

Contents

Part I	Introduction to neural networks	1
<hr/>		
1	General introduction	3
<hr/>		
1.1	Principles of neural information processing	3
1.2	Biological neurons and model neurons	7
1.3	Universality of McCulloch–Pitts neurons	21
1.4	Exercises	25
<hr/>		
2	Layered networks	29
<hr/>		
2.1	Linear separability	29
2.2	Multilayer networks	33
2.3	The perceptron	36
2.4	Learning in layered networks: error backpropagation	46
2.5	Learning dynamics in small learning rate perceptrons	52
2.6	Numerical simulations	58
2.7	Exercises	64
<hr/>		
3	Recurrent networks with binary neurons	69
<hr/>		
3.1	Noiseless recurrent networks	70
3.2	Synaptic symmetry and Lyapunov functions	77
3.3	Information processing in recurrent networks	81
3.4	Exercises	87
<hr/>		
4	Notes and suggestions for further reading	89
<hr/>		
Part II	Advanced neural networks	93
<hr/>		
5	Competitive unsupervised learning processes	95
<hr/>		
5.1	Vector quantization	95
5.2	Soft vector quantization	98

5.3	Time-dependent learning rates	110
5.4	Self-organizing maps	114
5.5	Exercises	122
<hr/>		
6	Bayesian techniques in supervised learning	127
<hr/>		
6.1	Preliminaries and introduction	127
6.2	Bayesian learning of network weights	136
6.3	Predictions with error bars: real-valued functions	144
6.4	Predictions with error bars: binary classification	152
6.5	Bayesian model selection	156
6.6	Practicalities: measuring curvature	163
6.7	Exercises	164
<hr/>		
7	Gaussian processes	169
<hr/>		
7.1	The underlying idea	169
7.2	Examples of networks reducing to Gaussian processes	172
7.3	The 'priors over functions' point of view	176
7.4	Stationary covariance functions	177
7.5	Learning and prediction with Gaussian processes	180
7.6	Exercises	182
<hr/>		
8	Support vector machines for binary classification	185
<hr/>		
8.1	Optimal separating plane for linearly separable tasks	185
8.2	Representation in terms of support vectors	189
8.3	Preprocessing and SVM kernels	197
8.4	Exercises	202
<hr/>		
9	Notes and suggestions for further reading	205
<hr/>		
Part III	Information theory and neural networks	207
<hr/>		
10	Measuring information	209
<hr/>		
10.1	Brute force: counting messages	209
10.2	Exploiting message likelihood differences via coding	212
10.3	Proposal for a measure of information	218

11 Identification of entropy as an information measure	223
11.1 Coding theory and the Kraft inequality	223
11.2 Entropy and optimal coding	229
11.3 Shannon's original proof	233
12 Building blocks of Shannon's information theory	235
12.1 Entropy	235
12.2 Joint and conditional entropy	240
12.3 Relative entropy and mutual information	245
12.4 Information measures for continuous random variables	251
12.5 Exercises	258
13 Information theory and statistical inference	261
13.1 Maximum likelihood estimation	261
13.2 The maximum entropy principle	264
13.3 Exercises	270
14 Applications to neural networks	273
14.1 Supervised learning: Boltzmann machines	273
14.2 Maximum information preservation	281
14.3 Neuronal specialization	285
14.4 Detection of coherent features	294
14.5 The effect of non-linearities	297
14.6 Introduction to Amari's information geometry	299
14.7 Simple applications of information geometry	306
14.8 Exercises	311
15 Notes and suggestions for further reading	313
Part IV Macroscopic analysis of dynamics	315
16 Network operation: macroscopic dynamics	317
16.1 Microscopic dynamics in probabilistic form	318
16.2 Sequential dynamics	324
16.3 Parallel dynamics	338
16.4 Exercises	345

17 Dynamics of online learning in binary perceptrons	349
17.1 Probabilistic definitions, performance measures	349
17.2 Explicit learning rules	353
17.3 Optimized learning rules	368
17.4 Exercises	382
18 Dynamics of online gradient descent learning	385
18.1 Online gradient descent	385
18.2 Learning from noisy examples	392
18.3 Exercises	394
19 Notes and suggestions for further reading	397
Part V Equilibrium statistical mechanics of neural networks	399
20 Basics of equilibrium statistical mechanics	401
20.1 Stationary distributions and ergodicity	401
20.2 Detailed balance and interaction symmetry	408
20.3 Equilibrium statistical mechanics: concepts, definitions	413
20.4 A simple example: storing a single pattern	419
20.5 Phase transitions and ergodicity breaking	425
20.6 Exercises	430
21 Network operation: equilibrium analysis	437
21.1 Hopfield model with finite number of patterns	438
21.2 Introduction to replica theory: the SK model	447
21.3 Hopfield model with an extensive number of patterns	460
21.4 Exercises	480
22 Gardner theory of task realizability	489
22.1 The space of interactions	489
22.2 Capacity of perceptrons—definition and toy example	494
22.3 Capacity of perceptrons—random inputs	498
23 Notes and suggestions for further reading	507

Appendix A: Probability theory in a nutshell	511
A.1 Discrete event sets	511
A.2 Continuous event sets	513
A.3 Averages of specific random variables	515

Appendix B: Conditions for the central limit theorem to apply	517
B.1 Moment condition	517
B.2 Lindeberg's theorem	518

Appendix C: Some simple summation identities	521
---	------------

Appendix D: Gaussian integrals and probability distributions	523
D.1 General properties of Gaussian integrals	523
D.2 Gaussian probability distributions	527
D.3 A list of specific Gaussian integrals	531

Appendix E: Matrix identities	537
E.1 Block inverses	537
E.2 The Woodbury formula	538

Appendix F: The δ-distribution	539
F.1 Definition	539
F.2 $\delta(x)$ as solution of the Liouville equation	540
F.3 Representations, relations, and generalizations	541

Appendix G: Inequalities based on convexity	543
--	------------

Appendix H: Metrics for parametrized probability distributions	549
H.1 Local distance definitions	549
H.2 The triangular inequality	550
H.3 Global distance definitions	551

Appendix I: Saddle-point integration	553
<hr/>	
References	555
<hr/>	
Index	563
