

Contents

<i>Preface</i>	vii
1. Introduction to Pattern Classification	1
1.1 Pattern Classification	2
1.2 Induction Algorithms	4
1.3 Rule Induction	5
1.4 Decision Trees	5
1.5 Bayesian Methods	8
1.5.1 Overview	8
1.5.2 Naïve Bayes	9
1.5.2.1 The Basic Naïve Bayes Classifier	9
1.5.2.2 Naïve Bayes Induction for Numeric Attributes	12
1.5.2.3 Correction to the Probability Estimation	12
1.5.2.4 Laplace Correction	13
1.5.2.5 No Match	14
1.5.3 Other Bayesian Methods	14
1.6 Other Induction Methods	14
1.6.1 Neural Networks	14
1.6.2 Genetic Algorithms	17
1.6.3 Instance-based Learning	17
1.6.4 Support Vector Machines	18
2. Introduction to Ensemble Learning	19
2.1 Back to the Roots	20
2.2 The Wisdom of Crowds	22

2.3	The Bagging Algorithm	22
2.4	The Boosting Algorithm	28
2.5	The AdaBoost Algorithm	28
2.6	No Free Lunch Theorem and Ensemble Learning	36
2.7	Bias-Variance Decomposition and Ensemble Learning	38
2.8	Occam's Razor and Ensemble Learning	40
2.9	Classifier Dependency	41
2.9.1	Dependent Methods	42
2.9.1.1	Model-guided Instance Selection	42
2.9.1.2	Basic Boosting Algorithms	42
2.9.1.3	Advanced Boosting Algorithms	44
2.9.1.4	Incremental Batch Learning	51
2.9.2	Independent Methods	51
2.9.2.1	Bagging	53
2.9.2.2	Wagging	54
2.9.2.3	Random Forest and Random Subspace Projection	55
2.9.2.4	Non-Linear Boosting Projection (NLBP)	56
2.9.2.5	Cross-validated Committees	58
2.9.2.6	Robust Boosting	59
2.10	Ensemble Methods for Advanced Classification Tasks	61
2.10.1	Cost-Sensitive Classification	61
2.10.2	Ensemble for Learning Concept Drift	63
2.10.3	Reject Driven Classification	63
3.	Ensemble Classification	65
3.1	Fusions Methods	65
3.1.1	Weighting Methods	65
3.1.2	Majority Voting	66
3.1.3	Performance Weighting	67
3.1.4	Distribution Summation	68
3.1.5	Bayesian Combination	68
3.1.6	Dempster-Shafer	69
3.1.7	Vogging	69
3.1.8	Naïve Bayes	69
3.1.9	Entropy Weighting	70
3.1.10	Density-based Weighting	70
3.1.11	DEA Weighting Method	70
3.1.12	Logarithmic Opinion Pool	71

3.1.13	Order Statistics	71
3.2	Selecting Classification	71
3.2.1	Partitioning the Instance Space	74
3.2.1.1	The K–Means Algorithm as a Decomposition Tool	75
3.2.1.2	Determining the Number of Subsets	78
3.2.1.3	The Basic K–Classifier Algorithm	78
3.2.1.4	The Heterogeneity Detecting K–Classifier (HDK–Classifier)	81
3.2.1.5	Running–Time Complexity	81
3.3	Mixture of Experts and Meta Learning	82
3.3.1	Stacking	82
3.3.2	Arbiter Trees	85
3.3.3	Combiner Trees	88
3.3.4	Grading	88
3.3.5	Gating Network	89
4.	Ensemble Diversity	93
4.1	Overview	93
4.2	Manipulating the Inducer	94
4.2.1	Manipulation of the Inducer’s Parameters	95
4.2.2	Starting Point in Hypothesis Space	95
4.2.3	Hypothesis Space Traversal	95
4.3	Manipulating the Training Samples	96
4.3.1	Resampling	96
4.3.2	Creation	97
4.3.3	Partitioning	100
4.4	Manipulating the Target Attribute Representation	101
4.4.1	Label Switching	102
4.5	Partitioning the Search Space	103
4.5.1	Divide and Conquer	104
4.5.2	Feature Subset-based Ensemble Methods	105
4.5.2.1	Random-based Strategy	106
4.5.2.2	Reduct-based Strategy	106
4.5.2.3	Collective-Performance-based Strategy	107
4.5.2.4	Feature Set Partitioning	108
4.5.2.5	Rotation Forest	111
4.6	Multi-Inducers	112
4.7	Measuring the Diversity	114

5. Ensemble Selection	119
5.1 Ensemble Selection	119
5.2 Pre Selection of the Ensemble Size	120
5.3 Selection of the Ensemble Size While Training	120
5.4 Pruning - Post Selection of the Ensemble Size	121
5.4.1 Ranking-based	122
5.4.2 Search based Methods	123
5.4.2.1 Collective Agreement-based Ensemble Pruning Method	124
5.4.3 Clustering-based Methods	129
5.4.4 Pruning Timing	129
5.4.4.1 Pre-combining Pruning	129
5.4.4.2 Post-combining Pruning	130
6. Error Correcting Output Codes	133
6.1 Code-matrix Decomposition of Multiclass Problems	135
6.2 Type I - Training an Ensemble Given a Code-Matrix	136
6.2.1 Error correcting output codes	138
6.2.2 Code-Matrix Framework	139
6.2.3 Code-matrix Design Problem	140
6.2.4 Orthogonal Arrays (OA)	144
6.2.5 Hadamard Matrix	146
6.2.6 Probabilistic Error Correcting Output Code	146
6.2.7 Other ECOC Strategies	147
6.3 Type II - Adapting Code-matrices to the Multiclass Problems	149
7. Evaluating Ensembles of Classifiers	153
7.1 Generalization Error	153
7.1.1 Theoretical Estimation of Generalization Error	154
7.1.2 Empirical Estimation of Generalization Error	155
7.1.3 Alternatives to the Accuracy Measure	157
7.1.4 The F-Measure	158
7.1.5 Confusion Matrix	160
7.1.6 Classifier Evaluation under Limited Resources	161
7.1.6.1 ROC Curves	163
7.1.6.2 Hit Rate Curve	163
7.1.6.3 Qrecall (Quota Recall)	164

7.1.6.4	Lift Curve	164
7.1.6.5	Pearson Correlation Coefficient	165
7.1.6.6	Area Under Curve (AUC)	166
7.1.6.7	Average Hit Rate	167
7.1.6.8	Average Qrecall	168
7.1.6.9	Potential Extract Measure (PEM)	170
7.1.7	Statistical Tests for Comparing Ensembles	172
7.1.7.1	McNemar's Test	173
7.1.7.2	A Test for the Difference of Two Proportions	174
7.1.7.3	The Resampled Paired t Test	175
7.1.7.4	The k -fold Cross-validated Paired t Test	176
7.2	Computational Complexity	176
7.3	Interpretability of the Resulting Ensemble	177
7.4	Scalability to Large Datasets	178
7.5	Robustness	179
7.6	Stability	180
7.7	Flexibility	180
7.8	Usability	180
7.9	Software Availability	180
7.10	Which Ensemble Method Should be Used?	181
	<i>Bibliography</i>	185
	<i>Index</i>	223