Contents

Preface 12

Acknowledgments 15

Abbreviations and Symbols 17

1 Introduction 23

1.1 What Is a Neural Network? 23
1.2 Human Brain 28
1.3 Models of a Neuron 32
1.4 Neural Networks Viewed as Directed Graphs 37
1.5 Feedback 40
1.6 Network Architectures 43
1.7 Knowledge Representation 45
1.8 Artificial Intelligence and Neural Networks 56
1.9 Historical Notes 60
Notes and References 67
Problems 67

2 Learning Processes 72

2.1 Introduction 72
2.2 Error-Correction Learning 73
2.3 Memory-Based Learning 75
2.4 Hebbian Learning 77
2.5 Competitive Learning 80
2.6 Boltzmann Learning 82

2.7 Credit Assignment Problem 84
2.8 Learning with a Teacher 85
2.9 Learning without a Teacher 86
2.10 Learning Tasks 88
2.11 Memory 97
2.12 Adaptation 105
2.13 Statistical Nature of the Learning Process 106
2.14 Statistical Learning Theory 111
2.15 Probably Approximately Correct Model of Learning 124
2.16 Summary and Discussion 127
Notes and References 128
Problems 133

3 Single Layer Perceptrons 139

3.1 Introduction 139
3.2 Adaptive Filtering Problem 140
3.3 Unconstrained Optimization Techniques 143
3.4 Linear Least-Squares Filters 148
3.5 Least-Mean-Square Algorithm 150
3.6 Learning Curves 155
3.7 Learning Rate Annealing Techniques 156
3.8 Perceptron 157
3.9 Perceptron Convergence Theorem 159
3.10 Relation Between the Perceptron and Bayes Classifier for a Gaussian Environment 165
3.11 Summary and Discussion 170
Notes and References 172
Problems 173

4 Multilayer Perceptrons 178

4.1 Introduction 178
4.2 Some Preliminaries 181
4.3 Back-Propagation Algorithm 183
4.4 Summary of the Back-Propagation Algorithm 195
4.5 XOR Problem 197
4.6 Heuristics for Making the Back-Propagation Algorithm Perform Better 200
4.7 Output Representation and Decision Rule 206
4.8 Computer Experiment 209
4.9 Feature Detection 221
4.10 Back-Propagation and Differentiation 224
4.11 Hessian Matrix 226
4.12 Generalization 227

4.13 Approximations of Functions 230
4.14 Cross-Validation 235
4.15 Network Pruning Techniques 240
4.16 Virtues and Limitations of Back-Propagation Learning 248
4.17 Accelerated Convergence of Back-Propagation Learning 255
4.18 Supervised Learning Viewed as an Optimization Problem 256
4.19 Convolutional Networks 267
4.20 Summary and Discussion 269
Notes and References 270
Problems 274

5 Radial-Basis Function Networks **278**

5.1 Introduction 278
5.2 Cover's Theorem on the Separability of Patterns 279
5.3 Interpolation Problem 284
5.4 Supervised Learning as an Ill-Posed Hypersurface Reconstruction Problem 287
5.5 Regularization Theory 289
5.6 Regularization Networks 299
5.7 Generalized Radial-Basis Function Networks 300
5.8 XOR Problem (Revisited) 304
5.9 Estimation of the Regularization Parameter 306
5.10 Approximation Properties of RBF Networks 312
5.11 Comparison of RBF Networks and Multilayer Perceptrons 315
5.12 Kernel Regression and Its Relation to RBF Networks 316
5.13 Learning Strategies 320
5.14 Computer Experiment 327
5.15 Summary and Discussion 330
Notes and References 330
Problems 334

6 Support Vector Machines **340**

6.1 Introduction 340
6.2 Optimal Hyperplane for Linearly Separable Patterns 341
6.3 Optimal Hyperplane for Nonseparable Patterns 348
6.4 How to Build a Support Vector Machine for Pattern Recognition 351
6.5 Example: XOR Problem (Revisited) 357
6.6 *Computer Experiment* 359
6.7 ϵ-Insensitive Loss Function 361
6.8 Support Vector Machines for Nonlinear Regression 362
6.9 Summary and Discussion 365
Notes and References 369
Problems 370

7 Committee Machines 373

7.1 Introduction 373
7.2 Ensemble Averaging 375
7.3 Computer Experiment I 377
7.4 Boosting 379
7.5 Computer Experiment II 386
7.6 Associative Gaussian Mixture Model 388
7.7 Hierarchical Mixture of Experts Model 394
7.8 Model Selection Using a Standard Decision Tree 396
7.9 A Priori and a Posteriori Probabilities 399
7.10 Maximum Likelihood Estimation 390
7.11 Learning Strategies for the HME Model 402
7.12 EM Algorithm 404
7.13 Application of the EM Algorithm to the HME Model 405
7.14 Summary and Discussion 408
Notes and References 409
Problems 411

8 Principal Components Analysis 414

8.1 Introduction 414
8.2 Some Intuitive Principles of Self-Organization 415
8.3 Principal Components Analysis 418
8.4 Hebbian-Based Maximum Eigenfilter 426
8.5 Hebbian-Based Principal Components Analysis 435
8.6 Computer Experiment: Image Coding 441
8.7 Adaptive Principal Components Analysis Using Lateral Inhibition 444
8.8 Two Classes of PCA Algorithms 452
8.9 Batch and Adaptive Methods of Computation 452
8.10 Kernel-Based Principal Components Analysis 454
8.11 Summary And Discussion 459
Notes And References 461
Problems 462

9 Self-Organizing Maps 465

9.1 Introduction 465
9.2 Two Basic Feature-Mapping Models 466
9.3 Self-Organizing Map 468
9.4 Summary of the SOM Algorithm 475
9.5 Properties of the Feature Map 476
9.6 Computer Simulations 483
9.7 Learning Vector Quantization 488
9.8 Computer Experiment: Adaptive Pattern Classification 490

9.9 Hierarchical Vector Quantization 492
9.10 Contextual Maps 496
9.11 Summary and Discussion 498
Notes and References 499
Problems 501

10 Information-Theoretic Models 506

10.1 Introduction 506
10.2 Entropy 507
10.3 Maximum Entropy Principle 512
10.4 Mutual Information 514
10.5 Kullback–Leibler Divergence 517
10.6 Mutual Information as an Objective Function To Be Optimized 520
10.7 Maximum Mutual Information Principle 521
10.8 Infomax and Redundancy Reduction 525
10.9 Spatially Coherent Features 528
10.10 Spatially Incoherent Features 530
10.11 Independent Components Analysis 532
10.12 Computer Experiment 545
10.13 Maximum Likelihood Estimation 547
10.14 Maximum Entropy Method 551
10.15 Summary and Discussion 555
Notes and References 557
Problems 563

11 Stochastic Machines And Their Approximates Rooted In Statistical Mechanics 567

11.1 Introduction 567
11.2 Statistical Mechanics 568
11.3 Markov Chains 570
11.4 Metropolis Algorithm 578
11.5 Simulated Annealing 580
11.6 Gibbs Sampling 583
11.7 Boltzmann Machine 584
11.8 Sigmoid Belief Networks 591
11.9 Helmholtz Machine 596
11.10 Mean-Field Theory 598
11.11 Deterministic Boltzmann Machine 600
11.12 Deterministic Sigmoid Belief Networks 601
11.13 Deterministic Annealing 608
11.14 Summary and Discussion 614
Notes and References 616
Problems 619

12 Neurodynamic Programming 625

12.1 Introduction 625
12.2 Markovian Decision Processes 626
12.3 Bellman's Optimality Criterion 629
12.4 Policy Iteration 632
12.5 Value Iteration 634
12.6 Neurodynamic Programming 639
12.7 Approximate Policy Iteration 640
12.8 Q-Learning 644
12.9 Computer Experiment 649
12.10 Summary and Discussion 651
Notes and References 653
Problems 654

13 Temporal Processing Using Feedforward Networks 657

13.1 Introduction 657
13.2 Short-term Memory Structures 658
13.3 Network Architectures for Temporal Processing 662
13.4 Focused Time Lagged Feedforward Networks 665
13.5 Computer Experiment 667
13.6 Universal Myopic Mapping Theorem 668
13.7 Spatio-Temporal Models of a Neuron 670
13.8 Distributed Time Lagged Feedforward Networks 673
13.9 Temporal Back-Propagation Algorithm 674
13.10 Summary and Discussion 681
Notes and References 682
Problems 682

14 Neurodynamics 686

14.1 Introduction 686
14.2 Dynamical Systems 688
14.3 Stability of Equilibrium States 691
14.4 Attractors 696
14.5 Neurodynamical Models 698
14.6 Manipulation of Attractors as a Recurrent Network Paradigm 702
14.7 Hopfield Models 702
14.8 Computer Experiment I 718
14.9 Cohen–Grossberg Theorem 723
14.10 Brain-State-in-a-Box Model 725
14.11 Computer Experiment II 731
14.12 Strange Attractors and Chaos 731
14.13 Dynamic Reconstruction of a Chaotic Process 736

14.14 Computer Experiment III 740
14.15 Summary and Discussion 744
Notes and References 747
Problems 749

15 Dynamically Driven Recurrent Networks **754**

15.1 Introduction 754
15.2 Recurrent Network Architectures 755
15.3 State-Space Model 761
15.4 Nonlinear Autoregressive with Exogenous Inputs Model 768
15.5 Computational Power of Recurrent Networks 769
15.6 Learning Algorithms 772
15.7 Back-Propagation Through Time 773
15.8 Real-Time Recurrent Learning 778
15.9 Kalman Filters 784
15.10 Decoupled Extended Kalman Filters 787
15.11 Computer Experiment 792
15.12 Vanishing Gradients in Recurrent Networks 795
15.13 System Identification 798
15.14 Model-Reference Adaptive Control 802
15.15 Summary and Discussion 804
Notes and References 805
Problems 807

Epilogue 812

Index 818