

CONTENTS

<i>Preface</i>	<i>xvii</i>
<i>Acknowledgements</i>	<i>xxi</i>
1. Introduction	1
1.1 Towards Intelligent Machines	1
1.2 Well-Posed Machine Learning Problems	5
1.3 Examples of Applications in Diverse Fields	7
1.4 Data Representation	12
1.4.1 Time Series Forecasting	15
1.4.2 Datasets for Toy (Unrealistically Simple) and Realistic Problems	17
1.5 Domain Knowledge for Productive use of Machine Learning	18
1.6 Diversity of Data: Structured/Unstructured	20
1.7 Forms of Learning	21
1.7.1 Supervised/Directed Learning	21
1.7.2 Unsupervised/Undirected Learning	22
1.7.3 Reinforcement Learning	22
1.7.4 Learning Based on Natural Processes: Evolution, Swarming, and Immune Systems	23
1.8 Machine Learning and Data Mining	25
1.9 Basic Linear Algebra in Machine Learning Techniques	26
1.10 Relevant Resources for Machine Learning	34

2. Supervised Learning: Rationale and Basics	36
2.1 Learning from Observations	36
2.2 Bias and Variance	42
2.3 Why Learning Works: Computational Learning Theory	46
2.4 Occam's Razor Principle and Overfitting Avoidance	49
2.5 Heuristic Search in Inductive Learning	51
2.5.1 Search through Hypothesis Space	52
2.5.2 Ensemble Learning	53
2.5.3 Evaluation of a Learning System	55
2.6 Estimating Generalization Errors	56
2.6.1 Holdout Method and Random Subsampling	56
2.6.2 Cross-validation	57
2.6.3 Bootstrapping	58
2.7 Metrics for Assessing Regression (Numeric Prediction) Accuracy	59
2.7.1 Mean Square Error	60
2.7.2 Mean Absolute Error	60
2.8 Metrics for Assessing Classification (Pattern Recognition) Accuracy	61
2.8.1 Misclassification Error	61
2.8.2 Confusion Matrix	62
2.8.3 Comparing Classifiers Based on ROC Curves	66
2.9 An Overview of the Design Cycle and Issues in Machine Learning	68
 3. Statistical Learning	 73
3.1 Machine Learning and Inferential Statistical Analysis	73
3.2 Descriptive Statistics in Learning Techniques	74
3.2.1 Representing Uncertainties in Data: Probability Distributions	75
3.2.2 Descriptive Measures of Probability Distributions	80
3.2.3 Descriptive Measures from Data Sample	83
3.2.4 Normal Distributions	84
3.2.5 Data Similarity	85
3.3 Bayesian Reasoning: A Probabilistic Approach to Inference	87
3.3.1 Bayes Theorem	88
3.3.2 Naive Bayes Classifier	93
3.3.3 Bayesian Belief Networks	98
3.4 k -Nearest Neighbor (k -NN) Classifier	102
3.5 Discriminant Functions and Regression Functions	106
3.5.1 Classification and Discriminant Functions	107
3.5.2 Numeric Prediction and Regression Functions	108
3.5.3 Practical Hypothesis Functions	109

3.6	Linear Regression with Least Square Error Criterion	112
3.6.1	Minimal Sum-of-Error-Squares and the Pseudoinverse	113
3.6.2	Gradient Descent Optimization Schemes	115
3.6.3	Least Mean Square (LMS) Algorithm	115
3.7	Logistic Regression for Classification Tasks	116
3.8	Fisher's Linear Discriminant and Thresholding for Classification	120
3.8.1	Fisher's Linear Discriminant	120
3.8.2	Thresholding	125
3.9	Minimum Description Length Principle	126
3.9.1	Bayesian Perspective	127
3.9.2	Entropy and Information	128
4.	Learning With Support Vector Machines (SVM)	130
4.1	Introduction	130
4.2	Linear Discriminant Functions for Binary Classification	132
4.3	Perceptron Algorithm	136
4.4	Linear Maximal Margin Classifier for Linearly Separable Data	141
4.5	Linear Soft Margin Classifier for Overlapping Classes	152
4.6	Kernel-Induced Feature Spaces	158
4.7	Nonlinear Classifier	162
4.8	Regression by Support Vector Machines	167
4.8.1	Linear Regression	169
4.8.2	Nonlinear Regression	172
4.9	Decomposing Multiclass Classification Problem Into Binary Classification Tasks	174
4.9.1	One-Against-All (OAA)	175
4.9.2	One-Against-One (OAO)	176
4.10	Variants of Basic SVM Techniques	177
5.	Learning With Neural Networks (NN)	181
5.1	Towards Cognitive Machine	181
5.1.1	From Perceptrons to Deep Networks	182
5.2	Neuron Models	184
5.2.1	Biological Neuron	184
5.2.2	Artificial Neuron	186
5.2.3	Mathematical Model	190
5.3	Network Architectures	193
5.3.1	Feedforward Networks	194
5.3.2	Recurrent Networks	199

5.4	Perceptrons	200
5.4.1	Limitations of Perceptron Algorithm for Linear Classification Tasks	201
5.4.2	Linear Classification using Regression Techniques	201
5.4.3	Standard Gradient Descent Optimization Scheme: Steepest Descent	203
5.5	Linear Neuron and the Widrow-Hoff Learning Rule	206
5.5.1	Stochastic Gradient Descent	207
5.6	The Error-Correction Delta Rule	208
5.6.1	Sigmoid Unit: Soft-Limiting Perceptron	211
5.7	Multi-Layer Perceptron (MLP) Networks and the Error-Backpropagation Algorithm	213
5.7.1	The Generalized Delta Rule	216
5.7.2	Convergence and Local Minima	226
5.7.3	Adding Momentum to Gradient Descent	227
5.7.4	Heuristic Aspects of the Error-backpropagation Algorithm	228
5.8	Multi-Class Discrimination with MLP Networks	232
5.9	Radial Basis Functions (RBF) Networks	235
5.9.1	Training the RBF Network	239
5.10	Genetic-Neural Systems	241

6. Fuzzy Inference Systems

245

6.1	Introduction	245
6.2	Cognitive Uncertainty and Fuzzy Rule-Base	248
6.3	Fuzzy Quantification of Knowledge	253
6.3.1	Fuzzy Logic	253
6.3.2	Fuzzy Sets	257
6.3.3	Fuzzy Set Operations	267
6.3.4	Fuzzy Relations	268
6.4	Fuzzy Rule-Base and Approximate Reasoning	277
6.4.1	Quantification of Rules via Fuzzy Relations	281
6.4.2	Fuzzification of Input	283
6.4.3	Inference Mechanism	284
6.4.4	Defuzzification of Inferred Fuzzy Set	298
6.5	Mamdani Model for Fuzzy Inference Systems	301
6.5.1	Mobile Robot Navigation Among Moving Obstacles	303
6.5.2	Mortgage Loan Assessment	308
6.6	Takagi-Sugeno Fuzzy Model	311
6.7	Neuro-Fuzzy Inference Systems	317
6.7.1	ANFIS Architecture	318
6.7.2	How Does an ANFIS Learn?	320
6.8	Genetic-Fuzzy Systems	324

7. Data Clustering and Data Transformations 328

- 7.1 Unsupervised Learning 328
 - 7.1.1 Clustering 329
- 7.2 Engineering the Data 331
 - 7.2.1 Exploratory Data Analysis: Learning about What is in the Data 333
 - 7.2.2 Cluster Analysis: Finding Similarities in the Data 334
 - 7.2.3 Data Transformations: Enhancing the Information Content of the Data 339
- 7.3 Overview of Basic Clustering Methods 341
 - 7.3.1 Partitional Clustering 341
 - 7.3.2 Hierarchical Clustering 344
 - 7.3.3 Spectral Clustering 345
 - 7.3.4 Clustering using Self-Organizing Maps 349
- 7.4 K-Means Clustering 352
- 7.5 Fuzzy K-Means Clustering 356
- 7.6 Expectation-Maximization (EM) Algorithm and Gaussian Mixtures Clustering 362
 - 7.6.1 EM Algorithm 362
 - 7.6.2 Gaussian Mixture Models 365
- 7.7 Some Useful Data Transformations 372
 - 7.7.1 Data Cleansing 372
 - 7.7.2 Derived Attributes 374
 - 7.7.3 Discretizing Numeric Attributes 375
 - 7.7.4 Attribute Reduction Techniques 377
- 7.8 Entropy-Based Method for Attribute Discretization 377
- 7.9 Principal Components Analysis (PCA) for Attribute Reduction 382
- 7.10 Rough Sets-Based Methods for Attribute Reduction 390
 - 7.10.1 Rough Set Preliminaries 392
 - 7.10.2 Analysis of Relevance of Attributes 397
 - 7.10.3 Reduction of Attributes 399

8. Decision Tree Learning 404

- 8.1 Introduction 404
- 8.2 Example of a Classification Decision Tree 406
- 8.3 Measures of Impurity for Evaluating Splits in Decision Trees 411
 - 8.3.1 Information Gain/Entropy reduction 411
 - 8.3.2 Gain Ratio 416
 - 8.3.3 Gini Index 417
- 8.4 ID3, C4.5, and CART Decision Trees 418
- 8.5 Pruning the Tree 427
- 8.6 Strengths and Weaknesses of Decision-Tree Approach 429
- 8.7 Fuzzy Decision Trees 433

9. Business Intelligence and Data Mining: Techniques and Applications	445
9.1 An Introduction to Analytics 445	
9.1.1 Machine Learning, Data Mining, and Predictive Analytics 448	
9.1.2 Basic Analytics Techniques 449	
9.2 The CRISP-DM (Cross Industry Standard Process for Data Mining) Model 451	
9.3 Data Warehousing and Online Analytical Processing 456	
9.3.1 Basic Concepts 456	
9.3.2 Databases 458	
9.3.3 Data Warehousing: A General Architecture, and OLAP Operations 461	
9.3.4 Data Mining in the Data Warehouse Environment 466	
9.4 Mining Frequent Patterns and Association Rules 467	
9.4.1 Basic Concepts 469	
9.4.2 Measures of Strength of Frequent Patterns and Association Rules 471	
9.4.3 Frequent Item Set Mining Methods 473	
9.4.4 Generating Association Rules from Frequent Itemsets 477	
9.5 Intelligent Information Retrieval Systems 479	
9.5.1 Text Retrieval 483	
9.5.2 Image Retrieval 486	
9.5.3 Audio Retrieval 488	
9.6 Applications and Trends 490	
9.6.1 Data Mining Applications 490	
9.6.2 Data Mining Trends 495	
9.7 Technologies for Big Data 498	
9.7.1 Emerging Analytic Methods 500	
9.7.2 Emerging Technologies for Higher Levels of Scalability 503	
 Appendix A Genetic Algorithm (GA) For Search Optimization	 508
A.1 A Simple Overview of Genetics 510	
A.2 Genetics on Computers 511	
A.3 The Basic Genetic Algorithm 515	
A.4 Beyond the Basic Genetic Algorithm 524	
 Appendix B Reinforcement Learning (RL)	 527
B.1 Introduction 527	
B.2 Elements of Reinforcement Learning 530	
B.3 Basics of Dynamic Programming 535	
B.3.1 Finding Optimal Policies 538	
B.3.2 Value Iteration 539	
B.3.3 Policy Iteration 540	

B.4	Temporal Difference Learning	542
B.4.1	Q-learning	544
B.4.2	Generalization	546
B.4.3	Sarsa-learning	548
Datasets from Real-Life Applications for Machine Learning Experiments		549
Problems		567
References		613
Index		623