

**A P R**

Advances in **P**attern **R**ecognition

Ian T. Nabney

**NETLAB**  
Algorithms for  
Pattern  
Recognition



Springer

**Software  
available  
on the web**

# Contents

<b>1. Introduction</b> .....	<b>1</b>
1.1 Introduction to MATLAB .....	2
1.1.1 MATLAB Basics .....	2
1.1.2 MATLAB Plotting .....	6
1.1.3 Programming in MATLAB .....	9
1.1.4 Programming Facilities .....	10
1.2 The NETLAB Toolbox .....	18
1.2.1 Overview of NETLAB .....	18
1.2.2 Generic Functions .....	21
1.3 Worked Example: Data Normalisation .....	29
Exercises .....	31
<b>2. Parameter Optimisation Algorithms</b> .....	<b>33</b>
2.1 Controlling the Algorithms .....	34
2.1.1 Information Display .....	35
2.1.2 Termination Criteria .....	36
2.1.3 Extra Arguments and Return Values .....	36
2.1.4 Demonstration Program .....	37
2.2 Quadratic Approximation at a Minimum .....	41
2.3 Line Search .....	42
2.3.1 Precision .....	42
2.3.2 Line Minimisation Algorithms .....	43
2.3.3 Bracketing the Minimum .....	48
2.4 Batch Gradient Descent .....	50
2.5 Conjugate Gradients .....	52
2.6 Scaled Conjugate Gradients .....	56
2.7 Quasi-Newton Methods .....	61
2.8 Optimisation and Neural Networks .....	66
2.8.1 General Purpose Optimisation Algorithms .....	66
2.8.2 On-line Gradient Descent .....	68
2.9 Worked Example: Constrained Optimisation .....	71
Exercises .....	75

<b>3. Density Modelling and Clustering</b> .....	<b>79</b>
3.1 Gaussian Mixture Models .....	79
3.1.1 Theory .....	79
3.1.2 NETLAB Implementation .....	84
3.2 Computing Probabilities .....	87
3.2.1 GMM Activations .....	87
3.2.2 GMM Probabilities .....	88
3.2.3 GMM Posteriors .....	89
3.3 EM Training Algorithm .....	89
3.3.1 Algorithm Theory .....	90
3.3.2 Initialisation .....	93
3.3.3 NETLAB EM Implementation .....	95
3.4 Demonstrations of GMM .....	97
3.4.1 EM Algorithm .....	97
3.4.2 Density Modelling .....	100
3.5 K-means Clustering .....	101
3.5.1 Algorithm and NETLAB Implementation .....	101
3.5.2 Demonstration Program .....	103
3.6 K-nearest-neighbour .....	104
3.6.1 Algorithm and NETLAB Implementation .....	105
3.6.2 Demonstration Program .....	107
3.7 Worked Examples .....	107
3.7.1 Classification with Density Models .....	107
3.7.2 Novelty Detection .....	110
Exercises .....	113
<b>4. Single Layer Networks</b> .....	<b>117</b>
4.1 The Single Layer Feed-forward Network .....	117
4.1.1 NETLAB Implementation .....	119
4.1.2 Forward Propagation .....	121
4.2 Error Functions .....	123
4.3 Error Gradient Calculation .....	127
4.4 Evaluating Other Derivatives .....	128
4.4.1 Network Activation Derivatives .....	128
4.4.2 The Hessian Matrix .....	129
4.5 Iterated Re-weighted Least Squares Training .....	132
4.6 Demonstration Programs .....	139
4.6.1 Two-class Problem .....	139
4.6.2 Three-class Problem .....	141
4.7 Worked Example: Training Regularised Models .....	142
Exercises .....	146

<b>5. The Multi-layer Perceptron</b> .....	149
5.1 The Two-layer Feed-forward Network .....	149
5.1.1 Definition .....	149
5.1.2 Network Creation and Initialisation .....	152
5.1.3 Manipulating Weights .....	154
5.1.4 Forward Propagation .....	155
5.2 Error Functions and Network Training .....	156
5.3 Error Gradient Calculation .....	156
5.4 Evaluating other Derivatives .....	159
5.5 The Hessian Matrix .....	160
5.5.1 Fast Multiplication by the Hessian .....	160
5.5.2 NETLAB Implementation .....	163
5.6 Demonstration Programs .....	163
5.6.1 Regression Demonstration .....	164
5.6.2 Classification Demonstration .....	165
5.7 Mixture Density Networks .....	167
5.7.1 Model Structure .....	169
5.7.2 Network Creation and Initialisation .....	172
5.7.3 Forward Propagation .....	174
5.7.4 Conditional Probabilities .....	175
5.7.5 Error and Gradient Calculation .....	179
5.7.6 Demonstration Program .....	181
5.8 Worked Example: Adding Direct Connections .....	184
Exercises .....	187
<b>6. Radial Basis Functions</b> .....	191
6.1 The RBF Network .....	192
6.1.1 Theory .....	192
6.1.2 NETLAB Implementation .....	195
6.2 Special Purpose Training Algorithms .....	199
6.2.1 Basis Function Optimisation .....	200
6.2.2 Output Weight Optimisation .....	201
6.2.3 NETLAB Implementation .....	202
6.3 Error and Error Gradient Calculation .....	203
6.4 Evaluating Other Derivatives .....	208
6.4.1 Output Derivatives .....	208
6.4.2 Network Jacobian .....	209
6.4.3 Network Hessian .....	211
6.5 Demonstration Program .....	213
6.6 Worked Examples: Linear Smoothing .....	214
6.6.1 Dual Basis Functions .....	214
6.6.2 Characterising Network Complexity .....	218
Exercises .....	221

<b>7. Visualisation and Latent Variable Models</b> .....	225
7.1 Principal Component Analysis .....	226
7.1.1 Theory .....	227
7.1.2 NETLAB Implementation .....	228
7.1.3 Examples .....	230
7.2 Probabilistic Principal Component Analysis .....	233
7.2.1 Probabilistic PCA .....	234
7.2.2 PPCA Implementation .....	236
7.2.3 Mixture of PPCA .....	237
7.2.4 Mixture of PPCA Implementation .....	239
7.3 Generative Topographic Mapping .....	243
7.3.1 The GTM Model .....	243
7.3.2 Model Creation and Initialisation .....	248
7.3.3 Computing Probabilities .....	249
7.3.4 EM Algorithm .....	250
7.3.5 Magnification Factors .....	256
7.3.6 Demonstration Programs .....	258
7.4 Topographic Projection .....	264
7.4.1 Neuroscale Algorithm .....	264
7.4.2 Neuroscale Implementation .....	268
7.4.3 Demonstration of Neuroscale .....	272
7.5 Worked Example: Canonical Variates .....	273
Exercises .....	277
<b>8. Sampling</b> .....	283
8.1 Monte Carlo Integration .....	284
8.2 Basic Sampling .....	285
8.2.1 Random Number Generators .....	286
8.2.2 Transformation Methods .....	287
8.2.3 Rejection Sampling .....	289
8.2.4 Importance Sampling .....	289
8.3 Markov Chain Sampling .....	290
8.3.1 Markov Chain Fundamentals .....	291
8.3.2 Gibbs Sampling .....	293
8.3.3 Metropolis-Hastings Algorithm .....	295
8.3.4 Hybrid Monte Carlo .....	300
8.4 Demonstration Programs .....	311
8.4.1 Metropolis-Hastings Sampling .....	312
8.4.2 Hybrid Monte Carlo Sampling .....	314
8.5 Worked Example: Convergence Diagnostics .....	317
Exercises .....	321

<b>9. Bayesian Techniques</b> .....	<b>325</b>
9.1 Principles of Bayesian Inference .....	327
9.2 Priors for Neural Networks .....	329
9.2.1 Theory of Priors .....	330
9.2.2 NETLAB Implementation of Priors .....	332
9.2.3 Demonstration of Gaussian Priors .....	335
9.2.4 Masks and Weight Manipulation .....	336
9.3 Computing Error and Gradient Functions .....	337
9.4 The Evidence Procedure .....	341
9.4.1 Theory .....	341
9.4.2 NETLAB Implementation .....	346
9.5 Predictions and Error Bars .....	350
9.5.1 Predictions for Regression .....	350
9.5.2 Predictions for Classification .....	351
9.5.3 NETLAB Implementation .....	353
9.6 Demonstrations of Evidence Procedure .....	354
9.6.1 The Evidence Procedure for Regression .....	354
9.6.2 The Evidence Procedure for Classification .....	356
9.6.3 Automatic Relevance Determination .....	359
9.7 Monte Carlo Methods .....	361
9.8 Demonstration of Hybrid Monte Carlo for MLPs .....	362
9.9 Worked Example: Improved Classification Approximation ...	364
Exercises .....	366
<b>10. Gaussian Processes</b> .....	<b>369</b>
10.1 Bayesian Regression .....	369
10.1.1 Weight Space View .....	370
10.1.2 Function Space View .....	372
10.2 Theory of Gaussian Processes .....	373
10.2.1 Model Definition .....	373
10.2.2 Learning Hyperparameters .....	377
10.3 NETLAB Implementation .....	378
10.3.1 Model Creation and Initialisation .....	379
10.3.2 Making Predictions .....	381
10.3.3 Model Training .....	384
10.4 Demonstration Programs .....	389
10.4.1 Regression Demonstration .....	389
10.4.2 ARD Demonstration .....	390
10.5 Worked Example: GPs for Classification .....	392
Exercises .....	396
<b>A. Linear Algebra and Matrices</b> .....	<b>399</b>
<b>B. Algorithm Error Analysis</b> .....	<b>403</b>

## XVIII Contents

<b>References</b> .....	<b>407</b>
<b>Function Index</b> .....	<b>413</b>
<b>Subject Index</b> .....	<b>416</b>