

Contents

1	Introduction	1
1.1	Major Events in Neural Networks Research	1
1.2	Neurons	3
1.2.1	The McCulloch–Pitts Neuron Model	5
1.2.2	Spiking Neuron Models	6
1.3	Neural Networks	8
1.4	Scope of the Book	12
	References	13
2	Fundamentals of Machine Learning	15
2.1	Learning Methods	15
2.2	Learning and Generalization	19
2.2.1	Generalization Error	21
2.2.2	Generalization by Stopping Criterion	21
2.2.3	Generalization by Regularization	23
2.2.4	Fault Tolerance and Generalization	24
2.2.5	Sparsity Versus Stability	25
2.3	Model Selection	25
2.3.1	Crossvalidation	26
2.3.2	Complexity Criteria	28
2.4	Bias and Variance	29
2.5	Robust Learning	31
2.6	Neural Network Processors	33
2.7	Criterion Functions	36
2.8	Computational Learning Theory	39
2.8.1	Vapnik-Chervonenkis Dimension	40
2.8.2	Empirical Risk-Minimization Principle	41
2.8.3	Probably Approximately Correct Learning	43
2.9	No-Free-Lunch Theorem	44
2.10	Neural Networks as Universal Machines	45
2.10.1	Boolean Function Approximation	45
2.10.2	Linear Separability and Nonlinear Separability	47
2.10.3	Continuous Function Approximation	49
2.10.4	Winner-Takes-All	50

2.11	Compressed Sensing and Sparse Approximation	51
2.11.1	Compressed Sensing	51
2.11.2	Sparse Approximation	53
2.11.3	LASSO and Greedy Pursuit	54
2.12	Bibliographical Notes	55
	References	59
3	Perceptrons	67
3.1	One-Neuron Perceptron	67
3.2	Single-Layer Perceptron	68
3.3	Perceptron Learning Algorithm.	69
3.4	Least-Mean Squares (LMS) Algorithm	71
3.5	P-Delta Rule	74
3.6	Other Learning Algorithms	76
	References	79
4	Multilayer Perceptrons: Architecture and Error Backpropagation	83
4.1	Introduction	83
4.2	Universal Approximation.	84
4.3	Backpropagation Learning Algorithm	85
4.4	Incremental Learning Versus Batch Learning.	90
4.5	Activation Functions for the Output Layer.	95
4.6	Optimizing Network Structure	96
4.6.1	Network Pruning Using Sensitivity Analysis	96
4.6.2	Network Pruning Using Regularization	99
4.6.3	Network Growing	101
4.7	Speeding Up Learning Process	102
4.7.1	Eliminating Premature Saturation	102
4.7.2	Adapting Learning Parameters	104
4.7.3	Initializing Weights.	108
4.7.4	Adapting Activation Function.	110
4.8	Some Improved BP Algorithms	112
4.8.1	BP with Global Descent.	113
4.8.2	Robust BP Algorithms.	115
4.9	Resilient Propagation (RProp)	115
	References	119
5	Multilayer Perceptrons: Other Learning Techniques	127
5.1	Introduction to Second-Order Learning Methods.	127
5.2	Newton's Methods	128
5.2.1	Gauss–Newton Method	129
5.2.2	Levenberg–Marquardt Method	130

5.3	Quasi-Newton Methods	133
5.3.1	BFGS Method	134
5.3.2	One-Step Secant Method	136
5.4	Conjugate-Gradient Methods	136
5.5	Extended Kalman Filtering Methods	141
5.6	Recursive Least Squares	143
5.7	Natural-Gradient Descent Method	144
5.8	Other Learning Algorithms	145
5.8.1	Layerwise Linear Learning	145
5.9	Escaping Local Minima	146
5.10	Complex-Valued MLPs and Their Learning	147
5.10.1	Split Complex BP	148
5.10.2	Fully Complex BP	148
	References	152
6	Hopfield Networks, Simulated Annealing, and Chaotic Neural Networks	159
6.1	Hopfield Model	159
6.2	Continuous-Time Hopfield Network	162
6.3	Simulated Annealing	165
6.4	Hopfield Networks for Optimization	168
6.4.1	Combinatorial Optimization Problems	169
6.4.2	Escaping Local Minima for Combinatorial Optimization Problems	172
6.4.3	Solving Other Optimization Problems	173
6.5	Chaos and Chaotic Neural Networks	175
6.5.1	Chaos, Bifurcation, and Fractals	175
6.5.2	Chaotic Neural Networks	176
6.6	Multistate Hopfield Networks	179
6.7	Cellular Neural Networks	180
	References	183
7	Associative Memory Networks	187
7.1	Introduction	187
7.2	Hopfield Model: Storage and Retrieval	189
7.2.1	Generalized Hebbian Rule	189
7.2.2	Pseudoinverse Rule	191
7.2.3	Perceptron-Type Learning Rule	191
7.2.4	Retrieval Stage	192
7.3	Storage Capability of the Hopfield Model	193
7.4	Increasing Storage Capacity	197
7.5	Multistate Hopfield Networks for Associative Memory	200
7.6	Multilayer Perceptrons as Associative Memories	201
7.7	Hamming Network	203

7.8	Bidirectional Associative Memories	205
7.9	Cohen–Grossberg Model	206
7.10	Cellular Networks	207
	References	211
8	Clustering I: Basic Clustering Models and Algorithms	215
8.1	Introduction	215
8.1.1	Vector Quantization	215
8.1.2	Competitive Learning	217
8.2	Self-Organizing Maps	218
8.2.1	Kohonen Network	220
8.2.2	Basic Self-Organizing Maps	221
8.3	Learning Vector Quantization	228
8.4	Nearest-Neighbor Algorithms	231
8.5	Neural Gas	234
8.6	ART Networks	237
8.6.1	ART Models	238
8.6.2	ART 1	239
8.7	C-Means Clustering	241
8.8	Subtractive Clustering	244
8.9	Fuzzy Clustering	247
8.9.1	Fuzzy C-Means Clustering	247
8.9.2	Other Fuzzy Clustering Algorithms	250
	References	253
9	Clustering II: Topics in Clustering	259
9.1	The Underutilization Problem	259
9.1.1	Competitive Learning with Conscience	259
9.1.2	Rival Penalized Competitive Learning	261
9.1.3	Softcompetitive Learning	263
9.2	Robust Clustering	264
9.2.1	Possibilistic C-Means	266
9.2.2	A Unified Framework for Robust Clustering	267
9.3	Supervised Clustering	268
9.4	Clustering Using Non-Euclidean Distance Measures	269
9.5	Partitional, Hierarchical, and Density-Based Clustering	271
9.6	Hierarchical Clustering	272
9.6.1	Distance Measures, Cluster Representations, and Dendrograms	272
9.6.2	Minimum Spanning Tree (MST) Clustering	274
9.6.3	BIRCH, CURE, CHAMELEON, and DBSCAN	276
9.6.4	Hybrid Hierarchical/Partitional Clustering	279

9.7	Constructive Clustering Techniques	280
9.8	Cluster Validity	282
9.8.1	Measures Based on Compactness and Separation of Clusters	282
9.8.2	Measures Based on Hypervolume and Density of Clusters	284
9.8.3	Crisp Silhouette and Fuzzy Silhouette	285
9.9	Projected Clustering	286
9.10	Spectral Clustering	288
9.11	Coclustering	289
9.12	Handling Qualitative Data	289
9.13	Bibliographical Notes	290
	References	291
10	Radial Basis Function Networks	299
10.1	Introduction	299
10.1.1	RBF Network Architecture	300
10.1.2	Universal Approximation of RBF Networks	301
10.1.3	RBF Networks and Classification	302
10.1.4	Learning for RBF Networks	302
10.2	Radial Basis Functions	303
10.3	Learning RBF Centers	306
10.4	Learning the Weights	308
10.4.1	Least-Squares Methods for Weight Learning	308
10.5	RBF Network Learning Using Orthogonal Least-Squares	310
10.5.1	Batch Orthogonal Least-Squares	310
10.5.2	Recursive Orthogonal Least-Squares	312
10.6	Supervised Learning of All Parameters	313
10.6.1	Supervised Learning for General RBF Networks	313
10.6.2	Supervised Learning for Gaussian RBF Networks	314
10.6.3	Discussion on Supervised Learning	316
10.6.4	Extreme Learning Machines	316
10.7	Various Learning Methods	317
10.8	Normalized RBF Networks	319
10.9	Optimizing Network Structure	320
10.9.1	Constructive Methods	320
10.9.2	Resource-Allocating Networks	322
10.9.3	Pruning Methods	324
10.10	Complex RBF Networks	324
10.11	A Comparison of RBF Networks and MLPs	326
10.12	Bibliographical Notes	328
	References	330

11	Recurrent Neural Networks	337
11.1	Introduction	337
11.2	Fully Connected Recurrent Networks	339
11.3	Time-Delay Neural Networks	340
11.4	Backpropagation for Temporal Learning	342
11.5	RBF Networks for Modeling Dynamic Systems	345
11.6	Some Recurrent Models	346
11.7	Reservoir Computing	348
	References	351
12	Principal Component Analysis	355
12.1	Introduction	355
12.1.1	Hebbian Learning Rule	356
12.1.2	Oja's Learning Rule	357
12.2	PCA: Conception and Model	358
12.2.1	Factor Analysis	361
12.3	Hebbian Rule-Based PCA	362
12.3.1	Subspace Learning Algorithms	362
12.3.2	Generalized Hebbian Algorithm	366
12.4	Least Mean Squared Error-Based PCA	368
12.4.1	Other Optimization-Based PCA	371
12.5	Anti-Hebbian Rule-Based PCA	372
12.5.1	APEX Algorithm	374
12.6	Nonlinear PCA	378
12.6.1	Autoassociative Network-Based Nonlinear PCA	379
12.7	Minor Component Analysis	380
12.7.1	Extracting the First Minor Component	380
12.7.2	Self-Stabilizing Minor Component Analysis	381
12.7.3	Oja-Based MCA	382
12.7.4	Other Algorithms	383
12.8	Constrained PCA	383
12.8.1	Sparse PCA	385
12.9	Localized PCA, Incremental PCA, and Supervised PCA	386
12.10	Complex-Valued PCA	387
12.11	Two-Dimensional PCA	388
12.12	Generalized Eigenvalue Decomposition	390
12.13	Singular Value Decomposition	391
12.13.1	Crosscorrelation Asymmetric PCA Networks	391
12.13.2	Extracting Principal Singular Components for Nonsquare Matrices	394
12.13.3	Extracting Multiple Principal Singular Components	395
12.14	Canonical Correlation Analysis	396
	References	399

13	Nonnegative Matrix Factorization	407
13.1	Introduction	407
13.2	Algorithms for NMF	408
13.2.1	Multiplicative Update Algorithm and Alternating Nonnegative Least Squares	409
13.3	Other NMF Methods	411
13.3.1	NMF Methods for Clustering	414
	References	415
14	Independent Component Analysis	419
14.1	Introduction	419
14.2	ICA Model	420
14.3	Approaches to ICA	421
14.4	Popular ICA Algorithms	424
14.4.1	Infomax ICA	424
14.4.2	EASI, JADE, and Natural-Gradient ICA	425
14.4.3	FastICA Algorithm	426
14.5	ICA Networks	431
14.6	Some ICA Methods	434
14.6.1	Nonlinear ICA	434
14.6.2	Constrained ICA	434
14.6.3	Nonnegativity ICA	435
14.6.4	ICA for Convolutional Mixtures	436
14.6.5	Other Methods	437
14.7	Complex-Valued ICA	439
14.8	Stationary Subspace Analysis and Slow Feature Analysis	441
14.9	EEG, MEG and fMRI	442
	References	446
15	Discriminant Analysis	451
15.1	Linear Discriminant Analysis	451
15.1.1	Solving Small Sample Size Problem	454
15.2	Fisherfaces	455
15.3	Regularized LDA	456
15.4	Uncorrelated LDA and Orthogonal LDA	457
15.5	LDA/GSVD and LDA/QR	459
15.6	Incremental LDA	460
15.7	Other Discriminant Methods	460
15.8	Nonlinear Discriminant Analysis	462
15.9	Two-Dimensional Discriminant Analysis	464
	References	465

16. Support Vector Machines	469
16.1 Introduction	469
16.2 SVM Model	472
16.3 Solving the Quadratic Programming Problem	475
16.3.1 Chunking	476
16.3.2 Decomposition	476
16.3.3 Convergence of Decomposition Methods	480
16.4 Least-Squares SVMs	481
16.5 SVM Training Methods	484
16.5.1 SVM Algorithms with Reduced Kernel Matrix	484
16.5.2 ν -SVM	485
16.5.3 Cutting-Plane Technique	486
16.5.4 Gradient-Based Methods	487
16.5.5 Training SVM in the Primal Formulation	488
16.5.6 Clustering-Based SVM	489
16.5.7 Other Methods	490
16.6 Pruning SVMs	493
16.7 Multiclass SVMs	495
16.8 Support Vector Regression	497
16.9 Support Vector Clustering	502
16.10 Distributed and Parallel SVMs	504
16.11 SVMs for One-Class Classification	506
16.12 Incremental SVMs	507
16.13 SVMs for Active, Transductive, and Semi-Supervised Learning	509
16.13.1 SVMs for Active Learning	509
16.13.2 SVMs for Transductive or Semi-Supervised Learning	509
16.14 Probabilistic Approach to SVM	512
16.14.1 Relevance Vector Machines	513
References	514
17 Other Kernel Methods	525
17.1 Introduction	525
17.2 Kernel PCA	527
17.3 Kernel LDA	531
17.4 Kernel Clustering	533
17.5 Kernel Autoassociators, Kernel CCA and Kernel ICA	534
17.6 Other Kernel Methods	536
17.7 Multiple Kernel Learning	537
References	540

18	Reinforcement Learning	547
18.1	Introduction	547
18.2	Learning Through Awards	549
18.3	Actor-Critic Model	551
18.4	Model-Free and Model-Based Reinforcement Learning	552
18.5	Temporal-Difference Learning	554
18.6	Q -Learning	556
18.7	Learning Automata	558
	References	560
19	Probabilistic and Bayesian Networks	563
19.1	Introduction	563
19.1.1	Classical Versus Bayesian Approach	564
19.1.2	Bayes' Theorem	565
19.1.3	Graphical Models	566
19.2	Bayesian Network Model.	567
19.3	Learning Bayesian Networks	570
19.3.1	Learning the Structure	570
19.3.2	Learning the Parameters.	575
19.3.3	Constraint-Handling	577
19.4	Bayesian Network Inference.	577
19.4.1	Belief Propagation.	578
19.4.2	Factor Graphs and the Belief Propagation Algorithm	580
19.5	Sampling (Monte Carlo) Methods	583
19.5.1	Gibbs Sampling	585
19.6	Variational Bayesian Methods	586
19.7	Hidden Markov Models.	588
19.8	Dynamic Bayesian Networks	591
19.9	Expectation–Maximization Algorithm	592
19.10	Mixture Models	594
19.10.1	Probabilistic PCA	595
19.10.2	Probabilistic Clustering	596
19.10.3	Probabilistic ICA	597
19.11	Bayesian Approach to Neural Network Learning	599
19.12	Boltzmann Machines.	601
19.12.1	Boltzmann Learning Algorithm.	602
19.12.2	Mean-Field-Theory Machine	604
19.12.3	Stochastic Hopfield Networks.	605
19.13	Training Deep Networks	606
	References	610

20	Combining Multiple Learners: Data Fusion and Emsemble Learning	621
20.1	Introduction	621
20.1.1	Ensemble Learning Methods	622
20.1.2	Aggregation	623
20.2	Boosting	624
20.2.1	AdaBoost	625
20.3	Bagging	628
20.4	Random Forests	629
20.5	Topics in Ensemble Learning	630
20.6	Solving Multiclass Classification	632
20.6.1	One-Against-All Strategy	633
20.6.2	One-Against-One Strategy	633
20.6.3	Error-Correcting Output Codes (ECOCs)	634
20.7	Dempster-Shafer Theory of Evidence	637
	References	640
21	Introduction to Fuzzy Sets and Logic	645
21.1	Introduction	645
21.2	Definitions and Terminologies	646
21.3	Membership Function	652
21.4	Intersection, Union, and Negation	653
21.5	Fuzzy Relation and Aggregation	655
21.6	Fuzzy Implication	657
21.7	Reasoning and Fuzzy Reasoning	658
21.7.1	Modus Ponens and Modus Tollens	659
21.7.2	Generalized Modus Ponens	659
21.7.3	Fuzzy Reasoning Methods	661
21.8	Fuzzy Inference Systems	662
21.8.1	Fuzzy Rules and Fuzzy Interference	663
21.8.2	Fuzzification and Defuzzification	664
21.9	Fuzzy Models	665
21.9.1	Mamdani Model	665
21.9.2	Takagi–Sugeno–Kang Model	667
21.10	Complex Fuzzy Logic	668
21.11	Possibility Theory	669
21.12	Case-Based Reasoning	670
21.13	Granular Computing and Ontology	671
	References	675
22	Neurofuzzy Systems	677
22.1	Introduction	677
22.1.1	Interpretability	678

22.2	Rule Extraction from Trained Neural Networks	679
22.2.1	Fuzzy Rules and Multilayer Perceptrons	679
22.2.2	Fuzzy Rules and RBF Networks	680
22.2.3	Rule Extraction from SVMs	681
22.2.4	Rule Generation from Other Neural Networks	682
22.3	Extracting Rules from Numerical data	683
22.3.1	Rule Generation Based on Fuzzy Partitioning	684
22.3.2	Other Methods	685
22.4	Synergy of Fuzzy Logic and Neural Networks	687
22.5	ANFIS Model	688
22.6	Fuzzy SVMs	693
22.7	Other Neurofuzzy Models	696
	References	700
23	Neural Circuits and Parallel Implementation	705
23.1	Introduction	705
23.2	Hardware/Software Codesign	707
23.3	Topics in Digital Circuit Designs	708
23.4	Circuits for Neural-Network Models	709
23.4.1	Circuits for MLPs	709
23.4.2	Circuits for RBF Networks	711
23.4.3	Circuits for Clustering	712
23.4.4	Circuits for SVMs	712
23.4.5	Circuits of Other Models	713
23.5	Fuzzy Neural Circuits	715
23.6	Graphic Processing Unit implementation	716
23.7	Implementation Using Systolic Algorithms	717
23.8	Implementation Using Parallel Computers	718
23.9	Implementation Using Cloud Computing	720
	References	721
24	Pattern Recognition for Biometrics and Bioinformatics	727
24.1	Biometrics	728
24.1.1	Physiological Biometrics and Recognition	728
24.1.2	Behavioral Biometrics and Recognition	731
24.2	Face Detection and Recognition	732
24.2.1	Face Detection	733
24.2.2	Face Recognition	734
24.3	Bioinformatics	736
24.3.1	Microarray Technology	739
24.3.2	Motif Discovery, Sequence Alignment, Protein Folding, and Coclustering	741
	References	743

25 Data Mining	747
25.1 Introduction	747
25.2 Document Representations for Text Categorization	748
25.3 Neural Network Approach to Data Mining	750
25.3.1 Classification-Based Data Mining	750
25.3.2 Clustering-Based Data Mining	752
25.3.3 Bayesian Network-Based Data Mining	755
25.4 Personalized Search	756
25.5 XML Format	759
25.6 Web Usage Mining	760
25.7 Association Mining	763
25.8 Ranking Search Results	761
25.8.1 Surfer Models	762
25.8.2 PageRank Algorithm	763
25.8.3 Hypertext Induced Topic Search (HITS)	766
25.9 Data Warehousing	767
25.10 Content-Based Image Retrieval	768
25.11 E-mail Anti-Spamming	771
References	773
 Appendix A: Mathematical Preliminaries	 779
 Appendix B: Benchmarks and Resources	 799
 About the Authors	 813
 Index	 815

<http://www.springer.com/978-1-4471-5570-6>

Neural Networks and Statistical Learning

Du, K.-L.; Swamy, M.N.S.

2014, XXVII, 824 p. 166 illus., 68 illus. in color.,

Hardcover

ISBN: 978-1-4471-5570-6