

Contents

Prologue	xv
1 Introduction	1
1.1 If Data Had Mass, the Earth Would Be a Black Hole	2
1.2 Learning	4
1.2.1 Machine Learning	5
1.3 Types of Machine Learning	6
1.4 Supervised Learning	7
1.4.1 Regression	8
1.4.2 Classification	9
1.5 The Brain and the Neuron	11
1.5.1 Hebb's Rule	12
1.5.2 McCulloch and Pitts Neurons	13
1.5.3 Limitations of the McCulloch and Pitts Neuronal Model	15
Further Reading	16
2 Linear Discriminants	17
2.1 Preliminaries	18
2.2 The Perceptron	19
2.2.1 The Learning Rate η	21
2.2.2 The Bias Input	22
2.2.3 The Perceptron Learning Algorithm	23
2.2.4 An Example of Perceptron Learning	24
2.2.5 Implementation	26
2.2.6 Testing the Network	31
2.3 Linear Separability	32
2.3.1 The Exclusive Or (XOR) Function	34
2.3.2 A Useful Insight	36
2.3.3 Another Example: The Pima Indian Dataset	37
2.4 Linear Regression	41
2.4.1 Linear Regression Examples	43
Further Reading	44
Practice Questions	45

3	The Multi-Layer Perceptron	47
3.1	Going Forwards	49
3.1.1	Biases	50
3.2	Going Backwards: Back-Propagation of Error	50
3.2.1	The Multi-Layer Perceptron Algorithm	54
3.2.2	Initialising the Weights	57
3.2.3	Different Output Activation Functions	58
3.2.4	Sequential and Batch Training	59
3.2.5	Local Minima	60
3.2.6	Picking Up Momentum	61
3.2.7	Other Improvements	62
3.3	The Multi-Layer Perceptron in Practice	63
3.3.1	Data Preparation	63
3.3.2	Amount of Training Data	63
3.3.3	Number of Hidden Layers	64
3.3.4	Generalisation and Overfitting	66
3.3.5	Training, Testing, and Validation	66
3.3.6	When to Stop Learning	68
3.3.7	Computing and Evaluating the Results	69
3.4	Examples of Using the MLP	70
3.4.1	A Regression Problem	70
3.4.2	Classification with the MLP	74
3.4.3	A Classification Example	75
3.4.4	Time-Series Prediction	77
3.4.5	Data Compression: The Auto-Associative Network	80
3.5	Overview	83
3.6	Deriving Back-Propagation	84
3.6.1	The Network Output and the Error	84
3.6.2	The Error of the Network	85
3.6.3	A Suitable Activation Function	87
3.6.4	Back-Propagation of Error	88
	Further Reading	90
	Practice Questions	91
4	Radial Basis Functions and Splines	95
4.1	Concepts	95
4.1.1	Weight Space	95
4.1.2	Receptive Fields	97
4.2	The Radial Basis Function (RBF) Network	100
4.2.1	Training the RBF Network	103
4.3	The Curse of Dimensionality	106
4.4	Interpolation and Basis Functions	108
4.4.1	Bases and Basis Expansion	108
4.4.2	The Cubic Spline	112
4.4.3	Fitting the Spline to the Data	112

4.4.4	Smoothing Splines	113
4.4.5	Higher Dimensions	114
4.4.6	Beyond the Bounds	116
	Further Reading	116
	Practice Questions	117
5	Support Vector Machines	119
5.1	Optimal Separation	120
5.2	Kernels	125
5.2.1	Example: XOR	128
5.2.2	Extensions to the Support Vector Machine	128
	Further Reading	130
	Practice Questions	131
6	Learning with Trees	133
6.1	Using Decision Trees	133
6.2	Constructing Decision Trees	134
6.2.1	Quick Aside: Entropy in Information Theory	135
6.2.2	ID3	136
6.2.3	Implementing Trees and Graphs in Python	139
6.2.4	Implementation of the Decision Tree	140
6.2.5	Dealing with Continuous Variables	143
6.2.6	Computational Complexity	143
6.3	Classification and Regression Trees (CART)	145
6.3.1	Gini Impurity	146
6.3.2	Regression in Trees	147
6.4	Classification Example	147
	Further Reading	150
	Practice Questions	151
7	Decision by Committee: Ensemble Learning	153
7.1	Boosting	154
7.1.1	AdaBoost	155
7.1.2	Stumping	160
7.2	Bagging	160
7.2.1	Subagging	162
7.3	Different Ways to Combine Classifiers	162
	Further Reading	164
	Practice Questions	165
8	Probability and Learning	167
8.1	Turning Data into Probabilities	167
8.1.1	Minimising Risk	171
8.1.2	The Naïve Bayes' Classifier	171
8.2	Some Basic Statistics	173

8.2.1	Averages	173
8.2.2	Variance and Covariance	174
8.2.3	The Gaussian	176
8.2.4	The Bias-Variance Tradeoff	177
8.3	Gaussian Mixture Models	178
8.3.1	The Expectation-Maximisation (EM) Algorithm	179
8.4	Nearest Neighbour Methods	183
8.4.1	Nearest Neighbour Smoothing	185
8.4.2	Efficient Distance Computations: the KD-Tree	186
8.4.3	Distance Measures	190
	Further Reading	192
	Practice Questions	193
9	Unsupervised Learning	195
9.1	The k -Means Algorithm	196
9.1.1	Dealing with Noise	200
9.1.2	The k -Means Neural Network	200
9.1.3	Normalisation	202
9.1.4	A Better Weight Update Rule	203
9.1.5	Example: The Iris Dataset Again	204
9.1.6	Using Competitive Learning for Clustering	205
9.2	Vector Quantisation	206
9.3	The Self-Organising Feature Map	207
9.3.1	The SOM Algorithm	210
9.3.2	Neighbourhood Connections	211
9.3.3	Self-Organisation	214
9.3.4	Network Dimensionality and Boundary Conditions	214
9.3.5	Examples of Using the SOM	215
	Further Reading	218
	Practice Questions	220
10	Dimensionality Reduction	221
10.1	Linear Discriminant Analysis (LDA)	223
10.2	Principal Components Analysis (PCA)	226
10.2.1	Relation with the Multi-Layer Perceptron	231
10.2.2	Kernel PCA	232
10.3	Factor Analysis	234
10.4	Independent Components Analysis (ICA)	237
10.5	Locally Linear Embedding	239
10.6	Isomap	242
10.6.1	Multi-Dimensional Scaling (MDS)	242
	Further Reading	245
	Practice Questions	246

11 Optimisation and Search	247
11.1 Going Downhill	248
11.2 Least-Squares Optimisation	251
11.2.1 Taylor Expansion	251
11.2.2 The Levenberg-Marquardt Algorithm	252
11.3 Conjugate Gradients	257
11.3.1 Conjugate Gradients Example	260
11.4 Search: Three Basic Approaches	261
11.4.1 Exhaustive Search	261
11.4.2 Greedy Search	262
11.4.3 Hill Climbing	262
11.5 Exploitation and Exploration	264
11.6 Simulated Annealing	265
11.6.1 Comparison	266
Further Reading	267
Practice Questions	267
12 Evolutionary Learning	269
12.1 The Genetic Algorithm (GA)	270
12.1.1 String Representation	271
12.1.2 Evaluating Fitness	272
12.1.3 Population	273
12.1.4 Generating Offspring: Parent Selection	273
12.2 Generating Offspring: Genetic Operators	275
12.2.1 Crossover	275
12.2.2 Mutation	277
12.2.3 Elitism, Tournaments, and Niching	277
12.3 Using Genetic Algorithms	279
12.3.1 Map Colouring	279
12.3.2 Punctuated Equilibrium	281
12.3.3 Example: The Knapsack Problem	281
12.3.4 Example: The Four Peaks Problem	282
12.3.5 Limitations of the GA	284
12.3.6 Training Neural Networks with Genetic Algorithms	285
12.4 Genetic Programming	285
12.5 Combining Sampling with Evolutionary Learning	286
Further Reading	289
Practice Questions	290
13 Reinforcement Learning	293
13.1 Overview	294
13.2 Example: Getting Lost	296
13.2.1 State and Action Spaces	298
13.2.2 Carrots and Sticks: the Reward Function	299
13.2.3 Discounting	300

13.2.4	Action Selection	301
13.2.5	Policy	302
13.3	Markov Decision Processes	302
13.3.1	The Markov Property	302
13.3.2	Probabilities in Markov Decision Processes	303
13.4	Values	305
13.5	Back on Holiday: Using Reinforcement Learning	309
13.6	The Difference between Sarsa and Q-Learning	310
13.7	Uses of Reinforcement Learning	311
	Further Reading	312
	Practice Questions	312
14	Markov Chain Monte Carlo (MCMC) Methods	315
14.1	Sampling	315
14.1.1	Random Numbers	316
14.1.2	Gaussian Random Numbers	317
14.2	Monte Carlo or Bust	319
14.3	The Proposal Distribution	320
14.4	Markov Chain Monte Carlo	325
14.4.1	Markov Chains	325
14.4.2	The Metropolis-Hastings Algorithm	326
14.4.3	Simulated Annealing (Again)	327
14.4.4	Gibbs Sampling	328
	Further Reading	331
	Practice Questions	332
15	Graphical Models	333
15.1	Bayesian Networks	335
15.1.1	Example: Exam Panic	335
15.1.2	Approximate Inference	339
15.1.3	Making Bayesian Networks	342
15.2	Markov Random Fields	344
15.3	Hidden Markov Models (HMMs)	347
15.3.1	The Forward Algorithm	349
15.3.2	The Viterbi Algorithm	352
15.3.3	The Baum-Welch or Forward-Backward Algorithm	353
15.4	Tracking Methods	356
15.4.1	The Kalman Filter	357
15.4.2	The Particle Filter	360
	Further Reading	361
	Practice Questions	362

16 Python	365
16.1 Installing Python and Other Packages	365
16.2 Getting Started	365
16.2.1 Python for MATLAB and R users	370
16.3 Code Basics	370
16.3.1 Writing and Importing Code	370
16.3.2 Control Flow	371
16.3.3 Functions	372
16.3.4 The doc String	373
16.3.5 map and lambda	373
16.3.6 Exceptions	374
16.3.7 Classes	374
16.4 Using NumPy and Matplotlib	375
16.4.1 Arrays	375
16.4.2 Random Numbers	379
16.4.3 Linear Algebra	379
16.4.4 Plotting	380
Further Reading	381
Practice Questions	382
 Index	 383