Contents

Foreword **xiii**

Acknowledgements **xv**

Abstract **xvii**

1 Introduction **1**
1.1 Objective 2
1.2 Roadmap and contributions 2
1.3 Rationale for the contributions 4
1.3.1 Characteristics of a classifier 4
1.3.2 Why do we need new classifiers? 6

2 Bayesian foundations of the learning process **9**
2.1 Probability theory as extended logic 10
2.1.1 The basic desiderata 10
2.1.2 Justifying the desiderata 10
2.2 The quantitative rules 12
2.2.1 The product rule 12
2.2.2 The sum rule 13
2.3 Prior probabilities 15
2.3.1 The principle of indifference 15
2.3.2 The entropy principle 16
2.3.3 Other methods 16
2.4 Summary 17

3 Bayesian network classifiers **19**
3.1 The problem of classification 20
3.2 Bayesian networks for classification 20
3.2.1 Dirichlet distributions 21
3.3 Naive Bayes 22
3.4 Learning with trees 23
3.4.1 Learning maximum likelihood TAN 23
3.5 Bayesian model averaging for classification 24

http://libros.csic.es

3.6 Summary . . . 25

4 A parallelizable distance-based discretization method 27
4.1 Introduction . . . 28
4.2 Discretization methods classification . . . 28
4.3 Some discretization methods . . . 29
4.3.1 Equal size . . . 29
4.3.2 Equal frequency . . . 29
4.3.3 ChiMerge . . . 29
4.3.4 Entropy . . . 30
4.3.5 D-2 . . . 32
4.3.6 Other discretization methods . . . 33
4.4 Distance-Based discretization method . . . 33
4.4.1 Cutpoint selection criterion . . . 34
4.4.2 The stopping criterion . . . 35
4.4.3 Computational complexity . . . 36
4.4.4 Parallelization of the method . . . 36
4.5 Empirical comparison . . . 38
4.5.1 Comparison design . . . 38
4.5.2 Comparison results . . . 38
4.6 Summary . . . 40

5 The Qualitative Bayesian Classifier 41
5.1 Introduction . . . 42
5.2 Introduction to Qualitative Probabilistic Networks 43
5.2.1 Wellman approach . . . 43
5.2.2 Neufeld approach . . . 44
5.3 Influences and synergies revisited . . . 44
5.4 The Qualitative Bayesian Classifier . . . 45
5.5 Empirical comparison . . . 48
5.5.1 Result analysis and justification . . . 49
5.6 Examples of explanations and characterizations . . . 49
5.6.1 Qualitative influences for characterization . . . 49
5.6.2 Explanation with qualitative influences and synergies . . 51
5.6.3 Comparison with c4.5rules results . . . 51
5.7 Summary . . . 52

6 The Indifferent Bayesian Classifier 55
6.1 The naive Bayes model . . . 56
6.1.1 The naive Bayes model as a Bayesian network . . . 56
6.1.2 The naive Bayes model as a Markov network . . . 56
6.1.3 Naive Bayes parameters . . . 57
6.2 Naive distributions . . . 60
6.2.1 Calculating probabilities with naive distributions . . . 60
6.2.2 Learning with naive distributions . . . 61
6.3 The Indifferent Naive Bayes Classifier . . . 61

http://libros.csic.es

6.4 Experimental results 63
6.4.1 Dataset description 63
6.4.2 Interpretation of the results 65
6.5 Summary 71

7 Empirical Local Bayesian model averaging of TAN classifiers 73
7.1 Local Bayesian model averaging for TAN induction 74
7.1.1 Local Bayesian model averaging 75
7.1.2 Empirical local Bayesian model averaging of TAN models 76
7.1.3 Computational complexity 78
7.2 Experimental results 80
7.2.1 Adjusting the algorithm to run 80
7.2.2 Experimental setting 80
7.2.3 Interpretation of the results 80
7.3 Summary 85
7.3.1 Further research 86

8 Tractable Bayesian Model Averaging of Tree Augmented Naive Bayes Classifiers 87
8.1 Decomposable distributions over tree belief networks 88
8.1.1 Definition 88
8.1.2 Meila and Jaakkola results and corrections to their results 90
8.2 Development of the Averaged Tree Augmented Naive Bayes . . . 92
8.2.1 Decomposable distributions over TANs 92
8.2.2 Calculating probabilities under decomposable distributions over TANs 95
8.2.3 Learning under decomposable distributions over TANs . . 95
8.2.4 Putting it all Together 96
8.3 Approximating TBMATAN 96
8.3.1 TBMATAN computational complexity 96
8.3.2 Computational problems 97
8.3.3 A solution to TBMATAN computational problems 97
8.4 Empirical Results 99
8.4.1 Interpretation of the Results 99
8.5 Conclusions and Future Work 110
8.5.1 Future work 110

9 Maximum a Posteriori Tree Augmented Naive Bayes Classifiers 111
9.1 Maximum a Posteriori results for decomposable distributions over trees 112
9.1.1 Calculating the most probable tree under a decomposable distribution over trees 112
9.1.2 Calculating the MAP tree given a prior decomposable distribution over trees 112
9.1.3 Calculating the k MAP trees and their relative weights given a prior decomposable distribution over trees 114

http://libros.csic.es

9.2 MAPTAN and MAPTAN+BMA classifiers . . . 115
9.2.1 Maximum a Posteriori results for decomposable distributions over TANs . . . 115
9.2.2 Constructing the MAPTAN and MAPTAN+BMA classifiers . 115
9.3 Empirical results . . . 119
9.3.1 Interpretation of the results . . . 120
9.4 Conclusions and future work . . . 132
9.4.1 Future work . . . 133

10 Conclusions **135**
10.1 Main contributions and its relevance . . . 135
10.1.1 A parallelizable discretization method . . . 135
10.1.2 Qualitative influences and synergies . . . 136
10.1.3 First Order Qualitative Bayesian Classifier . . . 136
10.1.4 Second Order Qualitative Bayesian Classifier . . . 136
10.1.5 Naive distributions . . . 136
10.1.6 The indifferent naive Bayes classifier . . . 136
10.1.7 Empirical local Bayesian model averaging of TAN . . . 137
10.1.8 Decomposable distributions over TAN models . . . 137
10.1.9 Tractable Bayesian model averaging of TAN . . . 137
10.1.10 Maximum a posteriori TAN classifier . . . 137
10.1.11 Maximum a posteriori local Bayesian model averaging of TAN . . . 138
10.2 Publication list . . . 138
10.3 Where to go from here? . . . 139

A Mathematical developments for the Indifferent Bayesian Classifier **141**
A.1 Preliminaries . . . 141
A.1.1 A multiple variable constrained integral . . . 141
A.1.2 A bit of notation . . . 142
A.2 Calculating probabilities with naive distributions . . . 142
A.3 Learning with naive distributions . . . 144
A.3.1 Computing the normalization constant . . . 145
A.3.2 Computing the posterior distribution . . . 145

B Mathematical developments for the Tractable Bayesian Model Averaging of Tree Augmented Naive Bayes Classifiers **147**
B.1 Preliminaries . . . 147
B.1.1 The matrix tree theorem . . . 147
B.1.2 The matrix tree theorem for decomposable distributions . 148
B.1.3 A useful result about Dirichlet distributions . . . 148
B.2 Detailed development for decomposable distributions over trees results . . . 149
B.2.1 Calculating probabilities under decomposable distributions over trees . . . 149

http://libros.csic.es