

Fundamentals of Artificial Neural Networks

Mohamad H. Hassoun

A Bradford Book

The MIT Press
Cambridge, Massachusetts
London, England

Contents

| | |
|----------------------------------------------------------------------------------------|-----------|
| Preface | xiii |
| Acknowledgments | xix |
| Abbreviations | xxi |
| Symbols | xxiii |
| 1 Threshold Gates | 1 |
| 1.1 Threshold Gates | 2 |
| 1.1.1 Linear Threshold Gates | 2 |
| 1.1.2 Quadratic Threshold Gates | 7 |
| 1.1.3 Polynomial Threshold Gates | 8 |
| 1.2 Computational Capabilities of Polynomial Threshold Gates | 9 |
| 1.3 General Position and the Function Counting Theorem | 15 |
| 1.3.1 Weierstrass's Approximation Theorem | 15 |
| 1.3.2 Points in General Position | 16 |
| 1.3.3 Function Counting Theorem | 17 |
| 1.3.4 Separability in ϕ -Space | 20 |
| 1.4 Minimal PTG Realization of Arbitrary Switching Functions | 21 |
| 1.5 Ambiguity and Generalization | 24 |
| 1.6 Extreme Points | 27 |
| 1.7 Summary | 29 |
| Problems | 30 |
| 2 Computational Capabilities of Artificial Neural Networks | 35 |
| 2.1 Some Preliminary Results on Neural Network Mapping Capabilities | 35 |
| 2.1.1 Network Realization of Boolean Functions | 35 |
| 2.1.2 Bounds on the Number of Functions Realizable by a Feedforward Network of LTG's | 38 |
| 2.2 Necessary Lower Bounds on the Size of LTG Networks | 41 |
| 2.2.1 Two Layer Feedforward Networks | 41 |
| 2.2.2 Three Layer Feedforward Networks | 44 |
| 2.2.3 Generally Interconnected Networks with No Feedback | 45 |
| 2.3 Approximation Capabilities of Feedforward Neural Networks for Continuous Functions | 46 |
| 2.3.1 Kolmogorov's Theorem | 46 |
| 2.3.2 Single-Hidden-Layer Neural Networks Are Universal Approximators | 47 |
| 2.3.3 Single-Hidden-Layer Neural Networks Are Universal Classifiers | 50 |

| | | |
|----------|---------------------------------------------------------------------------------------------------|-----------|
| 2.4 | Computational Effectiveness of Neural Networks | 51 |
| 2.4.1 | Algorithmic Complexity | 51 |
| 2.4.2 | Computational Energy | 52 |
| 2.5 | Summary | 53 |
| | Problems | 54 |
| 3 | Learning Rules | 57 |
| 3.1 | Supervised Learning in a Single-Unit Setting | 57 |
| 3.1.1 | Error Correction Rules | 58 |
| 3.1.2 | Other Gradient-Descent-Based Learning Rules | 67 |
| 3.1.3 | Extension of the μ -LMS Rule to Units with Differentiable Activation Functions: Delta Rule | 76 |
| 3.1.4 | Adaptive Ho-Kashyap (AHK) Learning Rules | 78 |
| 3.1.5 | Other Criterion Functions | 82 |
| 3.1.6 | Extension of Gradient-Descent-Based Learning to Stochastic Units | 87 |
| 3.2 | Reinforcement Learning | 88 |
| 3.2.1 | Associative Reward-Penalty Reinforcement Learning Rule | 89 |
| 3.3 | Unsupervised Learning | 90 |
| 3.3.1 | Hebbian Learning | 90 |
| 3.3.2 | Oja's Rule | 92 |
| 3.3.3 | Yuille et al. Rule | 92 |
| 3.3.4 | Linsker's Rule | 95 |
| 3.3.5 | Hebbian Learning in a Network Setting: Principal-Component Analysis (PCA) | 97 |
| 3.3.6 | Nonlinear PCA | 101 |
| 3.4 | Competitive Learning | 103 |
| 3.4.1 | Simple Competitive Learning | 103 |
| 3.4.2 | Vector Quantization | 109 |
| 3.5 | Self-Organizing Feature Maps: Topology-Preserving Competitive Learning | 112 |
| 3.5.1 | Kohonen's SOFM | 113 |
| 3.5.2 | Examples of SOFMs | 114 |
| 3.6 | Summary | 126 |
| | Problems | 134 |

| | | |
|----------|----------------------------------------------------------|-----|
| 4 | Mathematical Theory of Neural Learning | 143 |
| 4.1 | Learning as a Search/Approximation Mechanism | 143 |
| 4.2 | Mathematical Theory of Learning in a Single-Unit Setting | 145 |
| 4.2.1 | General Learning Equation | 146 |
| 4.2.2 | Analysis of the Learning Equation | 147 |
| 4.2.3 | Analysis of Some Basic Learning Rules | 148 |
| 4.3 | Characterization of Additional Learning Rules | 152 |
| 4.3.1 | Simple Hebbian Learning | 154 |
| 4.3.2 | Improved Hebbian Learning | 155 |
| 4.3.3 | Oja's Rule | 156 |
| 4.3.4 | Yuille et al. Rule | 158 |
| 4.3.5 | Hassoun's Rule | 161 |
| 4.4 | Principal-Component Analysis (PCA) | 163 |
| 4.5 | Theory of Reinforcement Learning | 165 |
| 4.6 | Theory of Simple Competitive Learning | 166 |
| 4.6.1 | Deterministic Analysis | 167 |
| 4.6.2 | Stochastic Analysis | 168 |
| 4.7 | Theory of Feature Mapping | 171 |
| 4.7.1 | Characterization of Kohonen's Feature Map | 171 |
| 4.7.2 | Self-Organizing Neural Fields | 173 |
| 4.8 | Generalization | 180 |
| 4.8.1 | Generalization Capabilities of Deterministic Networks | 180 |
| 4.8.2 | Generalization in Stochastic Networks | 185 |
| 4.9 | Complexity of Learning | 187 |
| 4.10 | Summary | 190 |
| | Problems | 190 |
| 5 | Adaptive Multilayer Neural Networks I | 197 |
| 5.1 | Learning Rule for Multilayer Feedforward Neural Networks | 197 |
| 5.1.1 | Error Backpropagation Learning Rule | 199 |
| 5.1.2 | Global-Descent-Based Error Backpropagation | 206 |
| 5.2 | Backprop Enhancements and Variations | 210 |
| 5.2.1 | Weights Initialization | 210 |
| 5.2.2 | Learning Rate | 211 |
| 5.2.3 | Momentum | 213 |

| | | |
|----------|-----------------------------------------------------------|------------|
| 5.2.4 | Activation Function | 219 |
| 5.2.5 | Weight Decay, Weight Elimination, and Unit Elimination | 221 |
| 5.2.6 | Cross-Validation | 226 |
| 5.2.7 | Criterion Functions | 230 |
| 5.3 | Applications | 234 |
| 5.3.1 | NETtalk | 234 |
| 5.3.2 | Glove-Talk | 236 |
| 5.3.3 | Handwritten ZIP Code Recognition | 240 |
| 5.3.4 | ALVINN: A Trainable Autonomous Land Vehicle | 244 |
| 5.3.5 | Medical Diagnosis Expert Net | 246 |
| 5.3.6 | Image Compression and Dimensionality Reduction | 247 |
| 5.4 | Extensions of Backprop for Temporal Learning | 253 |
| 5.4.1 | Time-Delay Neural Networks | 254 |
| 5.4.2 | Backpropagation Through Time | 259 |
| 5.4.3 | Recurrent Backpropagation | 267 |
| 5.4.4 | Time-Dependent Recurrent Backpropagation | 271 |
| 5.4.5 | Real-Time Recurrent Learning | 274 |
| 5.5 | Summary | 275 |
| | Problems | 276 |
| 6 | Adaptive Multilayer Neural Networks II | 285 |
| 6.1 | Radial Basis Function (RBF) Networks | 285 |
| 6.1.1 | RBF Networks versus Backprop Networks | 294 |
| 6.1.2 | RBF Network Variations | 296 |
| 6.2 | Cerebellar Model Articulation Controller (CMAC) | 301 |
| 6.2.1 | CMAC Relation to Rosenblatt's Perceptron and Other Models | 304 |
| 6.3 | Unit-Allocating Adaptive Networks | 310 |
| 6.3.1 | Hyperspherical Classifiers | 311 |
| 6.3.2 | Cascade-Correlation Network | 318 |
| 6.4 | Clustering Networks | 322 |
| 6.4.1 | Adaptive Resonance Theory (ART) Networks | 323 |
| 6.4.2 | Autoassociative Clustering Network | 328 |
| 6.5 | Summary | 337 |
| | Problems | 339 |

| | | |
|----------|--------------------------------------------------------------------------------|------------|
| 7 | Associative Neural Memories | 345 |
| 7.1 | Basic Associative Neural Memory Models | 345 |
| 7.1.1 | Simple Associative Memories and their Associated Recording Recipes | 346 |
| 7.1.2 | Dynamic Associative Memories (DAMs) | 353 |
| 7.2 | DAM Capacity and Retrieval Dynamics | 363 |
| 7.2.1 | Correlation DAMs | 363 |
| 7.2.2 | Projection DAMs | 369 |
| 7.3 | Characteristics of High-Performance DAMs | 374 |
| 7.4 | Other DAM Models | 375 |
| 7.4.1 | Brain-State-in-a-Box (BSB) DAM | 375 |
| 7.4.2 | Nonmonotonic Activations DAM | 381 |
| 7.4.3 | Hysteretic Activations DAM | 386 |
| 7.4.4 | Exponential-Capacity DAM | 389 |
| 7.4.5 | Sequence-Generator DAM | 391 |
| 7.4.6 | Heteroassociative DAM | 392 |
| 7.5 | The DAM as a Gradient Net and Its Application to Combinatorial Optimization | 394 |
| 7.6 | Summary Problems | 400 401 |
| 8 | Global Search Methods for Neural Networks | 417 |
| 8.1 | Local versus Global Search | 417 |
| 8.1.1 | A Gradient Descent/Ascent Search Strategy | 419 |
| 8.1.2 | Stochastic Gradient Search: Global Search via Diffusion | 421 |
| 8.2 | Simulated Annealing-Based Global Search | 424 |
| 8.3 | Simulated Annealing for Stochastic Neural Networks | 428 |
| 8.3.1 | Global Convergence in a Stochastic Recurrent Neural Net: The Boltzmann Machine | 429 |
| 8.3.2 | Learning in Boltzmann Machines | 431 |
| 8.4 | Mean-Field Annealing and Deterministic Boltzmann Machines | 436 |
| 8.4.1 | Mean-Field Retrieval | 437 |
| 8.4.2 | Mean-Field Learning | 438 |
| 8.5 | Genetic Algorithms in Neural Network Optimization | 439 |
| 8.5.1 | Fundamentals of Genetic Algorithms | 439 |
| 8.5.2 | Application of Genetic Algorithms to Neural Networks | 452 |

| | | |
|-------|------------------------------------------------------------------------------|-----|
| 8.6 | Genetic Algorithm–Assisted Supervised Learning | 454 |
| 8.6.1 | Hybrid GA/Gradient-Descent Method for Feedforward Multilayer Net Training | 455 |
| 8.6.2 | Simulations | 458 |
| 8.7 | Summary | 462 |
| | Problems | 463 |
| | References | 469 |
| | Index | 501 |