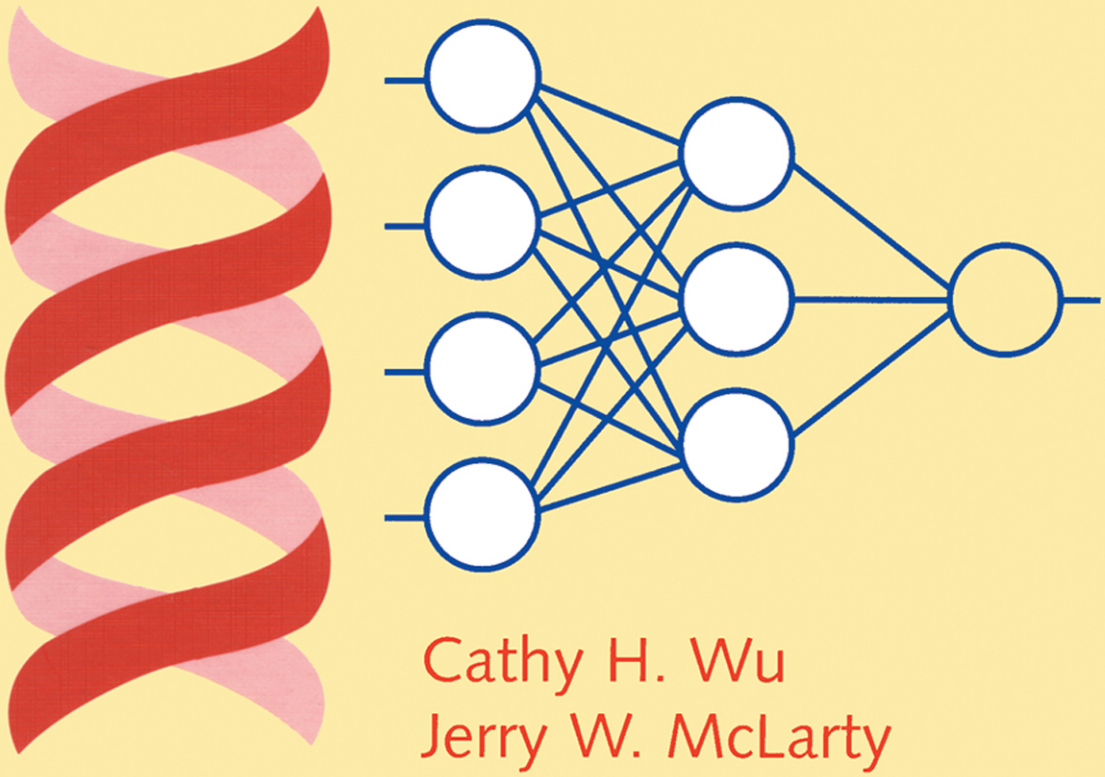
METHODS IN COMPUTATIONAL BIOLOGY AND BIOCHEMISTRY

Volume 1

SERIES EDITOR: A.K. KONOPKA

Neural Networks and Genome Informatics



Cathy H. Wu
Jerry W. McLarty

ELSEVIER

Neural Networks and Genome Informatics

METHODS IN COMPUTATIONAL BIOLOGY AND BIOCHEMISTRY

Volume 1

Series Editor

A.K. KONOPKA

Maryland, USA

ELSEVIER

Amsterdam - Lausanne - New York - Oxford - Shannon - Singapore - Tokyo

Neural Networks and Genome Informatics

C.H. Wu
J.W. McLarty

University of Texas Health Center at Tyler
Department of Epidemiology and Biomathematics
11937 U.S. Highway 271
Tyler, TX 75708-3154
USA

2000
ELSEVIER
Amsterdam - Lausanne - New York - Oxford - Shannon - Singapore - Tokyo

ELSEVIER SCIENCE Ltd
The Boulevard, Langford Lane
Kidlington, Oxford OX5 1GB, UK

First edition 2000

Library of Congress Cataloging in Publication Data
A catalog record from the Library of Congress has been applied for.

British Library Cataloguing in Publication Data
A catalogue record from the British Library has been applied for.

ISBN: 0 08 042800 2

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in The Netherlands.

This book is dedicated to:

My late father, my husband, daughter, and son, for their inspirations. – CHW.

The late Allen B. Cohen, whose passion for science, generous spirit, inherent fairness and integrity and gentle leadership helped create a wonderfully productive environment in which to work. - JWM.

This page intentionally left blank

Preface

The resurgence of interest in artificial neural networks fortunately coincided with the emergence of new technology in molecular biology and the explosion of information about the genomes of humans and other species. Many important problems in genome informatics have been successfully addressed with artificial neural networks, and a vast literature has developed within the last two decades. The purpose of this book is to introduce molecular biologists and other informatics scientists to artificial neural network technology and terminology; to review the major neural network applications in genome informatics; to address the important issues in applying neural network technology to informatics; and to identify significant remaining problems.

Part I of this book gives an overview of applications of artificial neural network technology. Part II contains a tutorial introduction to the most commonly used neural network architectures, network training methods, and applications and limitations of the different architectures. Part III reviews the current state of the art of neural network applications to genome informatics and discusses crucial issues such as input variable selection and preprocessing. Finally, Part IV identifies some of the remaining issues and future directions for research, including integration of statistical rigor into neural network applications, hybrid systems and knowledge extraction.

Acknowledgements

This work would not have been possible without the support of the National Library of Medicine, the National Biomedical Research Foundation in Georgetown, Washington, D.C., and the University of Texas Health Center at Tyler. The authors are grateful for their helpful discussions with Dr. Hongzhan Huang at the National Biomedical Research Foundation and for the expertise and efforts of Dr. Karen Sloan and Sara Shepherd.

Cathy H. Wu
Jerry McLarty
Georgetown and Tyler, May1999

This page intentionally left blank

Contents