *K*-Nearest-Neighbors (KNN) Methods 134
References 140

6 Ensemble Learning 141
6.1 Boosting 142
6.1.1 AdaBoost Classification 142
6.1.2 AdaBoost Regression and Gradient Boosting Machine (GBM) 150
6.1.3 The Second-Order Expansion of Loss Function 155
6.1.4 EXtreme Gradient Boosting (XGBoost) 156
6.2 Bagging 158
References 164

7 Bayesian Theorem and Expectation–Maximization (EM) Algorithm 167
7.1 Bayesian Theorem 167
7.2 Naive Bayes Classifier 168
7.3 Maximum Likelihood Estimation 177
7.3.1 Gaussian Distribution 177
7.3.2 Weibull Distribution 179
7.4 Bayesian Linear Regression 184
7.5 Expectation–Maximization (EM) Algorithm 193
7.5.1 Gaussian Mixture Model (GMM) 193
7.5.2 The Mixture of Lorentz and Gaussian Distributions 207
7.6 Gaussian Process (GP) Regression 218
References 228

8 Symbolic Regression 229
8.1 Overview of Evolutionary Computation 229
8.2 Genetic Programming 231
8.3 Grammar-Guided Genetic Programming and Grammatical Evolution 233
8.4 The Application of LASSO in Symbolic Regression 242
References 243

9 Neural Networks 245
9.1 Neural Networks and Perceptron 245
9.2 Back Propagation Algorithm 247
9.3 Regularization in NNs 257
9.3.1 L_1 Regularization 257
9.3.2 L_2 Regularization 266
9.4 Classification NNs 270
9.4.1 Binary Classification 270
9.4.2 Multiclassification of Multiply Grades in a Category 276
9.5 Autoencoders 280
9.5.1 Introduction 280
9.5.2 Denoising Autoencoder 281

9.5.3 Sparse Autoencoder 289
9.5.4 Variational Autoencoder 298
References 322

10 Hidden Markov Chains 325
10.1 Markov Chain 325
10.2 Stationary Markov Chain 329
10.3 Markov Chain Monte Carlo Methods 330
10.3.1 Metropolis Hastings (M-H) Algorithm 332
10.3.2 Gibbs Sampling Algorithm 333
10.4 Calculation Methods for the Probability of Observation Sequence 337
10.4.1 Direct Method 337
10.4.2 Forward Method 340
10.4.3 Backward Method 342
10.5 Estimation of Optimal State Sequence 345
10.5.1 Direct Method 345
10.5.2 Viterbi Algorithm 346
10.6 Estimation of Intrinsic Parameters—The Baum-Welch Algorithm 348
References 360

11 Data Preprocessing and Feature Selection 363
11.1 Reliable Data, Normals and Anomalies 364
11.1.1 Local Outlier Factor 364
11.1.2 Isolated Forest 368
11.1.3 One-Class Support Vector Machine 371
11.1.4 Support Vector Data Description 376
11.2 Feature Selection 381
11.2.1 Filter Approach 382
11.2.2 Wrapper Approach 411
11.2.3 Embedded Approach 419
References 426

12 Interpretative SHAP Value and Partial Dependence Plot 429
12.1 SHapley Additive ExPlanation Value 429
12.2 The Joint SHAP Value of Two Features 446
12.3 Partial Dependence Plot 447
12.3.1 Mono-Feature Partial Dependence Plot 448
12.3.2 Multifeature Partial Dependence Plot 453
References 461

Appendix A: Vector and Matrix 463

Appendix B: Basic Statistics 471

Index 477