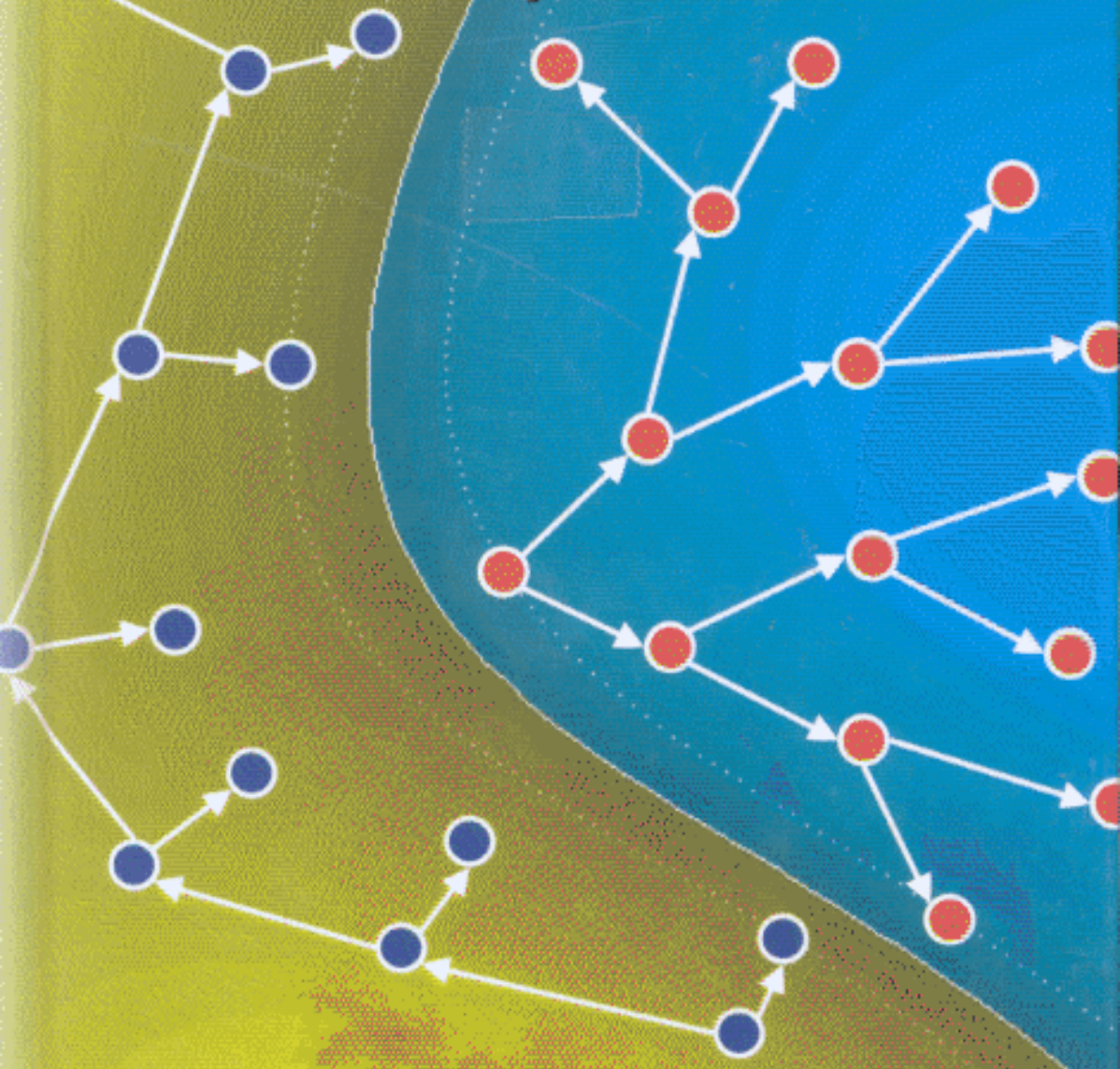
Machine Learning

Discriminative and Generative

Tony Jebara



Kluwer Academic Publishers

Contents

List of Figures ix
List of Tables xi
Preface xiii
Acknowledgments xvii

1. INTRODUCTION 1
   1 Machine Learning Roots 2
   2 Generative Learning 5
      2.1 Generative Models in AI 8
      2.2 Generative Models in Perception 8
      2.3 Generative Models in Tracking and Dynamics 9
   3 Why a Probability of Everything? 9
   4 Discriminative Learning 10
   5 Objective 12
   6 Scope and Organization 14
   7 Online Support 15

2. GENERATIVE VERSUS DISCRIMINATIVE LEARNING 17
   1 Two Schools of Thought 18
      1.1 Generative Probabilistic Models 19
      1.2 Discriminative Classifiers and Regressors 21
   2 Generative Learning 22
      2.1 Bayesian Inference 23
      2.2 Maximum Likelihood 24
      2.3 The Exponential Family 25
      2.4 Maximum Entropy 28
      2.5 Expectation Maximization and Mixtures 32

2.6 Graphical Models 36

3 Conditional Learning 42
3.1 Conditional Bayesian Inference 43
3.2 Maximum Conditional Likelihood 46
3.3 Logistic Regression 47

4 Discriminative Learning 48
4.1 Empirical Risk Minimization 48
4.2 Structural Risk Minimization 49
4.3 VC Dimension and Large Margins 50
4.4 Support Vector Machines 52
4.5 Kernel Methods 55

5 Averaged Classifiers 57

6 Joint Generative-Discriminative Learning 58

3. MAXIMUM ENTROPY DISCRIMINATION 61

1 Regularization Theory and Support Vector Machines 62
1.1 Solvability 64
1.2 Support Vector Machines and Kernels 65

2 A Distribution over Solutions 66

3 Augmented Distributions 69

4 Information and Geometry Interpretations 72

5 Computing the Partition Function 74

6 Margin Priors 75

7 Bias Priors 78
7.1 Gaussian Bias Priors 78
7.2 Non-Informative Bias Priors 78

8 Support Vector Machines 79
8.1 Single Axis SVM Optimization 80
8.2 Kernels 81

9 Generative Models 81
9.1 Exponential Family Models 82
9.2 Empirical Bayes Priors 84
9.3 Full Covariance Gaussians 86
9.4 Multinomials 91

10 Generalization Guarantees 93
10.1 VC Dimension 93
10.2 Sparsity 94
10.3 PAC-Bayes Bounds 95

4. EXTENSIONS TO MED 99
1 Multiclass Classification 100
2 Regression 102
2.1 SVM Regression 103
2.2 Generative Model Regression 105
3 Feature Selection and Structure Learning 105
3.1 Feature Selection in Classification 106
3.2 Feature Selection in Regression 110
3.3 Feature Selection in Generative Models 113
4 Kernel Selection 114
5 Meta-Learning 117
6 Transduction 120
6.1 Transductive Classification 121
6.2 Transductive Regression 125
7 Other Extensions 129
5. LATENT DISCRIMINATION 131
1 Mixture Models and Latent Variables 133
2 Iterative MED Projection 137
3 Bounding the Latent MED Constraints 138
4 Latent Decision Rules 143
5 Large Margin Mixtures of Gaussians 144
5.1 Parameter Distribution Update 145
5.2 Just a Support Vector Machine 148
5.3 Latent Distributions Update 149
5.4 Extension to Kernels 154
5.5 Extension to Non Gaussian Mixtures 155
6 Efficiency 155
6.1 Efficient Mixtures of Gaussians 160
7 Structured Latent Models 161
8 Factorization of Lagrange Multipliers 166
9 Mean Field for Intractable Models 168
6. CONCLUSION 171
1 A Generative and Discriminative Hybrid 172
2 Designing Models versus Designing Kernels 174
3 What's Next? 176

7. APPENDIX 179
1 Optimization in the MED Framework 179
1.1 Constrained Gradient Ascent 179
1.2 Axis-Parallel Optimization 181
1.3 Learning Axis Transitions 183
Index 199