

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels

William W. Hsieh

Frontmatter

[More information](#)

MACHINE LEARNING METHODS IN THE ENVIRONMENTAL SCIENCES

Neural Networks and Kernels

William W. Hsieh

Machine learning methods, having originated from computational intelligence (i.e. artificial intelligence), are now ubiquitous in the environmental sciences. This is the first single-authored textbook to give a unified treatment of machine learning methods and their applications in the environmental sciences.

Machine learning methods began to infiltrate the environmental sciences in the 1990s. Today, thanks to their powerful nonlinear modelling capability, they are no longer an exotic fringe species, as they are heavily used in satellite data processing, in general circulation models (GCM), in weather and climate prediction, air quality forecasting, analysis and modelling of environmental data, oceanographic and hydrological forecasting, ecological modelling, and in the monitoring of snow, ice and forests, etc. End-of-chapter review questions are included, allowing readers to develop their problem-solving skills and monitor their understanding of the material presented. An appendix lists websites available for downloading computer code and data sources. A resources website is available containing datasets for exercises, and additional material to keep the book completely up-to-date.

This book presents machine learning methods and their applications in the environmental sciences (including satellite remote sensing, atmospheric science, climate science, oceanography, hydrology and ecology), written at a level suitable for beginning graduate students and advanced undergraduates. It is also valuable for researchers and practitioners in environmental sciences interested in applying these new methods to their own work.

WILLIAM W. HSIEH is a Professor in the Department of Earth and Ocean Sciences and in the Department of Physics and Astronomy, as well as Chair of the Atmospheric Science Programme, at the University of British Columbia. He is internationally known for his pioneering work in developing and applying machine learning methods in the environmental sciences. He has published over 80 peer-reviewed journal publications covering areas of climate variability, machine learning, oceanography, atmospheric science and hydrology.

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)

MACHINE LEARNING METHODS IN THE ENVIRONMENTAL SCIENCES

Neural Networks and Kernels

WILLIAM W. HSIEH

*University of British Columbia
Vancouver, BC, Canada*



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521791922

© W. W. Hsieh 2009

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-79192-2 hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party Internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page</i> ix
<i>List of abbreviations</i>	xii
1 Basic notions in classical data analysis	1
1.1 Expectation and mean	1
1.2 Variance and covariance	2
1.3 Correlation	3
1.4 Regression	7
1.5 Bayes theorem	12
1.6 Discriminant functions and classification	14
1.7 Clustering	16
Exercises	18
2 Linear multivariate statistical analysis	20
2.1 Principal component analysis (PCA)	20
2.2 Rotated PCA	40
2.3 PCA for vectors	48
2.4 Canonical correlation analysis (CCA)	49
Exercises	57
3 Basic time series analysis	58
3.1 Spectrum	58
3.2 Windows	65
3.3 Filters	66
3.4 Singular spectrum analysis	68
3.5 Multichannel singular spectrum analysis	74
3.6 Principal oscillation patterns	75
3.7 Spectral principal component analysis	82
Exercises	85
4 Feed-forward neural network models	86
4.1 McCulloch and Pitts model	87

vi	<i>Contents</i>	
	4.2 Perceptrons	87
	4.3 Multi-layer perceptrons (MLP)	92
	4.4 Back-propagation	97
	4.5 Hidden neurons	102
	4.6 Radial basis functions (RBF)	105
	4.7 Conditional probability distributions	108
	Exercises	112
5	Nonlinear optimization	113
	5.1 Gradient descent method	115
	5.2 Conjugate gradient method	116
	5.3 Quasi-Newton methods	120
	5.4 Nonlinear least squares methods	121
	5.5 Evolutionary computation and genetic algorithms	124
	Exercises	126
6	Learning and generalization	127
	6.1 Mean squared error and maximum likelihood	127
	6.2 Objective functions and robustness	129
	6.3 Variance and bias errors	133
	6.4 Reserving data for validation	134
	6.5 Regularization	135
	6.6 Cross-validation	136
	6.7 Bayesian neural networks (BNN)	138
	6.8 Ensemble of models	145
	6.9 Approaches to predictive uncertainty	150
	6.10 Linearization from time-averaging	151
	Exercises	155
7	Kernel methods	157
	7.1 From neural networks to kernel methods	157
	7.2 Primal and dual solutions for linear regression	159
	7.3 Kernels	161
	7.4 Kernel ridge regression	164
	7.5 Advantages and disadvantages	165
	7.6 The pre-image problem	167
	Exercises	169
8	Nonlinear classification	170
	8.1 Multi-layer perceptron classifier	171
	8.2 Multi-class classification	175
	8.3 Bayesian neural network (BNN) classifier	176
	8.4 Support vector machine (SVM) classifier	177
	8.5 Forecast verification	187

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)

<i>Contents</i>		vii
8.6	Unsupervised competitive learning	193
	Exercises	195
9	Nonlinear regression	196
9.1	Support vector regression (SVR)	196
9.2	Classification and regression trees (CART)	202
9.3	Gaussian processes (GP)	206
9.4	Probabilistic forecast scores	211
	Exercises	212
10	Nonlinear principal component analysis	213
10.1	Auto-associative NN for nonlinear PCA	214
10.2	Principal curves	231
10.3	Self-organizing maps (SOM)	233
10.4	Kernel principal component analysis	237
10.5	Nonlinear complex PCA	240
10.6	Nonlinear singular spectrum analysis	244
	Exercises	251
11	Nonlinear canonical correlation analysis	252
11.1	MLP-based NLCCA model	252
11.2	Robust NLCCA	264
	Exercises	273
12	Applications in environmental sciences	274
12.1	Remote sensing	275
12.2	Oceanography	286
12.3	Atmospheric science	292
12.4	Hydrology	312
12.5	Ecology	314
	Exercises	317
<i>Appendices</i>		
	<i>A Sources for data and codes</i>	318
	<i>B Lagrange multipliers</i>	319
<i>References</i>		322
<i>Index</i>		345

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)

Preface

Machine learning is a major sub-field in computational intelligence (also called artificial intelligence). Its main objective is to use computational methods to extract information from data. Machine learning has a wide spectrum of applications including handwriting and speech recognition, object recognition in computer vision, robotics and computer games, natural language processing, brain–machine interfaces, medical diagnosis, DNA classification, search engines, spam and fraud detection, and stock market analysis. Neural network methods, generally regarded as forming the first wave of breakthrough in machine learning, became popular in the late 1980s, while kernel methods arrived in a second wave in the second half of the 1990s.

In the 1990s, machine learning methods began to infiltrate the environmental sciences. Today, they are no longer an exotic fringe species, since their presence is ubiquitous in the environmental sciences, as illustrated by the lengthy References section of this book. They are heavily used in satellite data processing, in general circulation models (GCM) for emulating physics, in post-processing of GCM model outputs, in weather and climate prediction, air quality forecasting, analysis and modelling of environmental data, oceanographic and hydrological forecasting, ecological modelling, and in monitoring of snow, ice and forests, etc.

This book presents machine learning methods (mainly neural network and kernel methods) and their applications in the environmental sciences, written at a level suitable for beginning graduate students and advanced undergraduates. It is also aimed at researchers and practitioners in environmental sciences, who having been intrigued by exotic terms like neural networks, support vector machines, self-organizing maps, evolutionary computation, etc., are motivated to learn more about these new methods and to use them in their own work. The reader is assumed to know multivariate calculus, linear algebra and basic probability.

Chapters 1–3, intended mainly as background material for students, cover the standard statistical methods used in environmental sciences. The machine learning methods of later chapters provide powerful nonlinear generalizations for many of these standard linear statistical methods. The reader already familiar with the background material of Chapters 1–3 can start directly with Chapter 4, which introduces neural network methods. While Chapter 5 is a relatively technical chapter on nonlinear optimization algorithms, Chapter 6 on learning and generalization is essential to the proper use of machine learning methods – in particular, Section 6.10 explains why a nonlinear machine learning method often outperforms a linear method in weather applications but fails to do so in climate applications. Kernel methods are introduced in Chapter 7. Chapter 8 covers nonlinear classification, Chapter 9, nonlinear regression, Chapter 10, nonlinear principal component analysis, and Chapter 11, nonlinear canonical correlation analysis. Chapter 12 broadly surveys applications of machine learning methods in the environmental sciences (remote sensing, atmospheric science, oceanography, hydrology, ecology, etc.). For exercises, the student could test the methods on data from their own area or from some of the websites listed in Appendix A. Codes for many machine learning methods are also available from sites listed in Appendix A. The book website www.cambridge.org/hsieh also provides datasets for some of the exercises given at the ends of the chapters.

On a personal note, writing this book has been both exhilarating and grueling. When I first became intrigued by neural networks through discussions with Dr Benyang Tang in 1992, I recognized that the new machine learning methods would have a major impact on the environmental sciences. However, I also realized that I had a steep learning curve ahead of me, as my background training was in physics, mathematics and environmental sciences, but not in statistics nor computer science. By the late 1990s I became convinced that the best way for me to learn more about machine learning was to write a book. What I thought would take a couple of years turned into a marathon of over eight years, as I desperately tried to keep pace with a rapidly expanding research field. I managed to limp past the finish line in pain, as repetitive strain injury from overusage of keyboard and mouse struck in the final months of intensive writing!

I have been fortunate in having supervised numerous talented graduate students, post-doctoral fellows and research associates, many of whom taught me far more than I taught them. I received helpful editorial assistance from the staff at the Cambridge University Press and from Max Ng. I am grateful for the support from my two university departments (Earth and Ocean Sciences, and Physics and Astronomy), the Peter Wall Institute of Advanced Studies, the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Climate and Atmospheric Sciences.

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)*Preface*

xi

Without the loving support from my family (my wife Jean and my daughters, Teresa and Serena), and the strong educational roots planted decades ago by my parents and my teachers, I could not have written this book.

Notation used in this book

In general, vectors are denoted by lower case bold letters (e.g. \mathbf{v}), matrices by upper case bold letters (e.g. \mathbf{A}) and scalar variables by italics (e.g. x or J). A column vector is denoted by \mathbf{v} , while its transpose \mathbf{v}^T is a row vector, i.e. $\mathbf{v}^T = (v_1, v_2, \dots, v_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$, and the inner or dot product of two vectors $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$. The elements of a matrix \mathbf{A} are written as A_{ij} or $(\mathbf{A})_{ij}$. The probability for discrete variables is denoted by upper case P , whereas the probability density for continuous variables is denoted by lower case p . The expectation is denoted by $E[\dots]$ or $\langle \dots \rangle$. The natural logarithm is denoted by \ln or \log .

Abbreviations

AO = Arctic Oscillation
BNN = Bayesian neural network
CART = classification and regression tree
CCA = canonical correlation analysis
CDN = conditional density network
EC = evolutionary computation
EEOF = extended empirical orthogonal function
ENSO = El Niño-Southern Oscillation
EOF = empirical orthogonal function
GA = genetic algorithm
GCM = general circulation model (or global climate model)
GP = Gaussian process model
IC = information criterion
LP = linear projection
MAE = mean absolute error
MCA = maximum covariance analysis
MJO = Madden-Julian Oscillation
MLP = multi-layer perceptron neural network
MLR = multiple linear regression
MOS = model output statistics
MSE = mean squared error
MSSA = multichannel singular spectrum analysis
NAO = North Atlantic Oscillation
NLCCA = nonlinear canonical correlation analysis
NLCPA = nonlinear complex PCA
NN = neural network
NLPC = nonlinear principal component
NLPCA = nonlinear principal component analysis

Cambridge University Press

978-0-521-79192-2 - Machine Learning Methods in the Environmental Sciences: Neural Networks and
Kernels

William W. Hsieh

Frontmatter

[More information](#)

Abbreviations

xiii

NLSSA = nonlinear singular spectrum analysis

PC = principal component

PCA = principal component analysis

PNA = Pacific-North American teleconnection

POP = principal oscillation pattern

QBO = Quasi-Biennial Oscillation

RBF = radial basis function

RMSE = root mean squared error

SLP = sea level pressure

SOM = self-organizing map

SSA = singular spectrum analysis

SST = sea surface temperature (sum of squares in Chapter 1)

SVD = singular value decomposition

SVM = support vector machine

SVR = support vector regression