

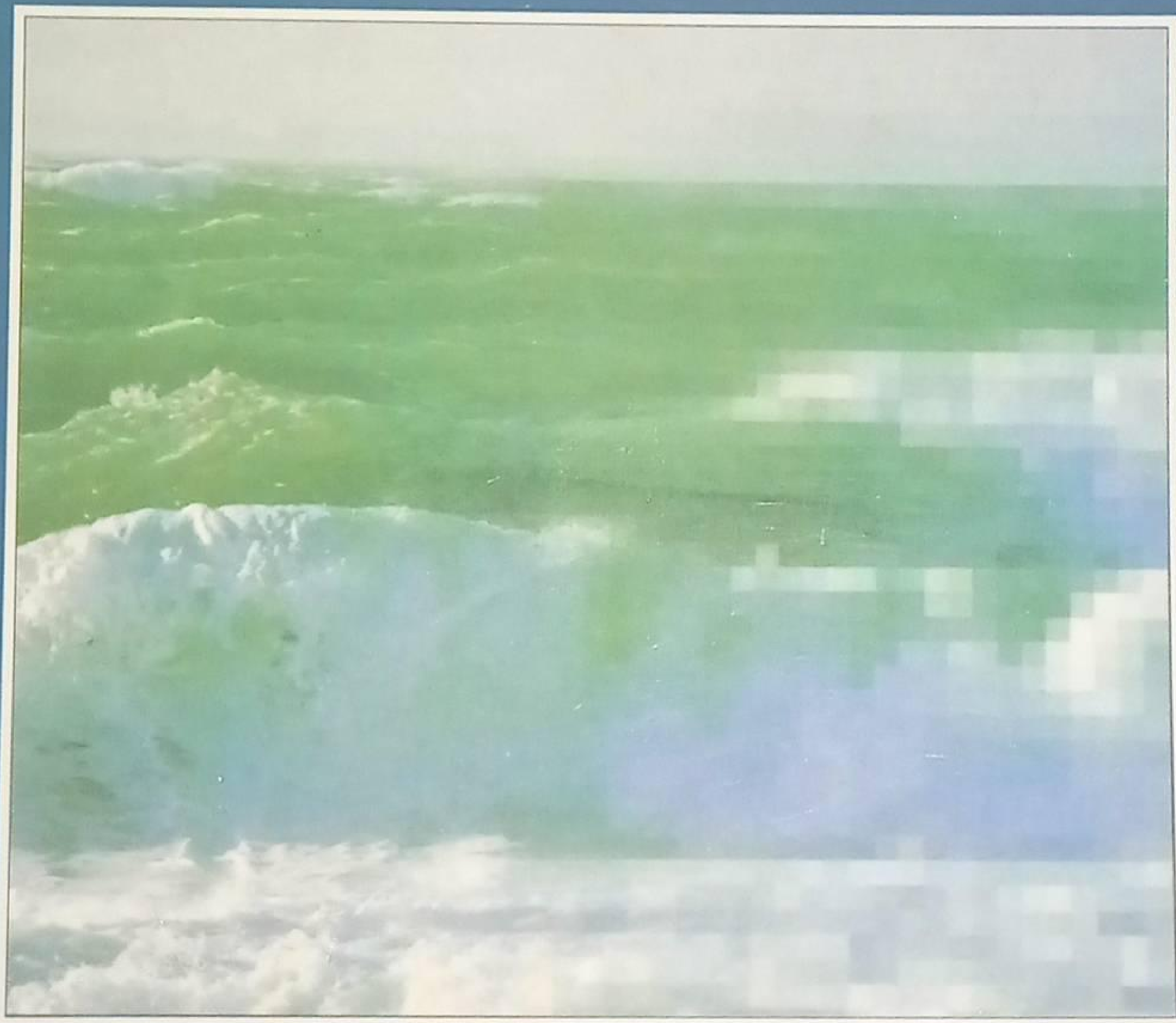
LOW PRICE EDITION

PEARSON  
Education

SECOND EDITION

# NEURAL NETWORKS

A COMPREHENSIVE FOUNDATION



SIMON HAYKIN

# Contents

Preface xii

Acknowledgments xv

Abbreviations and Symbols xvii

**1 Introduction 1**

---

- 1.1 What Is a Neural Network? 1
- 1.2 Human Brain 6
- 1.3 Models of a Neuron 10
- 1.4 Neural Networks Viewed as Directed Graphs 15
- 1.5 Feedback 18
- 1.6 Network Architectures 21
- 1.7 Knowledge Representation 23
- 1.8 Artificial Intelligence and Neural Networks 34
- 1.9 Historical Notes 38
- Notes and References 45
- Problems 45

**2 Learning Processes 50**

---

- 2.1 Introduction 50
- 2.2 Error-Correction Learning 51
- 2.3 Memory-Based Learning 53
- 2.4 Hebbian Learning 55
- 2.5 Competitive Learning 58
- 2.6 Boltzmann Learning 60

2.7	Credit Assignment Problem	62
2.8	Learning with a Teacher	63
2.9	Learning without a Teacher	64
2.10	Learning Tasks	66
2.11	Memory	75
2.12	Adaptation	83
2.13	Statistical Nature of the Learning Process	84
2.14	Statistical Learning Theory	89
2.15	Probably Approximately Correct Model of Learning	102
2.16	Summary and Discussion	105
	Notes and References	106
	Problems	111

### **3 Single Layer Perceptrons**

117

3.1	Introduction	117
3.2	Adaptive Filtering Problem	118
3.3	Unconstrained Optimization Techniques	121
3.4	Linear Least-Squares Filters	126
3.5	Least-Mean-Square Algorithm	128
3.6	Learning Curves	133
3.7	Learning Rate Annealing Techniques	134
3.8	Perceptron	135
3.9	Perceptron Convergence Theorem	137
3.10	Relation Between the Perceptron and Bayes Classifier for a Gaussian Environment	143
3.11	Summary and Discussion	148
	Notes and References	150
	Problems	151

### **4 Multilayer Perceptrons**

156

4.1	Introduction	156
4.2	Some Preliminaries	159
4.3	Back-Propagation Algorithm	161
4.4	Summary of the Back-Propagation Algorithm	173
4.5	XOR Problem	175
4.6	Heuristics for Making the Back-Propagation Algorithm Perform Better	178
4.7	Output Representation and Decision Rule	184
4.8	Computer Experiment	187
4.9	Feature Detection	199
4.10	Back-Propagation and Differentiation	202
4.11	Hessian Matrix	204
4.12	Generalization	205

4.13	Approximations of Functions	208
4.14	Cross-Validation	213
4.15	Network Pruning Techniques	218
4.16	Virtues and Limitations of Back-Propagation Learning	226
4.17	Accelerated Convergence of Back-Propagation Learning	233
4.18	Supervised Learning Viewed as an Optimization Problem	234
4.19	Convolutional Networks	245
4.20	Summary and Discussion	247
	Notes and References	248
	Problems	252

## **5 Radial-Basis Function Networks** **256**

---

5.1	Introduction	256
5.2	Cover's Theorem on the Separability of Patterns	257
5.3	Interpolation Problem	262
5.4	Supervised Learning as an Ill-Posed Hypersurface Reconstruction Problem	265
5.5	Regularization Theory	267
5.6	Regularization Networks	277
5.7	Generalized Radial-Basis Function Networks	278
5.8	XOR Problem (Revisited)	282
5.9	Estimation of the Regularization Parameter	284
5.10	Approximation Properties of RBF Networks	290
5.11	Comparison of RBF Networks and Multilayer Perceptrons	293
5.12	Kernel Regression and Its Relation to RBF Networks	294
5.13	Learning Strategies	298
5.14	Computer Experiment	305
5.15	Summary and Discussion	308
	Notes and References	308
	Problems	312

## **6 Support Vector Machines** **318**

---

6.1	Introduction	318
6.2	Optimal Hyperplane for Linearly Separable Patterns	319
6.3	Optimal Hyperplane for Nonseparable Patterns	326
6.4	How to Build a Support Vector Machine for Pattern Recognition	329
6.5	Example: XOR Problem (Revisited)	335
6.6	Computer Experiment	337
6.7	$\epsilon$ -Insensitive Loss Function	339
6.8	Support Vector Machines for Nonlinear Regression	340
6.9	Summary and Discussion	343
	Notes and References	347
	Problems	348

<b>7</b>	<b>Committee Machines</b>	<b>351</b>
7.1	Introduction	351
7.2	Ensemble Averaging	353
7.3	Computer Experiment I	355
7.4	Boosting	357
7.5	Computer Experiment II	364
7.6	Associative Gaussian Mixture Model	366
7.7	Hierarchical Mixture of Experts Model	372
7.8	Model Selection Using a Standard Decision Tree	374
7.9	A Priori and a Posteriori Probabilities	377
7.10	Maximum Likelihood Estimation	378
7.11	Learning Strategies for the HME Model	380
7.12	EM Algorithm	382
7.13	Application of the EM Algorithm to the HME Model	383
7.14	Summary and Discussion	386
	Notes and References	387
	Problems	389
<b>8</b>	<b>Principal Components Analysis</b>	<b>392</b>
8.1	Introduction	392
8.2	Some Intuitive Principles of Self-Organization	393
8.3	Principal Components Analysis	396
8.4	Hebbian-Based Maximum Eigenfilter	404
8.5	Hebbian-Based Principal Components Analysis	413
8.6	Computer Experiment: Image Coding	419
8.7	Adaptive Principal Components Analysis Using Lateral Inhibition	422
8.8	Two Classes of PCA Algorithms	430
8.9	Batch and Adaptive Methods of Computation	430
8.10	Kernel-Based Principal Components Analysis	432
8.11	Summary And Discussion	437
	Notes And References	439
	Problems	440
<b>9</b>	<b>Self-Organizing Maps</b>	<b>443</b>
9.1	Introduction	443
9.2	Two Basic Feature-Mapping Models	444
9.3	Self-Organizing Map	446
9.4	Summary of the SOM Algorithm	453
9.5	Properties of the Feature Map	454
9.6	Computer Simulations	461
9.7	Learning Vector Quantization	466
9.8	Computer Experiment: Adaptive Pattern Classification	468

- 9.9 Hierarchical Vector Quantization 470
- 9.10 Contextual Maps 474
- 9.11 Summary and Discussion 476
  - Notes and References 477
  - Problems 479

## **10 Information-Theoretic Models 484**

---

- 10.1 Introduction 484
- 10.2 Entropy 485
- 10.3 Maximum Entropy Principle 490
- 10.4 Mutual Information 492
- 10.5 Kullback–Leibler Divergence 495
- 10.6 Mutual Information as an Objective Function To Be Optimized 498
- 10.7 Maximum Mutual Information Principle 499
- 10.8 Infomax and Redundancy Reduction 503
- 10.9 Spatially Coherent Features 506
- 10.10 Spatially Incoherent Features 508
- 10.11 Independent Components Analysis 510
- 10.12 Computer Experiment 523
- 10.13 Maximum Likelihood Estimation 525
- 10.14 Maximum Entropy Method 529
- 10.15 Summary and Discussion 533
  - Notes and References 535
  - Problems 541

## **11 Stochastic Machines And Their Approximates Rooted In Statistical Mechanics 545**

---

- 11.1 Introduction 545
- 11.2 Statistical Mechanics 546
- 11.3 Markov Chains 548
- 11.4 Metropolis Algorithm 556
- 11.5 Simulated Annealing 558
- 11.6 Gibbs Sampling 561
- 11.7 Boltzmann Machine 562
- 11.8 Sigmoid Belief Networks 569
- 11.9 Helmholtz Machine 574
- 11.10 Mean-Field Theory 576
- 11.11 Deterministic Boltzmann Machine 578
- 11.12 Deterministic Sigmoid Belief Networks 579
- 11.13 Deterministic Annealing 586
- 11.14 Summary and Discussion 592
  - Notes and References 594
  - Problems 597

## 12 Neurodynamic Programming 603

---

- 12.1 Introduction 603
- 12.2 Markovian Decision Processes 604
- 12.3 Bellman's Optimality Criterion 607
- 12.4 Policy Iteration 610
- 12.5 Value Iteration 612
- 12.6 Neurodynamic Programming 617
- 12.7 Approximate Policy Iteration 618
- 12.8 Q-Learning 622
- 12.9 Computer Experiment 627
- 12.10 Summary and Discussion 629
- Notes and References 631
- Problems 632

## 13 Temporal Processing Using Feedforward Networks 635

---

- 13.1 Introduction 635
- 13.2 Short-term Memory Structures 636
- 13.3 Network Architectures for Temporal Processing 640
- 13.4 Focused Time Lagged Feedforward Networks 643
- 13.5 Computer Experiment 645
- 13.6 Universal Myopic Mapping Theorem 646
- 13.7 Spatio-Temporal Models of a Neuron 648
- 13.8 Distributed Time Lagged Feedforward Networks 651
- 13.9 Temporal Back-Propagation Algorithm 652
- 13.10 Summary and Discussion 659
- Notes and References 660
- Problems 660

## 14 Neurodynamics 664

---

- 14.1 Introduction 664
- 14.2 Dynamical Systems 666
- 14.3 Stability of Equilibrium States 669
- 14.4 Attractors 674
- 14.5 Neurodynamical Models 676
- 14.6 Manipulation of Attractors as a Recurrent Network Paradigm 680
- 14.7 Hopfield Models 680
- 14.8 Computer Experiment I 696
- 14.9 Cohen-Grossberg Theorem 701
- 14.10 Brain-State-in-a-Box Model 703
- 14.11 Computer Experiment II 709
- 14.12 Strange Attractors and Chaos 709
- 14.13 Dynamic Reconstruction of a Chaotic Process 714

- 14.14 Computer Experiment III 718
- 14.15 Summary and Discussion 722
- Notes and References 725
- Problems 727

## **15 Dynamically Driven Recurrent Networks 732**

---

- 15.1 Introduction 732
- 15.2 Recurrent Network Architectures 733
- 15.3 State-Space Model 739
- 15.4 Nonlinear Autoregressive with Exogenous Inputs Model 746
- 15.5 Computational Power of Recurrent Networks 747
- 15.6 Learning Algorithms 750
- 15.7 Back-Propagation Through Time 751
- 15.8 Real-Time Recurrent Learning 756
- 15.9 Kalman Filters 762
- 15.10 Decoupled Extended Kalman Filters 765
- 15.11 Computer Experiment 770
- 15.12 Vanishing Gradients in Recurrent Networks 773
- 15.13 System Identification 776
- 15.14 Model-Reference Adaptive Control 780
- 15.15 Summary and Discussion 782
- Notes and References 783
- Problems 785

## **Epilogue 790**

## **Bibliography 796**

## **Index 837**