
Contents

Preface

vii

Notations

ix

1 Introduction

1

1.1	Basic Concepts	1
1.2	Popular Learning Algorithms	3
1.2.1	Linear Discriminant Analysis	3
1.2.2	Decision Trees	4
1.2.3	Neural Networks	6
1.2.4	Naïve Bayes Classifier	8
1.2.5	k -Nearest Neighbor	9
1.2.6	Support Vector Machines and Kernel Methods	9
1.3	Evaluation and Comparison	12
1.4	Ensemble Methods	15
1.5	Applications of Ensemble Methods	17
1.6	Further Readings	20

2 Boosting

23

2.1	A General Boosting Procedure	23
2.2	The AdaBoost Algorithm	24
2.3	Illustrative Examples	28
2.4	Theoretical Issues	32
2.4.1	Initial Analysis	32
2.4.2	Margin Explanation	32
2.4.3	Statistical View	35
2.5	Multiclass Extension	38
2.6	Noise Tolerance	41
2.7	Further Readings	44

3 Bagging

47

3.1	Two Ensemble Paradigms	47
3.2	The Bagging Algorithm	48
3.3	Illustrative Examples	50
3.4	Theoretical Issues	53
3.5	Random Tree Ensembles	57
3.5.1	Random Forest	57

3.5.2	Spectrum of Randomization	59
3.5.3	Random Tree Ensembles for Density Estimation	61
3.5.4	Random Tree Ensembles for Anomaly Detection	64
3.6	Further Readings	66
4	Combination Methods	67
4.1	Benefits of Combination	67
4.2	Averaging	68
4.2.1	Simple Averaging	68
4.2.2	Weighted Averaging	70
4.3	Voting	71
4.3.1	Majority Voting	72
4.3.2	Plurality Voting	73
4.3.3	Weighted Voting	74
4.3.4	Soft Voting	75
4.3.5	Theoretical Issues	77
4.4	Combining by Learning	83
4.4.1	Stacking	83
4.4.2	Infinite Ensemble	86
4.5	Other Combination Methods	87
4.5.1	Algebraic Methods	87
4.5.2	Behavior Knowledge Space Method	88
4.5.3	Decision Template Method	89
4.6	Relevant Methods	89
4.6.1	Error-Correcting Output Codes	90
4.6.2	Dynamic Classifier Selection	93
4.6.3	Mixture of Experts	93
4.7	Further Readings	95
5	Diversity	99
5.1	Ensemble Diversity	99
5.2	Error Decomposition	100
5.2.1	Error-Ambiguity Decomposition	100
5.2.2	Bias-Variance-Covariance Decomposition	102
5.3	Diversity Measures	105
5.3.1	Pairwise Measures	105
5.3.2	Non-Pairwise Measures	106
5.3.3	Summary and Visualization	109
5.3.4	Limitation of Diversity Measures	110
5.4	Information Theoretic Diversity	111
5.4.1	Information Theory and Ensemble	111
5.4.2	Interaction Information Diversity	112
5.4.3	Multi-Information Diversity	113
5.4.4	Estimation Method	114
5.5	Diversity Generation	116

5.6	Further Readings	118
6	Ensemble Pruning	119
6.1	What Is Ensemble Pruning	119
6.2	Many Could Be Better Than All	120
6.3	Categorization of Pruning Methods	123
6.4	Ordering-Based Pruning	124
6.5	Clustering-Based Pruning	127
6.6	Optimization-Based Pruning	128
6.6.1	Heuristic Optimization Pruning	128
6.6.2	Mathematical Programming Pruning	129
6.6.3	Probabilistic Pruning	131
6.7	Further Readings	133
7	Clustering Ensembles	135
7.1	Clustering	135
7.1.1	Clustering Methods	135
7.1.2	Clustering Evaluation	137
7.1.3	Why Clustering Ensembles	139
7.2	Categorization of Clustering Ensemble Methods	141
7.3	Similarity-Based Methods	142
7.4	Graph-Based Methods	144
7.5	Relabeling-Based Methods	147
7.6	Transformation-Based Methods	152
7.7	Further Readings	155
8	Advanced Topics	157
8.1	Semi-Supervised Learning	157
8.1.1	Usefulness of Unlabeled Data	157
8.1.2	Semi-Supervised Learning with Ensembles	159
8.2	Active Learning	163
8.2.1	Usefulness of Human Intervention	163
8.2.2	Active Learning with Ensembles	165
8.3	Cost-Sensitive Learning	166
8.3.1	Learning with Unequal Costs	166
8.3.2	Ensemble Methods for Cost-Sensitive Learning	167
8.4	Class-Imbalance Learning	171
8.4.1	Learning with Class Imbalance	171
8.4.2	Performance Evaluation with Class Imbalance	172
8.4.3	Ensemble Methods for Class-Imbalance Learning	176
8.5	Improving Comprehensibility	179
8.5.1	Reduction of Ensemble to Single Model	179
8.5.2	Rule Extraction from Ensembles	180
8.5.3	Visualization of Ensembles	181
8.6	Future Directions of Ensembles	182

8.7 Further Readings	184
--------------------------------	-----

References	187
-------------------	------------

Index	219
--------------	------------