

Advanced Information and Knowledge Processing

Also in this series

Gregoris Mentzas, Dimitris Apostolou, Andreas Abecker and Ron Young
Knowledge Asset Management
1-85233-583-1

Michalis Vazirgiannis, Maria Halkidi and Dimitrios Gunopulos
Uncertainty Handling and Quality Assessment in Data Mining
1-85233-655-2

Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho
Ontological Engineering
1-85233-551-3

Arno Scharl (Ed.)
Environmental Online Communication
1-85233-783-4

Shichao Zhang, Chengqi Zhang and Xindong Wu
Knowledge Discovery in Multiple Databases
1-85233-703-6

Jason T.L. Wang, Mohammed J. Zaki, Hannu T.T. Toivonen and Dennis Shasha (Eds)
Data Mining in Bioinformatics
1-85233-671-4

C.C. Ko, Ben M. Chen and Jianping Chen
Creating Web-based Laboratories
1-85233-837-7

Manuel Grana, Richard Duro, Alicia d'Anjou and Paul P. Wang (Eds)
Information Processing with Evolutionary Algorithms
1-85233-886-0

Nikhil R. Pal and Lakhmi Jain (Eds)
Advanced Techniques in Knowledge Discovery and Data Mining
1-85233-867-9

Colin Fyfe

Hebbian Learning and Negative Feedback Networks

With 117 Figures



Springer

Colin Fyfe
Applied Computational Intelligence Research Unit, The University of Paisley, UK

Series Editors

Xindong Wu
Lakhmi Jain

British Library Cataloguing in Publication Data

Fyfe, Colin

Hebbian learning and negative feedback networks.–

(Advanced information and knowledge processing)

1. Neural Networks (Computer science) 2. Data mining

I. Title

006.3'2

ISBN 1852338830

Library of Congress Cataloging-in-Publication Data

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

AI&KP ISSN 1610-3947

ISBN 1-85233-883-0

Springer is a part of Springer Science+Business Media
springeronline.com

© Springer-Verlag London Limited 2005

Printed in the United States of America

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Typesetting: Electronic text files prepared by authors

34/3830-543210 Printed on acid-free paper SPIN 11006701

To my wife, Mary Teresa, for her unswerving help and support.

Contents

1	Introduction	1
1.1	Artificial Neural Networks	2
1.2	The Organisation of this Book	3

Part I Single Stream Networks

2	Background	11
2.1	Hebbian Learning	11
2.2	Quantification of Information	13
2.2.1	Entropy and the Gaussian Distribution	14
2.3	Principal Component Analysis	16
2.4	Weight Decay in Hebbian Learning	18
2.4.1	Principal Components and Weight Decay	19
2.5	ANNs and PCA	21
2.5.1	Oja's One Neuron Model	21
2.5.2	Oja's Subspace Algorithm	22
2.5.3	Oja's Weighted Subspace Algorithm	22
2.5.4	Sanger's Generalized Hebbian Algorithm	23
2.6	Anti-Hebbian Learning	23
2.7	Independent Component Analysis	25
2.7.1	A Restatement of the Problem	26
2.7.2	A Neural Model for ICA	28
2.8	Conclusion	29
3	The Negative Feedback Network	31
3.1	Introduction	31
3.1.1	Equivalence to Oja's Subspace Algorithm	31
3.1.2	Algorithm for PCA	34
3.1.3	Plasticity and Continuity	35
3.1.4	Speed of Learning and Information Content	36

3.1.5	Analysis of Convergence	37
3.2	The VW Model	41
3.2.1	Properties of the VW Network	42
3.2.2	Theoretical Discussion	43
3.3	Using Distance Differences	46
3.3.1	Equivalence to Sanger's Algorithm	47
3.4	Minor Components Analysis	48
3.4.1	Regression	49
3.4.2	Use of Minor Components Analysis	50
3.4.3	Robustness of Regression Solutions	52
3.4.4	Application to ICA	53
3.5	Conclusion	54
4	Peer-Inhibitory Neurons	57
4.1	Analysis of Differential Learning Rates	59
4.1.1	The GW Anomaly	68
4.1.2	Simulations	69
4.2	Differential Activation Functions	70
4.2.1	Model 1: Lateral Activation Functions	70
4.2.2	Model 2: Lateral and Feedforward Activation Functions	73
4.2.3	Model 3: Feedforward Activation Functions	75
4.2.4	Summary	77
4.2.5	Other Models	80
4.3	Emergent Properties of the Peer-Inhibition Network	82
4.4	Conclusion	83
5	Multiple Cause Data	85
5.0.1	A Typical Experiment	87
5.1	Non-negative Weights	88
5.1.1	Other Data Sets	89
5.1.2	Theoretical Analysis	91
5.1.3	Conclusion	92
5.2	Factor Analysis	92
5.2.1	Principal Factor Analysis	93
5.2.2	The Varimax Rotation	94
5.2.3	Relation to Non-negativity	94
5.2.4	The Bars Data	95
5.2.5	Continuous Data	96
5.2.6	Generalised PCA	97
5.2.7	Non-negative Outputs	99
5.2.8	Additive Noise	100
5.2.9	Dimensionality of the Output Space	101
5.2.10	The Minimum Overcomplete Basis	102
5.2.11	Simulations	104
5.3	Conclusion	108

6	Exploratory Data Analysis	111
6.1	Exploratory Projection Pursuit	112
6.1.1	Interesting Directions	112
6.2	The Data and Sphering	113
6.3	The Projection Pursuit Network	114
6.3.1	Extending PCA	114
6.3.2	The Projection Pursuit Indices	116
6.3.3	Principal Component Analysis	117
6.3.4	Convergence of the Algorithm	117
6.3.5	Experimental Results	119
6.3.6	Using Hyperbolic Functions	121
6.3.7	Simulations	122
6.4	Other Indices	123
6.4.1	Indices Based on Information Theory	123
6.4.2	Friedman's Index	125
6.4.3	Intrator's Index	127
6.5	Using Exploratory Projection Pursuit	127
6.5.1	Hierarchical Exploratory Projection Pursuit	128
6.5.2	World Bank Data	131
6.6	Independent Component Analysis	133
6.7	Conclusion	136
7	Topology Preserving Maps	137
7.1	Background	137
7.1.1	Competitive Learning	138
7.1.2	The Kohonen Feature Map	138
7.2	The Classification Network	140
7.2.1	Results	141
7.2.2	Stochastic Neurons	143
7.3	The Scale Invariant Map	143
7.3.1	An Example	144
7.3.2	Comparison with Kohonen Feature Maps	145
7.3.3	Discussion	145
7.3.4	Self-Organisation on Voice Data	148
7.3.5	Vowel Classification	148
7.4	The Subspace Map	149
7.4.1	Summary of the Training Algorithm	151
7.4.2	Training and Results	153
7.4.3	Discussion	157
7.5	The Negative Feedback Coding Network	158
7.5.1	Results	159
7.5.2	Statistics and Weights	160
7.5.3	Reconstruction Error	162
7.5.4	Topology Preservation	163
7.5.5	Approximate Topological Equivalence	164

7.5.6	A Hierarchical Feature Map	165
7.5.7	A Biological Implementation	166
7.6	Conclusion	167
8	Maximum Likelihood Hebbian Learning	169
8.1	The Negative Feedback Network and Cost Functions	169
8.1.1	Is This a Hebbian Rule?	171
8.2	ϵ -Insensitive Hebbian Learning	171
8.2.1	Principal Component Analysis	171
8.2.2	Anti-Hebbian Learning	174
8.2.3	Other Negative Feedback Networks	175
8.3	The Maximum Likelihood EPP Algorithm	176
8.3.1	Minimum Likelihood Hebbian Learning	177
8.3.2	Experimental Results	178
8.3.3	Skewness	180
8.4	A Combined Algorithm	181
8.4.1	Astronomical Data	182
8.4.2	Wine	182
8.4.3	Independent Component Analysis	183
8.5	Conclusion	186

Part II Dual Stream Networks

9	Two Neural Networks for Canonical Correlation Analysis ..	191
9.1	Statistical Canonical Correlation Analysis	191
9.2	The First Canonical Correlation Network	192
9.3	Experimental Results	194
9.3.1	Artificial Data	195
9.3.2	Real Data	195
9.3.3	Random Dot Stereograms	196
9.3.4	Equal Correlations	198
9.3.5	More Than Two Data Sets	199
9.3.6	Many Correlations	200
9.4	A Second Neural Implementation of CCA	202
9.5	Simulations	204
9.5.1	Artificial Data	204
9.5.2	Real Data	205
9.6	Linear Discriminant Analysis	206
9.7	Discussion	207
10	Alternative Derivations of CCA Networks	209
10.1	A Probabilistic Perspective	209
10.1.1	Putting Priors on the Probabilities	210
10.2	Robust CCA	211

10.3	A Model Derived from Becker's Model 1	212
10.3.1	Who Is Telling the Truth?	213
10.3.2	A Model Derived from Becker's Second Model	214
10.4	Discussion	215
11	Kernel and Nonlinear Correlations	217
11.1	Nonlinear Correlations	217
11.1.1	Experiment Results	217
11.2	The Search for Independence	221
11.2.1	Using Minimum Correlation to Extract Independent Sources	222
11.2.2	Experiments	223
11.2.3	Forecasting	223
11.3	Kernel Canonical Correlation Analysis	225
11.3.1	Kernel Principal Correlation Analysis	226
11.3.2	Kernel Canonical Correlation Analysis	227
11.3.3	Simulations	230
11.3.4	ICA using KCCA	231
11.4	Relevance Vector Regression	234
11.4.1	Application to CCA	237
11.5	Appearance-Based Object Recognition	237
11.6	Mixtures of Linear Correlations	240
11.6.1	Many Locally Linear Correlations	240
11.6.2	Stone's Data	241
11.6.3	Discussion	246
12	Exploratory Correlation Analysis	247
12.1	Exploratory Correlation Analysis	247
12.2	Experiments	251
12.2.1	Artificial Data	251
12.2.2	Dual Stream Blind Source Separation	252
12.3	Connection to CCA	253
12.4	FastECA	254
12.4.1	FastECA for Several Units	255
12.4.2	Comparison of ECA and FastECA	256
12.5	Local Filter Formation From Natural Stereo Images	256
12.5.1	Biological Vision	256
12.5.2	Sparse Coding of Natural Images	259
12.5.3	Stereo Experiments	260
12.6	Twinned Maximum Likelihood Learning	266
12.7	Unmixing of Sound Signals	270
12.8	Conclusion	271

13 Multicollinearity and Partial Least Squares	275
13.1 The Ridge Model	276
13.2 Application to CCA	276
13.2.1 Relation to Partial Least Squares	279
13.3 Extracting Multiple Canonical Correlations	280
13.4 Experiments on Multicollinear Data	281
13.4.1 Artificial Data	281
13.4.2 Examination Data	281
13.4.3 Children's Gait Data	281
13.5 A Neural Implementation of Partial Least Squares	284
13.5.1 Introducing Nonlinear Correlations	284
13.5.2 Simulations	285
13.5.3 Linear Neural PLS	285
13.5.4 Mixtures of Linear Neural PLS	286
13.5.5 Nonlinear Neural PLS Regression	287
13.6 Conclusion	288
14 Twinned Principal Curves	291
14.1 Twinned Principal Curves	291
14.2 Properties of Twinned Principal Curves	293
14.2.1 Comparison with Single Principal Curves	293
14.2.2 Illustrative Examples	295
14.2.3 Intersecting Curves	297
14.2.4 Termination Criteria: MSE	299
14.2.5 Termination Criteria: Using Derivative Information	299
14.2.6 Alternative Twinned Principal Curves	301
14.3 Twinned Self-Organising Maps	305
14.3.1 Predicting Student's Exam Marks	306
14.4 Discussion	307
15 The Future	309
15.1 Review	309
15.2 Omissions	312
15.3 Current and Future Work	312
Negative Feedback Artificial Neural Networks	315
A.1 The Interneuron Model	315
A.2 Other Models	317
A.2.1 Static Models	318
A.2.2 Dynamic Models	319
A.3 Related Biological Models	320

Previous Factor Analysis Models	323
B.1 Földiák's Sixth Model	323
B.1.1 Implementation Details	324
B.1.2 Results	325
B.2 Competitive Hebbian Learning	326
B.3 Multiple Cause Models	327
B.3.1 Saund's Model	328
B.3.2 Dayan and Zemel	330
B.4 Predictability Minimisation	330
B.5 Mixtures of Experts	332
B.5.1 An Example	334
B.6 Probabilistic Models	334
B.6.1 Mixtures of Gaussians	335
B.6.2 A Logistic Belief Network	337
B.6.3 The Helmholtz Machine and the EM Algorithm	337
B.6.4 The Wake-Sleep Algorithm	338
B.6.5 Olshausen and Field's Sparse Coding Network	339
Related Models for ICA	341
C.1 Jutten and Herault	341
C.1.1 An Example Separation	342
C.1.2 Learning the Weights	343
C.2 Nonlinear PCA	344
C.2.1 Simulations and Discussion	345
C.3 Information Maximisation	345
C.3.1 The Learning Algorithm	347
C.4 Penalised Minimum Reconstruction Error	350
C.4.1 Adding Competition	350
C.5 FastICA	351
C.5.1 FastICA for One Unit	352
Previous Dual Stream Approaches	353
D.1 The I-Max Model	353
D.2 Stone's Model	355
D.3 Kay's Neural Models	356
D.4 Borga's Algorithm	358
Data Sets	363
E.1 Artificial Data Sets	363
E.1.1 Gaussian Data	363
E.1.2 Bars Data	363
E.1.3 Sinusoids	364
E.1.4 Random Dot Stereograms	365
E.1.5 Nonlinear Manifolds	365
E.2 Real Data Sets	366

E.2.1	Wine Data	366
E.2.2	Astronomical Data	366
E.2.3	The Cetin Data Set	366
E.2.4	Exam Data	366
E.2.5	Children’s Gait Data	367
E.2.6	Speech Signals	367
E.2.7	Bank Database.....	368
E.2.8	World Bank Data	368
E.2.9	Exchange Rate Data	368
E.2.10	Power Load Data.....	368
E.2.11	Image Data	368
References		371
Index		381

Acronyms

ANN Artificial Neural Network
ASSOM Adaptive Subspace Self-Organising Map
CCA Canonical Correlation Analysis
ECA Exploratory Correlation Analysis
EPP Exploratory Projection Pursuit
FA Factor Analysis
HEPP Hierarchical Exploratory Projection Pursuit
ICA Independent Component Analysis
KCCA Kernel Canonical Correlation Analysis
KPCA Kernel Principal Component Analysis
LDA Linear Discriminant Analysis
MCA Minor Components Analysis
ML Maximum Likelihood
MOB Minimum Overcomplete Basis
MSE Mean Square Error
NLCCA Nonlinear Canonical Correlation Analysis
NLPCA Nonlinear Principal Components Analysis
NLPLS Nonlinear Partial Least Squares
PCA Principal Component Analysis
PFA Principal Factor Analysis
PLS Partial Least Squares
SIM Scale Invariant Map
SOM Self-Organising Map

Preface

This book is the outcome of a decade's research into a specific architecture and associated learning mechanism for an artificial neural network: the architecture involves negative feedback and the learning mechanism is simple Hebbian learning. The research began with my own thesis at the University of Strathclyde, Scotland, under Professor Douglas McGregor which culminated with me being awarded a PhD in 1995 [52], the title of which was "Negative Feedback as an Organising Principle for Artificial Neural Networks".

Naturally enough, having established this theme, when I began to supervise PhD students of my own, we continued to develop this concept and this book owes much to the research and theses of these students at the Applied Computational Intelligence Research Unit in the University of Paisley. Thus we discuss work from

- Dr. Darryl Charles [24] in Chapter 5.
- Dr. Stephen McGlinchey [127] in Chapter 7.
- Dr. Donald MacDonald [121] in Chapters 6 and 8.
- Dr. Emilio Corchado [29] in Chapter 8.

We briefly discuss one simulation from the thesis of Dr. Mark Girolami [58] in Chapter 6 but do not discuss any of the rest of his thesis since it has already appeared in book form [59]. We also must credit Cesar Garcia Osorio, a current PhD student, for the comparative study of the two Exploratory Projection Pursuit networks in Chapter 8. All of Chapters 3 to 8 deal with single stream artificial neural networks.

Chapters 9-14 discuss research into dual stream artificial neural networks at the Applied Computational Intelligence Research Unit in the University of Paisley. This work has resulted in four PhDs [60, 67, 108, 113]. I must therefore acknowledge the work done by:

- Dr. Pei Ling Lai [113] in Chapters 9, 10 and 11.
- Dr. Jos Koetsier [108] in Chapters 12 and 14.
- Dr. Zhenkun Gou [60] in Chapters 9, 11 and 13.

XVIII Preface

- Dr. Ying Han [67] in Chapters 12, 13 and 14.

Without their assistance this book could not have been written. We must also credit current PhD student, Emilio Corchado (yes, his second PhD) with some of the results in Chapter 14. We must also acknowledge other research students whose work does not form part of this book, but whose overall contribution to the life of the group was invaluable: these include Dr. Shang-Jen Chuang, Dr. Tzai-Der Wang, Dr. Juan Corchado, Dr. Danny Livingstone and Dr. Lina Petrakieva.

This book also has sections which were used for undergraduate teaching at the University of Paisley and I must credit these undergraduates with inspiring me to write more clearly.

Needless to say, we cannot cover all the work in these theses in a single book and so the interested reader is invited to consult the originals for more detailed description of any work which commands his or her interest.

Colin Fyfe
Paisley, 2004.