

Contents

<i>Preface</i>	<i>page ix</i>
1 Getting Started	1
1.1 Artificial neural networks	1
1.2 A simple example	4
1.3 General setup	8
1.4 Problems	13
2 Perceptron Learning – Basics	14
2.1 Gibbs learning	14
2.2 The annealed approximation	18
2.3 The Gardner analysis	22
2.4 Summary	27
2.5 Problems	29
3 A Choice of Learning Rules	33
3.1 The Hebb rule	33
3.2 The perceptron rule	36
3.3 The pseudo-inverse rule	37
3.4 The adaline rule	39
3.5 Maximal stability	40
3.6 The Bayes rule	42
3.7 Summary	46
3.8 Problems	46
4 Augmented Statistical Mechanics Formulation	49
4.1 Maximal stabilities	49
4.2 Gibbs learning at non-zero temperature	52
4.3 General statistical mechanics formulation	56
4.4 Learning rules revisited	59
4.5 The optimal potential	63
4.6 Summary	64

vi	<i>Contents</i>	
	4.7 Problems	65
5	Noisy Teachers	69
	5.1 Sources of noise	69
	5.2 Trying perfect learning	72
	5.3 Learning with errors	78
	5.4 Refinements	80
	5.5 Summary	82
	5.6 Problems	83
6	The Storage Problem	85
	6.1 The storage capacity	85
	6.2 Counting dichotomies: the Cover analysis	89
	6.3 Galilean pastiche: the Ising perceptron	93
	6.4 The distribution of stabilities	98
	6.5 Beyond the storage capacity	102
	6.6 Problems	104
7	Discontinuous Learning	109
	7.1 Smooth networks	109
	7.2 The Ising perceptron	111
	7.3 The reversed wedge perceptron	114
	7.4 The dynamics of discontinuous learning	118
	7.5 Summary	121
	7.6 Problems	122
8	Unsupervised Learning	125
	8.1 Supervised versus unsupervised learning	125
	8.2 The deceptions of randomness	129
	8.3 Learning a symmetry-breaking direction	133
	8.4 Clustering through competitive learning	137
	8.5 Clustering by tuning the temperature	142
	8.6 Discussion	145
	8.7 Problems	147
9	On-line Learning	149
	9.1 Stochastic gradient descent	149
	9.2 Specific examples	152
	9.3 Optimal on-line learning	155
	9.4 Perceptron with a smooth transfer function	159
	9.5 Queries	160
	9.6 Unsupervised on-line learning	165
	9.7 The natural gradient	169
	9.8 Discussion	170
	9.9 Problems	171

	<i>Contents</i>	vii
10	Making Contact with Statistics	176
10.1	Convergence of frequencies to probabilities	176
10.2	Sauer’s lemma	178
10.3	The Vapnik–Chervonenkis theorem	180
10.4	Comparison with statistical mechanics	182
10.5	The Cramér–Rao inequality	186
10.6	Discussion	189
10.7	Problems	190
11	A Bird’s Eye View: Multifractals	193
11.1	The shattered coupling space	193
11.2	The multifractal spectrum of the perceptron	195
11.3	The multifractal organization of internal representations	203
11.4	Discussion	207
11.5	Problems	207
12	Multilayer Networks	209
12.1	Basic architectures	210
12.2	Bounds	214
12.3	The storage problem	218
12.4	Generalization with a parity tree	222
12.5	Generalization with a committee tree	225
12.6	The fully connected committee machine	228
12.7	Summary	230
12.8	Problems	232
13	On-line Learning in Multilayer Networks	237
13.1	The committee tree	237
13.2	The parity tree	243
13.3	Soft committee machine	246
13.4	Back-propagation	251
13.5	Bayesian on-line learning	253
13.6	Discussion	255
13.7	Problems	256
14	What Else?	259
14.1	Support vector machines	259
14.2	Complex optimization	263
14.3	Error-correcting codes	266
14.4	Game theory	270
	Appendices	275
A1	Basic Mathematics	275
A2	The Gardner Analysis	282
A3	Convergence of the Perceptron Rule	289

viii	<i>Contents</i>	
A4	Stability of the Replica Symmetric Saddle Point	291
A5	One-step Replica Symmetry Breaking	300
A6	The Cavity Approach	304
A7	The VC theorem	310
	<i>Bibliography</i>	313
	<i>Index</i>	327