

Neural Networks and Learning Machines

Third Edition

Simon Haykin

McMaster University

Hamilton, Ontario, Canada

PEARSON

Upper Saddle River Boston Columbus San Francisco

New York Indianapolis London Toronto Sydney Singapore

Tokyo Montreal Dubai Madrid Hong Kong Mexico City Munich

Paris Amsterdam Cape Town

Contents

Preface 10

Introduction 1

1. What is a Neural Network? 31
2. The Human Brain 36
3. Models of a Neuron 40
4. Neural Networks Viewed As Directed Graphs 45
5. Feedback 48
6. Network Architectures 51
7. Knowledge Representation 54
8. Learning Processes 64
9. Learning Tasks 68
10. Concluding Remarks 75
- Notes and References 76

Chapter 1 Rosenblatt's Perceptron 77

- 1.1 Introduction 77
- 1.2 Perceptron 78
- 1.3 The Perceptron Convergence Theorem 80
- 1.4 Relation Between the Perceptron and Bayes Classifier for a Gaussian Environment 85
- 1.5 Computer Experiment: Pattern Classification 90
- 1.6 The Batch Perceptron Algorithm 92
- 1.7 Summary and Discussion 95
- Notes and References 96
- Problems 96

Chapter 2 Model Building through Regression 98

- 2.1 Introduction 98
- 2.2 Linear Regression Model: Preliminary Considerations 99
- 2.3 Maximum a Posteriori Estimation of the Parameter Vector 101
- 2.4 Relationship Between Regularized Least-Squares Estimation and MAP Estimation 106
- 2.5 Computer Experiment: Pattern Classification 107
- 2.6 The Minimum-Description-Length Principle 109
- 2.7 Finite Sample-Size Considerations 112
- 2.8 The Instrumental-Variables Method 116
- 2.9 Summary and Discussion 118
- Notes and References 119
- Problems 119

Chapter 3 The Least-Mean-Square Algorithm 121

- 3.1 Introduction 121
- 3.2 Filtering Structure of the LMS Algorithm 122
- 3.3 Unconstrained Optimization: a Review 124
- 3.4 The Wiener Filter 130
- 3.5 The Least-Mean-Square Algorithm 132
- 3.6 Markov Model Portraying the Deviation of the LMS Algorithm from the Wiener Filter 134
- 3.7 The Langevin Equation: Characterization of Brownian Motion 136
- 3.8 Kushner's Direct-Averaging Method 137
- 3.9 Statistical LMS Learning Theory for Small Learning-Rate Parameter 138
- 3.10 Computer Experiment I: Linear Prediction 140
- 3.11 Computer Experiment II: Pattern Classification 142
- 3.12 Virtues and Limitations of the LMS Algorithm 143
- 3.13 Learning-Rate Annealing Schedules 145
- 3.14 Summary and Discussion 147
- Notes and References 148
- Problems 149

Chapter 4 Multilayer Perceptrons 152

- 4.1 Introduction 153
- 4.2 Some Preliminaries 154
- 4.3 Batch Learning and On-Line Learning 156
- 4.4 The Back-Propagation Algorithm 159
- 4.5 XOR Problem 171
- 4.6 Heuristics for Making the Back-Propagation Algorithm Perform Better 174
- 4.7 Computer Experiment: Pattern Classification 180
- 4.8 Back Propagation and Differentiation 183
- 4.9 The Hessian and Its Role in On-Line Learning 185
- 4.10 Optimal Annealing and Adaptive Control of the Learning Rate 187
- 4.11 Generalization 194
- 4.12 Approximations of Functions 196
- 4.13 Cross-Validation 201
- 4.14 Complexity Regularization and Network Pruning 205
- 4.15 Virtues and Limitations of Back-Propagation Learning 210
- 4.16 Supervised Learning Viewed as an Optimization Problem 216
- 4.17 Convolutional Networks 231
- 4.18 Nonlinear Filtering 233
- 4.19 Small-Scale Versus Large-Scale Learning Problems 239
- 4.20 Summary and Discussion 247
- Notes and References 249
- Problems 251

Chapter 5 Kernel Methods and Radial-Basis Function Networks 258

- 5.1 Introduction 258
- 5.2 Cover's Theorem on the Separability of Patterns 259
- 5.3 The Interpolation Problem 264
- 5.4 Radial-Basis-Function Networks 267
- 5.5 K-Means Clustering 270
- 5.6 Recursive Least-Squares Estimation of the Weight Vector 273
- 5.7 Hybrid Learning Procedure for RBF Networks 277
- 5.8 Computer Experiment: Pattern Classification 278
- 5.9 Interpretations of the Gaussian Hidden Units 280

5.10	Kernel Regression and Its Relation to RBF Networks	283
5.11	Summary and Discussion	287
	Notes and References	289
	Problems	291

Chapter 6 Support Vector Machines 296

6.1	Introduction	296
6.2	Optimal Hyperplane for Linearly Separable Patterns	297
6.3	Optimal Hyperplane for Nonseparable Patterns	304
6.4	The Support Vector Machine Viewed as a Kernel Machine	309
6.5	Design of Support Vector Machines	312
6.6	XOR Problem	314
6.7	Computer Experiment: Pattern Classification	317
6.8	Regression: Robustness Considerations	317
6.9	Optimal Solution of the Linear Regression Problem	321
6.10	The Representer Theorem and Related Issues	324
6.11	Summary and Discussion	330
	Notes and References	332
	Problems	335

Chapter 7 Regularization Theory 341

7.1	Introduction	341
7.2	Hadamard's Conditions for Well-Posedness	342
7.3	Tikhonov's Regularization Theory	343
7.4	Regularization Networks	354
7.5	Generalized Radial-Basis-Function Networks	355
7.6	The Regularized Least-Squares Estimator: Revisited	359
7.7	Additional Notes of Interest on Regularization	363
7.8	Estimation of the Regularization Parameter	364
7.9	Semisupervised Learning	370
7.10	Manifold Regularization: Preliminary Considerations	371
7.11	Differentiable Manifolds	373
7.12	Generalized Regularization Theory	376
7.13	Spectral Graph Theory	378
7.14	Generalized Representer Theorem	380
7.15	Laplacian Regularized Least-Squares Algorithm	382
7.16	Experiments on Pattern Classification Using Semisupervised Learning	384
7.17	Summary and Discussion	387
	Notes and References	389
	Problems	391

Chapter 8 Principal-Components Analysis 395

8.1	Introduction	395
8.2	Principles of Self-Organization	396
8.3	Self-Organized Feature Analysis	400
8.4	Principal-Components Analysis: Perturbation Theory	401
8.5	Hebbian-Based Maximum Eigenfilter	411
8.6	Hebbian-Based Principal-Components Analysis	420
8.7	Case Study: Image Coding	426
8.8	Kernel Principal-Components Analysis	429
8.9	Basic Issues Involved in the Coding of Natural Images	434
8.10	Kernel Hebbian Algorithm	435
8.11	Summary and Discussion	440
	Notes and References	443
	Problems	446

Chapter 9 Self-Organizing Maps 453

- 9.1 Introduction 453
- 9.2 Two Basic Feature-Mapping Models 454
- 9.3 Self-Organizing Map 456
- 9.4 Properties of the Feature Map 465
- 9.5 Computer Experiments I: Disentangling Lattice Dynamics Using SOM 473
- 9.6 Contextual Maps 475
- 9.7 Hierarchical Vector Quantization 478
- 9.8 Kernel Self-Organizing Map 482
- 9.9 Computer Experiment II: Disentangling Lattice Dynamics Using Kernel SOM 490
- 9.10 Relationship Between Kernel SOM and Kullback-Leibler Divergence 492
- 9.11 Summary and Discussion 494
- Notes and References 496
- Problems 498

Chapter 10 Information-Theoretic Learning Models 503

- 10.1 Introduction 504
- 10.2 Entropy 505
- 10.3 Maximum-Entropy Principle 509
- 10.4 Mutual Information 512
- 10.5 Kullback-Leibler Divergence 514
- 10.6 Copulas 517
- 10.7 Mutual Information as an Objective Function to be Optimized 521
- 10.8 Maximum Mutual Information Principle 522
- 10.9 Infomax and Redundancy Reduction 527
- 10.10 Spatially Coherent Features 529
- 10.11 Spatially Incoherent Features 532
- 10.12 Independent-Components Analysis 536
- 10.13 Sparse Coding of Natural Images and Comparison with ICA Coding 542
- 10.14 Natural-Gradient Learning for Independent-Components Analysis 544
- 10.15 Maximum-Likelihood Estimation for Independent-Components Analysis 554
- 10.16 Maximum-Entropy Learning for Blind Source Separation 557
- 10.17 Maximization of Negentropy for Independent-Components Analysis 562
- 10.18 Coherent Independent-Components Analysis 569
- 10.19 Rate Distortion Theory and Information Bottleneck 577
- 10.20 Optimal Manifold Representation of Data 581
- 10.21 Computer Experiment: Pattern Classification 588
- 10.22 Summary and Discussion 589
- Notes and References 592
- Problems 600

Chapter 11 Stochastic Methods Rooted in Statistical Mechanics 607

- 11.1 Introduction 608
- 11.2 Statistical Mechanics 608
- 11.3 Markov Chains 610
- 11.4 Metropolis Algorithm 619
- 11.5 Simulated Annealing 622
- 11.6 Gibbs Sampling 624
- 11.7 Boltzmann Machine 626
- 11.8 Logistic Belief Nets 632
- 11.9 Deep Belief Nets 634
- 11.10 Deterministic Annealing 638

10.15	Ordinamento geometrico	401
10.16	Liste concatenate	401
10.17	Ordinamento dei triangoli	404
Parte 3	Linee nascoste	409
10.18	Eliminazione delle linee nascoste	409
10.19	Copertura delle linee da parte dei triangoli	411
10.20	Suddivisione di una linea	418
10.21	Controllo delle linee nascoste per i lati del triangolo	423
Parte 4	Superfici nascoste concave	427
10.22	Generalizzazione ai poligoni concavi	427
	Esercizi	435
	Problemi di programmazione	435
Capitolo 11	Ombreggiatura	437
	Introduzione	437
11.1	Illuminazione diffusa	437
11.2	Illuminazione da sorgente puntiforme	440
11.3	Riflessione speculare	442
11.4	Algoritmi di ombreggiatura	446
11.5	Trasparenza e ombre	456
	Esercizi	456
	Problemi di programmazione	457
Capitolo 12	Le curve	459
	Introduzione	459
12.1	Generazione di curve	459
12.2	Interpolazione	462
12.3	Algoritmi di interpolazione	467
12.4	Arrotondamento di poligoni	473
12.5	B-spline	475
12.6	B-spline e angoli	481
	Applicazione pratica	483
	Esercizi	486
	Problemi di programmazione	486
Appendice A	Il sistema grafico CORE	487
Appendice B	Grafica su un terminale non grafico	495
Appendice C	Interfacciamento con il PLOT10	501
Appendice D	Interfacciamento con le routine CALCOMP	505
	Bibliografia	509
	Indice analitico	511

- 11.11 Analogy of Deterministic Annealing with Expectation-Maximization Algorithm 644
- 11.12 Summary and Discussion 645
- Notes and References 647
- Problems 649

Chapter 12 Dynamic Programming 655

- 12.1 Introduction 655
- 12.2 Markov Decision Process 657
- 12.3 Bellman's Optimality Criterion 659
- 12.4 Policy Iteration 663
- 12.5 Value Iteration 665
- 12.6 Approximate Dynamic Programming: Direct Methods 670
- 12.7 Temporal-Difference Learning 671
- 12.8 Q-Learning 676
- 12.9 Approximate Dynamic Programming: Indirect Methods 680
- 12.10 Least-Squares Policy Evaluation 683
- 12.11 Approximate Policy Iteration 688
- 12.12 Summary and Discussion 691
- Notes and References 693
- Problems 696

Chapter 13 Neurodynamics 700

- 13.1 Introduction 700
- 13.2 Dynamic Systems 702
- 13.3 Stability of Equilibrium States 706
- 13.4 Attractors 712
- 13.5 Neurodynamic Models 714
- 13.6 Manipulation of Attractors as a Recurrent Network Paradigm 717
- 13.7 Hopfield Model 718
- 13.8 The Cohen-Grossberg Theorem 731
- 13.9 Brain-State-In-A-Box Model 733
- 13.10 Strange Attractors and Chaos 739
- 13.11 Dynamic Reconstruction of a Chaotic Process 744
- 13.12 Summary and Discussion 750
- Notes and References 752
- Problems 755

Chapter 14 Bayesian Filtering for State Estimation of Dynamic Systems 759

- 14.1 Introduction 759
- 14.2 State-Space Models 760
- 14.3 Kalman Filters 764
- 14.4 The Divergence-Phenomenon and Square-Root Filtering 772
- 14.5 The Extended Kalman Filter 778
- 14.6 The Bayesian Filter 783
- 14.7 Cubature Kalman Filter: Building on the Kalman Filter 787
- 14.8 Particle Filters 793
- 14.9 Computer Experiment: Comparative Evaluation of Extended Kalman and Particle Filters 803
- 14.10 Kalman Filtering in Modeling of Brain Functions 805
- 14.11 Summary and Discussion 808
- Notes and References 810
- Problems 812

Chapter 15 Dynamically Driven Recurrent Networks 818

- 15.1 Introduction 818
- 15.2 Recurrent Network Architectures 819
- 15.3 Universal Approximation Theorem 825
- 15.4 Controllability and Observability 827
- 15.5 Computational Power of Recurrent Networks 832
- 15.6 Learning Algorithms 834
- 15.7 Back Propagation Through Time 836
- 15.8 Real-Time Recurrent Learning 840
- 15.9 Vanishing Gradients in Recurrent Networks 846
- 15.10 Supervised Training Framework for Recurrent Networks Using Nonlinear Sequential State Estimators 850
- 15.11 Computer Experiment: Dynamic Reconstruction of Mackay-Glass Attractor 857
- 15.12 Adaptivity Considerations 859
- 15.13 Case Study: Model Reference Applied to Neurocontrol 861
- 15.14 Summary and Discussion 863
 - Notes and References 867
 - Problems 870

Bibliography 875

Index 916