

Pattern Recognition



Unsupervised



Statistical Theory



Feature
Selection



Linear Methods



Flexible
Discriminants



Neural
Networks



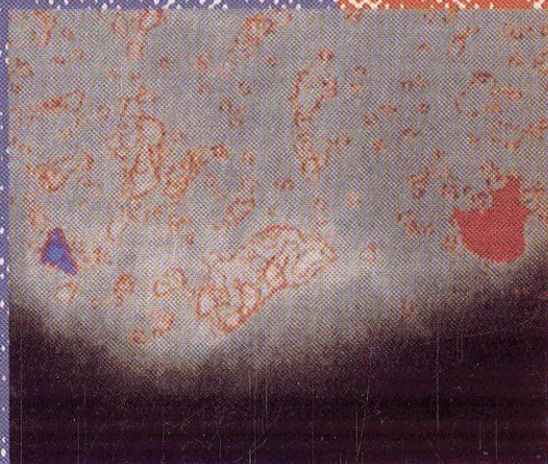
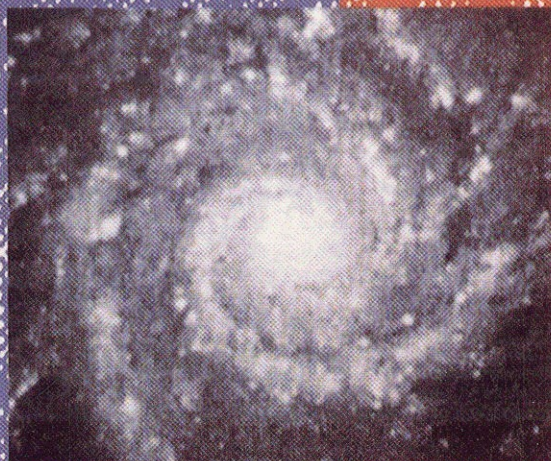
Non-
parametric
Methods



Belief Networks



Classification Trees



Pattern Recognition AND Neural Networks

B. D. Ripley

Contents

Preface	ix
Notation	xii
1 Introduction and Examples	1
1.1 How do neural methods differ?	4
1.2 The pattern recognition task	5
1.3 Overview of the remaining chapters	9
1.4 Examples	10
1.5 Literature	15
2 Statistical Decision Theory	17
2.1 Bayes rules for known distributions	18
2.2 Parametric models	26
2.3 Logistic discrimination	43
2.4 Predictive classification	45
2.5 Alternative estimation procedures	55
2.6 How complex a model do we need?	59
2.7 Performance assessment	66
2.8 Computational learning approaches	77
3 Linear Discriminant Analysis	91
3.1 Classical linear discrimination	92
3.2 Linear discriminants via regression	101
3.3 Robustness	105
3.4 Shrinkage methods	106

3.5	Logistic discrimination	109
3.6	Linear separation and perceptrons	116
4	Flexible Discriminants	121
4.1	Fitting smooth parametric functions	122
4.2	Radial basis functions	131
4.3	Regularization	136
5	Feed-forward Neural Networks	143
5.1	Biological motivation	145
5.2	Theory	147
5.3	Learning algorithms	148
5.4	Examples	160
5.5	Bayesian perspectives	163
5.6	Network complexity	168
5.7	Approximation results	173
6	Non-parametric Methods	181
6.1	Non-parametric estimation of class densities	181
6.2	Nearest neighbour methods	191
6.3	Learning vector quantization	201
6.4	Mixture representations	207
7	Tree-structured Classifiers	213
7.1	Splitting rules	216
7.2	Pruning rules	221
7.3	Missing values	231
7.4	Earlier approaches	235
7.5	Refinements	237
7.6	Relationships to neural networks	240
7.7	Bayesian trees	241
8	Belief Networks	243
8.1	Graphical models and networks	246
8.2	Causal networks	262
8.3	Learning the network structure	275

8.4	Boltzmann machines	279
8.5	Hierarchical mixtures of experts	283
9	Unsupervised Methods	287
9.1	Projection methods	288
9.2	Multidimensional scaling	305
9.3	Clustering algorithms	311
9.4	Self-organizing maps	322
10	Finding Good Pattern Features	327
10.1	Bounds for the Bayes error	328
10.2	Normal class distributions	329
10.3	Branch-and-bound techniques	330
10.4	Feature extraction	331
A	Statistical Sidelines	333
A.1	Maximum likelihood and MAP estimation	333
A.2	The EM algorithm	334
A.3	Markov chain Monte Carlo	337
A.4	Axioms for conditional independence	339
A.5	Optimization	342
	Glossary	347
	References	355
	Author Index	391
	Subject Index	399