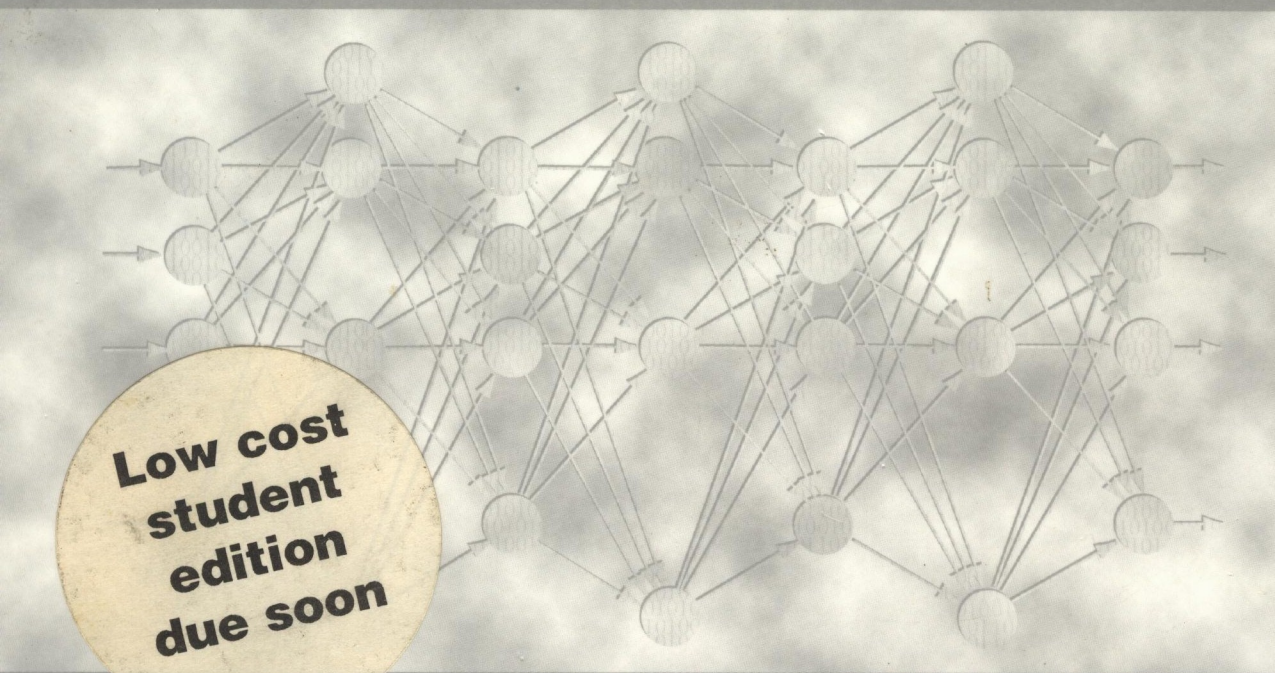
ARTIFICIAL NEURAL NETWORKS



Low cost student edition due soon

Robert J. Schalkoff

TABLE OF CONTENTS

Preface xix

1 Overview: Artificial Neural Networks and Neural Computing 1

1.1 What is Neural Computing? 1
1.1.1 Computing Architectures / 1.1.2 Chapter Overview / 1.1.3 Definition of Artificial Neural Network / 1.1.4 Fundamental Neural Network Concepts / 1.1.5 Introductory Terminology and Notational Conventions

1.2 Neural Computing Applications 6
1.2.1 Characteristics of Problems Suitable for ANNs / 1.2.2 Sample ANN Applications

1.3 A Brief Overview of Neural Computing 7
1.3.1 Background / 1.3.2 What Are the Relevant Computational Properties of the Human Brain? / 1.3.3 Neural Approaches to Computation / 1.3.4 Advantages and Disadvantages of ANNs

1.4 Engineering Approaches to Neural Computing 10
1.4.1 Initial Questions / 1.4.2 Neural Engineering Procedures: Replacing Design with Training / 1.4.3 Procedures for ANN System Engineering

1.5 ANNs: The Mappings Viewpoint 12
1.5.1 The Basic Perceptual System and Stimulus-Response Approaches / 1.5.2 Network Inputs and Outputs / 1.5.3 Vector Representations for S-R Characteristics / 1.5.4 Parameters, Weights, and Constraints

1.6 ANNs: The Structure Viewpoint 17
1.6.1 ANN Functions / 1.6.2 Neural Network Structure / 1.6.3 Network Topologies and Characterization / 1.6.4 Interconnection Complexity and Problem Scale / 1.6.5 Feedback Interconnections and Network Stability / 1.6.6 Combinations of Nets and Variable Topologies

1.7 ANN Learning Approaches 22
1.7.1 Training Sets and Test Sets / 1.7.2 Generalization / 1.7.3 Learning Curves / 1.7.4 Error Measures and Error Trajectories

1.8 Relationship of ANNs to Other Technologies 25

1.9 Historical Efforts 27
1.9.1 Perceptron and Earlier / 1.9.2 Post-Perceptron / 1.9.3 The Third Generation of ANNs / 1.9.4 Future Directions and Open Issues

1.10 Overview of ANN Literature and Resources 30
1.10.1 Books / 1.10.2 Journals / 1.10.3 Conferences / 1.10.4 Internet Resources

1.11 Overview of This Book 32

References 32

2 Mathematical Fundamentals for ANN Study 36

2.1 Vector and Matrix Fundamentals 36
2.1.1 Elementary Matrices / 2.1.2 Vectors / 2.1.3 Linearity / 2.1.4 Inner and Outer Products and Applications / 2.1.5 Measures of Similarity in Vector Space / 2.1.6 Differentiation of Matrices and Vectors / 2.1.7 The Chain Rule / 2.1.8 Multidimensional Taylor Series Expansions / 2.1.9 The Pseudoinverse of a Matrix and Least Squares Techniques (Deterministic) / 2.1.10 Eigenvalues and Eigenvectors

2.2 Geometry for State-Space Visualization 50
2.2.1 Geometric Interpretation of ANN Mappings / 2.2.2 Hypercubes / 2.2.3 ANN Mappings, Decision Regions and Boundaries, and Discriminant Functions / 2.2.4 Quadric Surfaces and Boundaries

2.3 Optimization 55
2.3.1 Gradient Descent–Based Procedures / 2.3.2 Error Function Contours and Trajectories

2.4 Graphs and Digraphs 60

2.5 Bibliography 61

References 61

3 Elementary ANN Building Blocks 62

3.1 Overview and Objectives 62

3.2 Biological Neural Units 62
3.2.1 Physical (Biological) Neurons / 3.2.2 The Scale of Biological Systems / 3.2.3 Biophysical Mechanisms and Equivalent Neural Operations / 3.2.4 Neural System Hierarchies and Examples

3.3 Artificial Unit Structures 70
3.3.1 Linear Unit Structures / 3.3.2 Generalizing the Unit Model / 3.3.3 Two-Part Unit Models: Activation and Squashing / 3.3.4 McCulloch-Pitts (MP) Units / 3.3.5 Threshold Logic with Weighted Linear Input Combination

3.4 Unit Net Activation to Output Characteristics 77
3.4.1 Activation Functions and Squashing / 3.4.2 The Sigmoid (Logistic) Squashing Function / 3.4.3 Other Squashing Functions / 3.4.4 Exceptions to the Two-Part Model / 3.4.5 "Memory" or Individual Unit Activation Dynamics

3.5 Artificial Unit Model Extensions 86
3.5.1 Adding (an Optional) Bias to the Artificial Neuron Model / 3.5.2 Inhibitory Inputs / 3.5.3 Individual Unit Dynamics versus Network Dynamics / 3.5.4 Activation Functions and Network Training

3.6 Bibliography 88
References 88
Problems 89

4 Single-Unit Mappings and the Perceptron 93

4.1 Introduction 93

4.2 Linear Separability 93
4.2.1 Probability of a Linearly Separable Problem / 4.2.2 Linear Programming and Linear Separability / 4.2.3 Examples / 4.2.4 Alternative Constraint on a Nonlinearly Separable H

4.3 Techniques to Directly Obtain Linear Unit Parameters 103
4.3.1 Batch (Pseudoinverse) Solution to Single-Unit Design / 4.3.2 Iterative (Gradient Descent) Solution Procedures

4.4 Perceptrons and Adaline/Madaline Units and Networks 105
4.4.1 Perceptron/Adaline Overview / 4.4.2 α-LMS Algorithms

4.5 Multilayer Perceptrons (MLPs) 109

4.6 Gradient Descent Training Using Sigmoidal Activation Functions 112

4.7 Bibliography 113
References 114
Problems 115

5 Introduction to Neural Mappings and Pattern Associator Applications 120

5.1 Neural Network-Based Pattern Associators 120
5.1.1 Black-Box Pattern Associator Structure / 5.1.2 Desirable Pattern Associator Properties / 5.1.3 Autocorrelator versus Heterocorrelator Structures

5.2 The Influence of Psychology on PA Design and Evaluation 122
5.2.1 Discrimination, Association, and Principles of Connection / 5.2.2 Relevant Principles of Association

5.3 Linear Associative Mappings, Training, and Examples 125
5.3.1 An Elementary Linear Network Structure and Mathematical Representation / 5.3.2 Training a Single Layer of Units / 5.3.3 Multiple Layers (Linear Units) / 5.3.4 Hetero-associators Directly from Generalized Inverses / 5.3.5 Extended Examples: Hetero-associative Memory Design / 5.3.6 Training Nonlinear Mappings Using Linear Solution Techniques

5.4 Hebbian or Correlation-Based Learning 138
5.4.1 A Storage Prescription for a Single Pattern / 5.4.2 Assessment of Storage Properties / 5.4.3 Alternative Formulation for the Hebbian Prescription / 5.4.4 Hebbian Example 1: Problem Formulation for Ternary-Valued Output (Single Pattern Pair) / 5.4.5 Example 2: Three-Input, Three-Output PA Using Hebbian Training

5.5 Bibliography 142
References 142
Problems 143

6 Feedforward Networks and Training: Part 1 146

6.1 Multilayer Feedforward Network Structure 146
6.1.1 Introduction / 6.1.2 Feedforward Structures

6.2 The Delta Rule and Generalized Delta Rule 151
6.2.1 Overview of the Generalized Delta Rule / 6.2.2 Derivation of the Delta Rule / 6.2.3 Extension of the DR for Hidden Units: The Generalized Delta Rule / 6.2.4 Pattern Presentation and Weight-Updating Strategies

6.3 Architecture and Training Extensions 162
6.3.1 Error Trajectories during Training / 6.3.2 Adding Momentum to the Training Procedure / 6.3.3 Training the Unit Bias Inputs / 6.3.4 Extension of the GDR to Jump-ahead or Shortcut Connections

6.4 Ramifications of Hidden Units 166
6.4.1 Hidden Layers and Network Mapping Ability / 6.4.2 Training Time and Mapping Accuracy

6.5 General Multilayer FF Network Mapping Capability 167
6.5.1 Exact Mappings, Approximations, and Limiting Cases / 6.5.2 Boolean Functions / 6.5.3 A Fourier Series Approach / 6.5.4 An Intuitive Argument for the Approximation Capability of a Two-Hidden-Layer Network / 6.5.5 Approximation by Superpositions of Sigmoid Functions / 6.5.6 The Stone-Weierstrass Theorem / 6.5.7 Kolmogorov's Mapping Neural Network Existence Theorem / 6.5.8 FF ANNs as Universal Approximators

6.6 Examples of FF Network Design 175
6.6.1 XOR Logic Function / 6.6.2 The (Exactly) 2-of-3 Detector / 6.6.3 Network Design for Recognition of Digits (and Extensions)

6.7 Bibliography 182

References 182

Problems 183

7 Feedforward Networks, Part 2: Extensions and Advanced Topics 189

7.1 Feedforward Pattern Associator Design: Achieving Desired Mappings 189
7.1.1 Solutions: Existence vs. Realization / 7.1.2 Combinations of Training Algorithms

7.2 Weight Space, Error Surfaces, and Search 190
7.2.1 Weight Space / 7.2.2 The Search for $\underline{w}$ / 7.2.3 Error Surface / 7.2.4 GDR Starting Points / 7.2.5 Effect of the Error Surface on Training Algorithms / 7.2.6 Symmetries in Weight Space / 7.2.7 Characteristics of a Training Algorithm

7.3 Generalization 194
7.3.1 Objective / 7.3.2 The Test Set, S_T / 7.3.3 Cross-Validation for Generalization Prediction

7.4 Non-Euclidean (Output) Error Norms 196
7.4.1 Cost Functions / 7.4.2 Weighted L_2 Norm / 7.4.3 Other L_p, $p \neq 2$, Norms / 7.4.4 Modification of the GDR/Backpropagation Equations for the L_∞ Norm

7.5 Higher-Order Derivative–Based Training 199
7.5.1 Background and Error Formulation / 7.5.2 Taylor Series Expansions of E / 7.5.3 First-Order Methods / 7.5.4 Second-Order Approaches / 7.5.5 Computational Cost of Higher-Order Approaches

7.6 Stochastic Optimization for Weight Determination 207
7.6.1 Random Elements of Weight Determination Algorithms / 7.6.2 Objective Function and Region of Search / 7.6.3 The Random Optimization Method (ROM) / 7.6.4 Example: Four-Input Parity Detector

7.7 The Network Architecture Determination Problem 209
7.7.1 Ontogenic Neural Networks / 7.7.2 Partition Complexity / 7.7.3 Constructive vs. Destructive Network Topology Modification / 7.7.4 Weight Costs and Minimizations

7.8 Genetic Algorithms for Network Training 211
7.8.1 Introduction to Genetic Algorithms / 7.8.2 Critical Aspects of a Genetic Programming Solution / 7.8.3 Genetic Algorithm Simulation Parameters / 7.8.4 Topology Optimization and/or

Weight Optimization / 7.8.5 Application to ANN Design and Training

7.9 Cascade Correlation Networks and Algorithms 219
7.9.1 Overview / 7.9.2 Major Features of the CC Approach / 7.9.3 CC Architecture Characteristics / 7.9.4 CC Applications and Assessment

7.10 Network Minimization 223
7.10.1 Motivation / 7.10.2 Pruning / 7.10.3 Determining Superfluous Units / 7.10.4 Pruning by Weight Decay / 7.10.5 Sensitivity-Based Pruning

7.11 Network Inversion 231
7.11.1 Methodology / 7.11.2 Relation of Inversion to Backpropagation / 7.11.3 Example: Character Recognition Revisited (and Inverted)

7.12 Bibliography 236

References 241

Problems 243

8 Recurrent Networks 246

8.1 Introduction 246
8.1.1 Recurrent Networks as Mapping Networks / 8.1.2 E (Error) versus E (Energy) / 8.1.3 Review of Nonlinear Feedback Systems

8.2 Basic Parameters and Recurrent Network Design 252
8.2.1 Network Parameters / 8.2.2 Network Dynamics

8.3 Weight Storage Prescription and Network Capacity 254
8.3.1 Weight Prescriptions / 8.3.2 Network Capacity Estimation

8.4 Network Synthesis Procedures and Examples 257
8.4.1 General Design Procedure / 8.4.2 Design of a Simple Hopfield Network: Storing and Accessing Stable States / 8.4.3 CAM Example: Association of Simple 2-D Patterns / 8.4.4 Network Parameters and Effects

8.5 Energy Function Characterization 267
8.5.1 Energy Analysis / 8.5.2 Example: $d = 4$ Network Energy Function / 8.5.3 Eigenvalue-Eigenvector Analysis of W and Relation to E / 8.5.4 Alternative Recurrent Network Synthesis Strategies: Training Recurrent Networks Using FF Networks with Feedback / 8.5.5 Summary of the Generalized Hopfield Network Equations

8.6 Constraint Satisfaction and Optimization Applications 281
8.6.1 Applying Hopfield Nets / 8.6.2 Mapping Constraints and Objectives into Recurrent Networks

8.7 Bidirectional Associative Memory 285
8.7.1 BAM Examples / 8.7.2 BAM as Partitioned Bipolar Hopfield and Other Connection Matrices
8.8 Relating Feedforward and Recurrent Networks 291
8.9 Nondeterministic Network Formulations 294
8.9.1 Introduction / 8.9.2 Unit Probabilistic Characterizations / 8.9.3 Simulated Annealing / 8.9.4 The Metropolis Algorithm / 8.9.5 Boltzmann Machines
8.10 Bibliography 298
References 299
Problems 302

9 Competitive and Self-Organizing Networks 308
9.1 Introduction 308
9.2 Formal Characterization and General Clustering Procedures 309
9.2.1 Clustering Similarity Measures / 9.2.2 Clustering Complexity / 9.2.3 c-Means Algorithm / 9.2.4 Learning Vector Quantization (LVQ) / 9.2.5 Other Clustering Strategies
9.3 Competitive Learning Architectures and Algorithms 314
9.3.1 Examples
9.4 Self-Organizing Feature Maps 316
9.4.1 Unit Topologies / 9.4.2 Defining Topological Neighborhoods / 9.4.3 Network Learning Algorithm / 9.4.4 Network Coordinate Systems: Topological and Weight Spaces / 9.4.5 Algorithm Properties and Discussion / 9.4.6 Sample Results: 2-D Input Space / 9.4.7 Sample Results: Higher-Dimensional Input Spaces
9.5 Adaptive Resonance Architectures 325
9.5.1 Background / 9.5.2 Overall Structure / 9.5.3 Neural-Layer Structure Details / 9.5.4 Network Dynamics / 9.5.5 Algorithm Specifics and Equations / 9.5.6 Sample Results
9.6 Bibliography 331
References 332
Problems 334

10 Radial Basis Function (RBF) Networks and Time Delay Neural Networks (TDNNs) 337
10.1 Introduction 337
10.2 Radial Basis Function (RBF) Networks 337
10.2.1 Introduction / 10.2.2 RBF Network Structure / 10.2.3 RBF Unit Characteristics / 10.2.4 Basis Function

Interpretation / 10.2.5 RBF Network Design and Training / 10.2.6 RBF Applications / 10.2.7 RBF Design and Application Examples

10.3 Time Delay Neural Nets (TDNNs) 344
10.3.1 Introduction / 10.3.2 The Concept of Invariant Recognition / 10.3.3 Shift-Invariant (SI) Recognition and Properties / 10.3.4 Temporal Significance of Inputs / 10.3.5 The TDNN Concept / 10.3.6 Input and Unit Representations and Equivalences / 10.3.7 Evolution of a Single-Unit TDNN Structure / 10.3.8 TDNN Layered Network Structure: Extension of the Single-Unit Model into One or More Layers of Delayed Units / 10.3.9 TDNN Training

10.4 Bibliography 353
References 355
Problems 356

11 Fuzzy Neural Networks Including Fuzzy Sets and Logic and ANN Implementations 360

11.1 Introduction to Neuro-Fuzzy Systems 360
11.1.1 Fuzzy Systems and Neural Nets / 11.1.2 Sample Applications

11.2 Fuzzy Sets and Logic Background 361
11.2.1 Representing and Manipulating Uncertainty / 11.2.2 Types of Fuzzy Neural Nets / 11.2.3 Fundamental Fuzzy System Concepts and Fuzzy Logic

11.3 Fuzzy System Design Procedures 370
11.3.1 Main Tasks / 11.3.2 Applying Compositional Rules of Inference (CRIs) in Fuzzy Systems / 11.3.3 Examples of Combined Fuzzy CRI and Defuzzification Strategies / 11.3.4 Fuzzy Propeller Speed Controller Example

11.4 Fuzzy/ANN Design and Implementation 376
11.4.1 Neuro-Fuzzy Systems / 11.4.2 A Fuzzy MP Neuron Unit / 11.4.3 Neural Components of a Fuzzy System / 11.4.4 ANN-Implementable Approximations / 11.4.5 Examples of Fuzzy Neural Network Controllers / 11.4.6 Expert System Implementations

11.5 Bibliography 382
References 383
Problems 384

12 ANN Hardware and Implementation Concerns 386

12.1 Practical ANN Implementation 386
12.1.1 Background / 12.1.2 Theoretical and Practical Concerns / 12.1.3 Fundamental ANN Computations /

12.1.4 Emulation of Neural Architecture (Topology) versus Emulation of Computation (Functionality) / 12.1.5 Hardware Modularity, Flexibility, and Scaling / 12.1.6 Summary of Implementation-Related Concerns

12.2 Related Elements of Computer Architecture 392
12.2.1 Definitions and Salient Concepts / 12.2.2 Classes of Parallelism and Architectural Partitions / 12.2.3 Architectural Performance

12.3 Numerical Accuracy, Weight Resolution, and Fault Tolerance 397

12.4 Hardware Realization of ANNs 398
12.4.1 Existing and Previous Chip/Device Designs / 12.4.2 Array Approaches / 12.4.3 VLSI and ULSI Approaches / 12.4.4 Optical Techniques

12.5 Bibliography 412

References 413

Index 417

Permissions 421