

Digital Neural Networks



73.918
K95

Digital Neural Networks

S. Y. Kung

*Department of Electrical Engineering
Princeton University*



9550041



*PTR Prentice Hall
Englewood Cliffs, New Jersey 07632*

9550041

Library of Congress Cataloging-in-Publication Data

Kung, S. Y. (Sun Yuan)
Digital neural networks / S. Y. Kung.
p. cm. -- (Prentice-Hall information and system sciences series)

Includes bibliographical references and index.

ISBN 0-13-612326-0

1. Neural networks (Computer science) I. Title. II. Series.

QA76.87.K86 1993

006.3--dc20

92-42737

CIP

Editorial/production supervision: *Kim Gueterman*

Acquisitions editor: *Karen Gettman*

Cover design: *George Cornell*

Buyer: *Mary Elizabeth McCartney*

Copy editor: *Peter J. Zurita*

Series logo design: *A. M. Bruckstein*



©1993 by PTR Prentice-Hall, Inc.
A Simon & Schuster Company
Englewood Cliffs, New Jersey 07632

ED92/12

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-612326-0

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Digital Neural Networks



**PRENTICE HALL INFORMATION
AND SYSTEM SCIENCES SERIES**

Thomas Kailath, Editor

- ANDERSON & MOORE *Optimal Control: Linear Quadratic Methods*
ANDERSON & MOORE *Optimal Filtering*
ASTROM & WIT TENMARK *Computer-Controlled Systems: Theory and Design, second edition*
BOYD & BARRATT *Linear Controller Design: Limits of Performance*
GARDNER *Statistical Spectral Analysis: A Nonprobabilistic Theory*
GOODWIN & SIN *Adaptive Filtering, Prediction, and Control*
GRAY & DAVISSON *Random Processes: A Mathematical Approach for Engineers*
HAYKIN *Adaptive Filter Theory*
JAIN *Fundamentals of Digital Image Processing*
JAMSHIDI, TAROKH, & SHAFAI *Computer-Aided Analysis and Design of Linear Control Systems*
JOHNSON *Lectures on Adaptive Parameter Estimation*
KAILATH *Linear Systems*
KUMAR & VARAIYA *Stochastic Systems*
KUNG *Digital Neural Networks*
KUNG *VLSI Array Processors*
KUNG, WHITEHOUSE, & KAILATH, EDS. *VLSI and Modern Signal Processing*
KWAKERNAAK & SIVAN *Signals & Systems*
LANDAU *System Identification and Control Design Using P.I.M. + Software*
LJUNG *System Identification: Theory for the User*
MACOVSKI *Medical Imaging Systems*
MELSA & SAGE *An Introduction to Probability and Stochastic Processes*
MIDDLETON & GOODWIN *Digital Control & Estimation*
NARENDRA & ANNASWAMY *Stable Adaptive Systems*
SASTRY & BODSON *Adaptive Control: Stability, Convergence, and Robustness*
SOLIMAN & SRINATH *Continuous and Discrete Signals and Systems*
SPILKER *Digital Communications by Satellite*
WILLIAMS *Designing Digital Filters*

Preface

Unification. — Confucius, 551-479 B.C

The field of *digital neural networks* brings together scientists who study how the brain works and engineers who build highly parallel and intelligent supercomputers. The main objective of this book is to provide a synergistic and systematic exploration of several fundamental issues on *digital neural networks*. The word *digital* refers to *discrete-time* digital processing systems. The term *neural networks* embodies information processing systems using a combination of *nonlinear, adaptive, network, and parallel processing* technologies. Digital neural networks provide a natural link between the rapid advance of (VLSI) microelectronics and parallel processing technology and the increasing need for intelligent processing techniques. This represents a truly interdisciplinary research field which will ultimately lead to the future intelligent processing systems.

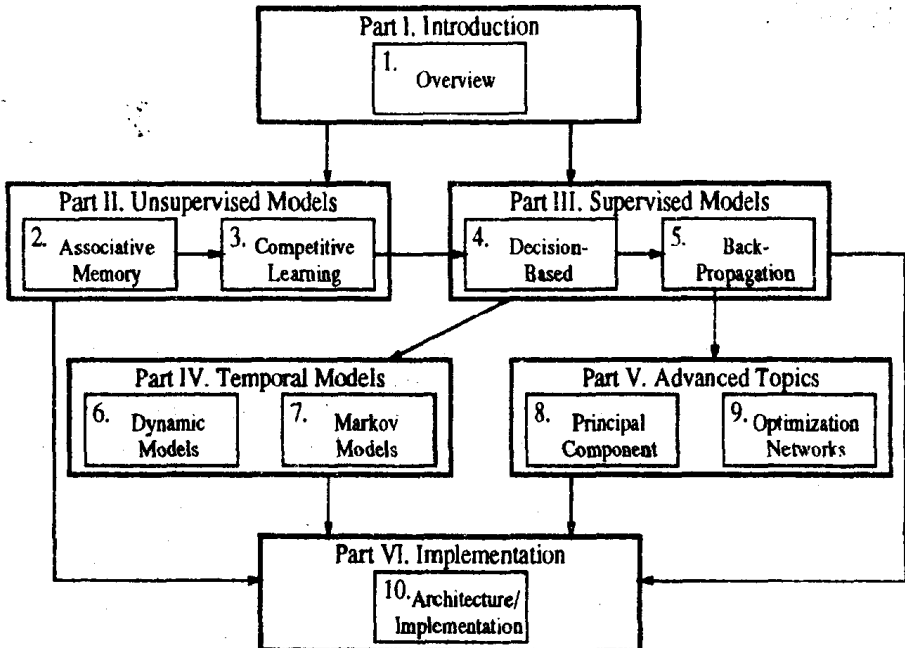
Many excellent reference books on neural networks exist; some place emphasis on the biological view, and others take a physics perspective. In contrast, this book is intended to be a textbook prepared for the discipline of information science and electrical engineering.

The book contains six parts:

- **Part I: Introduction** (Chapter 1);
- **Part II: Unsupervised Models** — associative memory (Chapter 2) and competitive learning networks (Chapter 3);
- **Part III: Supervised Models** — decision and approximation/optimization based neural networks (Chapters 4 & 5);

- **Part IV: Temporal Models** — deterministic and stochastic temporal networks (Chapters 6 & 7);
- **Part V: Advanced Topics** — principal component neural networks (Chapter 8) and stochastic annealing networks (Chapter 9);
- **Part VI: Implementation** (Chapter 10).

The following road map serves as a guide for teaching or reading.



The topics covered in Chapters 1 - 5 and Chapter 9 follow a rather orthodox neural network presentation, while the topics addressed in Chapters 6 through 8 are somewhat unorthodox in that they stem from the conventional signal processing and system theoretical perspectives. Chapter 10 provides a methodology for mapping neural models to parallel architecture and implementation.

Some unique features of the book:

- It aims at a coherent treatment which links the orthodox and unorthodox approaches to neural models. For this, we introduce what can be called *neural network system theory*, which unifies many diversified neural models.

- It employs an application-driven perspective, particularly signal/image processing and pattern recognition applications. Moreover, it goes on to present an integrated study of the application, algorithm, and architecture aspects of neural networks.
- It provides a working knowledge of various neural models, including theoretical basis, potential applications, and digital implementations. With this cross-disciplinary background, the reader would be better equipped to tackle research and application projects related to neural information processing.

This book is intended to be a textbook for a one-semester graduate or undergraduate course. (It is assumed that the reader has the basic background of advanced calculus, linear system theory, and signal processing. Otherwise, some mathematical treatments in some sections may be too advanced for uninitiated students.) Lectures for an undergraduate or graduate class can be organized from the following recommended sections:

- Chapters 1 - 7, (except Section 5.4)
- Sections 8.2, 8.3, 9.2, and 10.2.

ACKNOWLEDGMENTS

Without enthusiastic support and encouragement from so many friends, it would be impossible to attempt a book on such a dynamic and diversified subject. First, my thanks go to the researchers who have contributed to this rich research literature of neural networks. Their works have provided the foundation for the book. Also, I must apologize for having overlooked many critical topics and references in bibliographical research due to time constraints.

This book is an outgrowth of many semesters' lectures on the subject. I am very much indebted to many of my in-class students for their invaluable questions and contributions on examples and exercises. The advanced part of the book is largely extracted from several Ph. D. dissertations under my supervision. I wish to thank all of them: Drs. Kostas Diamantaras, Mark Goudreau, Jen-Neng Hwang, Wei-Dong Mao, John Vlontzos, and especially Mr. Wei-Hsin Chou and Mr. Jin-Shiuh Taur. They deserve to be named co-authors of the book.

I have benefited greatly the frequent and enlightening exchange and collaboration with many of my colleagues. To name a few: Drs. Alice Chiang, Frank Fallside, Lee Giles, Robert Hecht-Nielsen, Yuhon Hu, Fred Juang, Teuvo Kohonen, Gary Kuhn, H.C. Lee, Lenart Ljung, Christoph von der Malsburg, Alastair McAulay, John Moody, Ulrich Ramacher, and John Sorensen. Their enthusiastic participation in this project has made this prolonged process more bearable and indeed very enjoyable and rewarding. My deep gratitude goes to the reviewers for their critical and constructive suggestions on many drafts of the book. They are Drs. Joshua Alspector, Brad Dickinson, Tony Kuh, Guanghan Xu, Mr. Jerry Murphy, and other anonymous reviewers. My sincere

appreciation goes to my professor and the series editor, Professor Tom Kailath, for many years of inspiration and friendship.

I have been very fortunate to work with Ms. Karen Gettman and Ms. Kim Gueterman at Prentice Hall, who have provided very professional assistance throughout this project. I am grateful to the Department of Electrical Engineering, Princeton University, for making available a very scholarly environment for both teaching and research. I also want to thank the secretarial staff for their tireless and cheerful assistance during this long process: Susan Gafgen, Kim Hegelbach, Emmalee Morrison, and Lynn O'Shaughnessy. Finally, I am pleased to acknowledge that some work presented in this book was supported in part by Grants and Contracts from National Science Foundation, Air Force Office of Scientific Research, and Defense Advanced Research Project Agency.

Princeton, New Jersey

S.Y. Kung

Contents

Preface

xv

PART I INTRODUCTION

1	Overview	1
1.1	Introduction	1
1.1.1	Biological-Type Neural Networks	2
1.1.2	Application-Driven Neural Networks	4
1.2	Applications, Algorithms, and Architectures	6
1.2.1	Application Paradigms of Neural Models	7
1.2.2	Algorithmic Study on Neural Networks	11
1.2.3	Architectures of Neural Networks	14
1.2.4	Total Information Processing Systems	16
1.2.5	Representation and Feature Extraction	21
1.3	Taxonomy of Neural Networks	24
1.3.1	Supervised and Unsupervised Networks	25
1.3.2	Basis Function and Activation Function	27
1.3.3	Structures of Neural Networks	30
1.3.4	Mutual and Individual Training Strategies	33
1.3.5	Static and Temporal Pattern Recognitions	36
1.3.6	Decision and Approximation/Optimization Formulations	38
1.4	Concluding Remarks	39
1.5	Problems	40

PART II UNSUPERVISED MODELS		
2	Fixed-Weight Associative Memory Networks	43
2.1	Introduction	43
2.2	Feedforward Associative Memory Networks	44
2.2.1	Linear Associative Memory	44
2.2.2	Nonlinear Associative Memory for Holographic Retrieval	48
2.2.3	Hamming Networks	50
2.3	Feedback Associative Memory Networks	51
2.3.1	Sequential (Asynchronous) Hopfield Model	52
2.3.2	Parallel (Synchronous) Hopfield Model	57
2.3.3	Capacities of Hopfield and Hamming Networks	59
2.3.4	Bidirectional Associative Memory	62
2.3.5	Discrete-Time Continuous-State Hopfield Model	64
2.3.6	Stochastic Models and the Boltzmann Machine	66
2.4	Concluding Remarks	66
2.5	Problems	67
3	Competitive Learning Networks	73
3.1	Introduction	73
3.2	Basic Competitive Learning Networks	74
3.2.1	Minimal Learning Model	75
3.2.2	Training Rules Based on Normalized Weights	76
3.2.3	Training Rules for Leaky Learning	78
3.3	Adaptive Clustering Techniques: VQ and ART	78
3.3.1	Vector Quantizer (VQ)	78
3.3.2	Binary-Valued ART	79
3.3.3	Continuous-Valued ART	80
3.4	Self-Organizing Feature Map: Sensitivity to Neighborhood and History	85
3.4.1	Self-Organizing Feature Map	85
3.4.2	Competitive Learning with History Sensitivity	87
3.5	Neocognitron: Hierarchically Structured Model	91
3.6	Concluding Remarks	95
3.7	Problems	95

PART III SUPERVISED MODELS

4	Decision-Based Neural Networks	99
4.1	Introduction	99
4.2	Linear Perceptron Networks	104
4.2.1	Linear Perceptron for Binary Classification	105
4.2.2	Linear Perceptron for Multiple Classification	110
4.3	Decision-Based Neural Networks	115
4.3.1	Decision-Based Learning Rule	116
4.3.2	Hierarchical DBNN Structure	118
4.3.3	Static and Temporal DBNNs	122
4.3.4	Fuzzy Decision Neural Networks	124
4.4	Applications to Signal/Image Classifications	130
4.4.1	Texture Classification	130
4.4.2	Optical Character Recognition (OCR) Application	132
4.4.3	DTW Temporal Networks for Speech Recognition	133
4.5	Concluding Remarks	139
4.6	Problems	140
5	Approximation/Optimization Neural Networks	145
5.1	Introduction	145
5.2	Linear Approximation Networks	148
5.2.1	Delta Learning Rule: ADALINE	149
5.2.2	Kaczmarz Projection Method and Learning Rates	150
5.3	Nonlinear Multilayer Back-Propagation Networks	152
5.3.1	Back-Propagation Algorithm	154
5.3.2	Numerical Back-Propagation Methods	156
5.4	Training Versus Generalization Performance	167
5.4.1	Approximation Analysis and Training Performance	168
5.4.2	Generalization Performance	171
5.4.3	RBF Regularization Networks	175
5.4.4	Network Pruning and Growing Techniques	179
5.5	Applications of Back-Propagation Networks	184
5.5.1	Approximation Formulation: Image Compression	185
5.5.2	Optimization Formulation: Surface Reconstruction	186
5.5.3	OCON Applications: OCR, Speech, and Texture	187

5.5.4	Control Application: Robotic Path Control	190
5.6	Concluding Remarks	192
5.7	Problems	194

PART IV TEMPORAL MODELS

6	Deterministic Temporal Neural Networks	203
6.1	Introduction	203
6.2	Linear Temporal Dynamic Models	205
6.2.1	Non-recursive Linear Predictive Filter	207
6.2.2	Recursive (IIR) Adaptive Filters	208
6.2.3	State-Space Representations	209
6.3	Nonlinear Temporal Dynamic Models	215
6.3.1	Nonrecurrent Temporal Dynamic Model (TDNN)	215
6.3.2	Recurrent Neural Networks (RNN)	219
6.3.3	Hamiltonian Theoretical Approach to General Learning	222
6.4	Prediction-Based Temporal Networks	224
6.4.1	Prediction-Based Independent Training Model	225
6.4.2	Compare Static and Temporal Models for ECG Analysis	226
6.5	Concluding Remarks	233
6.6	Problems	234
7	Stochastic Temporal Networks: Hidden Markov Models	237
7.1	Introduction	237
7.2	From Markov Model to Hidden Markov Model	239
7.2.1	Markov Model	239
7.2.2	Hidden Markov Model	242
7.3	Learning Phase of Hidden Markov Models	245
7.3.1	Learning Formulas for MM	245
7.3.2	BP-Type HMM Learning Rule	247
7.4	Retrieving Phase of Hidden Markov Models	253
7.4.1	Viterbi Algorithm for Model-Scoring Problem	254
7.4.2	Pattern-Completion Problem	257
7.5	Applications to Speech, ECG, and Character Recognition	258

7.5.1	Application to Isolated Digit Recognition	258
7.5.2	Application to ECG Recognition	258
7.5.3	Application to Optical Character Recognition	259
7.6	Concluding Remarks	266
7.7	Problems	267

PART V ADVANCED TOPICS

8	Principal Component Neural Networks	269
8.1	Introduction	269
8.2	From Wiener Filtering to PCA	271
8.2.1	Symmetric Principal Component Analysis (PCA)	273
8.2.2	Asymmetric Principal Component Analysis	274
8.3	Symmetric Principal Component Analysis	277
8.3.1	Extraction of the Single Principal Component	277
8.3.2	Multiple Principal Components: Lateral Orthogonalization Network	281
8.4	BP Network for Asymmetric PCA Problems	289
8.4.1	Extraction of the Single Principal Component	289
8.4.2	Lateral Networks for Multiple Components	290
8.5	Applications to Signal/Image Processing	296
8.5.1	Rotational Compensation Applications	297
8.5.2	Data-Compression Applications: Extraction of Innovative Components	297
8.5.3	Signal-Separation Applications: Separate Signal from Noise	298
8.5.4	Sinusoidal Retrieval Applications: Harmonic Spectrum Analysis	300
8.5.5	Signal-Restoration Applications: Remove or Cancel Unwanted Noises	301
8.5.6	Subspace Classification Applications	301
8.6	Concluding Remarks	304
8.7	Problems	305
9	Stochastic Annealing Networks for Optimization	311
9.1	Introduction	311

9.2	Stochastic Neural Networks	313
9.2.1	Equilibrium in Stochastic Networks	313
9.2.2	Stochastic Simulated Annealing (SSA)	315
9.2.3	Mean-Field Annealing and the Continuous-Valued Hopfield Model	317
9.3	Applications to Combinatorial Optimization and Image Restoration	320
9.3.1	Combinatorial Optimization	321
9.3.2	Image-Restoration Model: Markov Random Field (MRF)	325
9.4	Boltzmann Machine	328
9.4.1	Training Phase	329
9.4.2	Retrieving Phases	332
9.5	Concluding Remarks	334
9.6	Problems	334

PART VI IMPLEMENTATION

10	Architecture and Implementation	337
10.1	Introduction	337
10.2	Mapping Neural Nets to Array Architectures	340
10.2.1	Mapping Design Methodology	340
10.2.2	Design for Multilayer Networks: Retrieving Phase	345
10.2.3	Design for Multilayer Networks: Training Phase	350
10.2.4	System Design and Simulation	362
10.3	Dedicated Neural Processing Circuits	368
10.3.1	Analog Electronic Circuits	371
10.3.2	Digital (or Mostly Digital) ASIC Chips	374
10.3.3	Digital Design Based on FPGA Chips	376
10.3.4	Optical Implementation of Neural Nets	378
10.4	General-Purpose Digital Neurocomputers	380
10.4.1	Overall System Configuration	381
10.4.2	Processor Element Architecture	384
10.4.3	Parallel Array Architectures	392

10.5 Concluding Remarks

405

10.6 Problems

406

Bibliography

413

Index

435

PART I: INTRODUCTION

Chapter 1

Overview

1.1 Introduction

Neural networks have become a very popular field of research in cognitive science, neurobiology, computer engineering/science, signal processing, optics, and physics. They represent a very broad range of neural processing models. The purpose of this book is to provide an integrated and cohesive exploration of the fundamental issues on *digital neural networks*. The word “digital” refers to “discrete-time” processing. In terms of data format, it may involve discrete-value or continuous-value signals. The term “neural networks” is characterized by a combined adaptive network and parallel processing technologies. Therefore, the field requires a coherent study and understanding of multi-disciplinary issues, including application-needs, neural models, and parallel processing.

An artificial neural network is an abstract simulation of a real nervous system that contains a collection of *neuron units* communicating with each other via *axon connections*. Such a model bears a strong resemblance to axons and dendrites in a nervous system. Due to its self-organizing and adaptive