Neural Networks and Statistical Learning

Ke-Lin Du · M. N. S. Swamy

Neural Networks and Statistical Learning

Second Edition

Springer

Ke-Lin Du
Department of Electrical
and Computer Engineering
Concordia University
Montreal, QC, Canada

Xonlink Inc.
Hangzhou, China

M. N. S. Swamy
Department of Electrical
and Computer Engineering
Concordia University
Montreal, QC, Canada

ISBN 978-1-4471-7451-6 ISBN 978-1-4471-7452-3 (eBook)
https://doi.org/10.1007/978-1-4471-7452-3

This Springer imprint is published by the registered company Springer-Verlag London Ltd. part of Springer Nature.
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

To Falong Xing and Jie Zeng
—Ke-Lin Du

To my teachers and my students
—M. N. S. Swamy

Preface to the Second Edition

Since the publication of the first edition in December 2013, the rapid rise of deep learning and AI has resulted in a wave of research activities and numerous new results. During the past few years, there have been several breakthroughs in deep learning and AI. At the same time, research and application of big data are widespread. Machine learning has become the brains behind big data.

In such a setting, this book has become one of the best sellers of Springer books. Under the suggestion of Anthony Doyle at Springer London Ltd., we decided to publish this second edition.

In this second edition, we will add six new chapters to the first edition:

- Chapter 3 focuses on computation learning theory. Part of its content is split from Chap. 2 of the first edition.
- Chapter 18 introduces compressed sensing and sparse coding. In this approach, a datum is represented as a linear combination of basis functions, and the coefficients are assumed to be sparse.
- Chapter 19 deals with matrix completion. Recovery of a data matrix from a subset of its entries is an extension of compressed sensing and sparse approximation.
- Chapter 23 introduces the Boltzmann machine. Part of its content is split from Chap. 19 of the first edition.
- Chapter 24 describes deep learning and deep neural networks. Deep learning is the state-of-the-art approach to solving complex problems.
- Chapter 31 introduces big data, cloud computing, and Internet of Things. These topics go hand in hand. Machine learning functions as the major tool for data analytics.

We also update each chapter by including major contributions published in the past 6 years.

The authors wish to thank Jie Zeng (Richcon MC, Inc., China), Biaobiao Zhang (Yuantu Internet Technology Ltd., China), Li Yu (Zhejiang University of Technology, China), Zhijiang Xu (Zhejiang University of Technology, China), and Renwang Li (Zhejiang Sci-Tech University, China) for their help during the preparation of this second edition.

Hangzhou, China/Montreal, Canada — Ke-Lin Du
Montreal, Canada — M. N. S. Swamy
May 2019

Preface to the First Edition

The human brain, consisting of nearly 10^{11} neurons, is the center of human intelligence. Human intelligence has been simulated in various ways. Artificial intelligence (AI) pursues exact logical reasoning based on symbol manipulation. Fuzzy logics model the highly uncertain behavior of decision-making. Neural networks model the highly nonlinear infrastructure of brain networks. Evolutionary computation models the evolution of intelligence. Chaos theory models the highly nonlinear and chaotic behaviors of human intelligence.

Soft computing is an evolving collection of methodologies for the representation of ambiguity in human thinking; it exploits the tolerance for imprecision and uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions. The major methodologies of soft computing are fuzzy logic, neural networks, and evolutionary computation.

Conventional model-based data processing methods require experts' knowledge for the modeling of a system. Neural network methods provide a model-free, adaptive, fault-tolerant, parallel, and distributed processing solution. A neural network is a black box that directly learns the internal relations of an unknown system, without guessing functions for describing cause-and-effect relationships. The neural network approach is a basic methodology of information processing. Neural network models may be used for function approximation, classification, nonlinear mapping, associative memory, vector quantization, optimization, feature extraction, clustering, and approximate inference. Neural networks have wide applications in almost all areas of science and engineering.

Fuzzy logic provides a means for treating uncertainty and computing with words. This mimics human recognition, which skillfully copes with uncertainty. Fuzzy systems are conventionally created from explicit knowledge expressed in the form of fuzzy rules, which are designed based on experts' experience. A fuzzy system can explain its action by fuzzy rules. Neurofuzzy systems, as a synergy of fuzzy logic and neural networks, possess both learning and knowledge representation capabilities.

This book is our attempt to bring together the major advances in neural networks and machine learning, and to explain them in a statistical framework. While some mathematical details are needed, we emphasize the practical aspects of the models and methods rather than the theoretical details. To us, neural networks are merely some statistical methods that can be represented by graphs and networks. They can iteratively adjust the network parameters. As a statistical model, a neural network can learn the probability density function from the given samples, and then predict, by generalization according to the learnt statistics, outputs for new samples that are not included in the learning sample set.

The neural network approach is a general statistical computational paradigm. Neural network research solves two problems: the direct problem and the inverse problem. The direct problem employs computer and engineering techniques to model biological neural systems of the human brain. This problem is investigated by cognitive scientists and can be useful in neuropsychiatry and neurophysiology. The inverse problem simulates biological neural systems for their problem-solving capabilities for application in scientific or engineering fields. Engineering and computer scientists have conducted an extensive investigation in this area. This book concentrates mainly on the inverse problem, although the two areas often shed light on each other. The biological and psychological plausibility of the neural network models have not been seriously treated in this book, though some background material is discussed.

This book is intended to be used as a textbook for advanced undergraduate and graduate students in engineering, science, computer science, business, arts, and medicine. It is also a good reference book for scientists, researchers, and practitioners in a wide variety of fields, and assumes no previous knowledge of neural network or machine learning concepts.

This book is divided into 25 chapters and 2 appendices. It contains almost all the major neural network models and statistical learning approaches. We also give an introduction to fuzzy sets and logic, and neurofuzzy models. Hardware implementations of the models are discussed. Two chapters are dedicated to the applications of neural network and statistical learning approaches to biometrics/bioinformatics and data mining. Finally, in the appendices, some mathematical preliminaries are given, and benchmarks for validating all kinds of neural network methods and some web resources are provided.

First and foremost, we would like to thank the supporting staff from Springer London, especially Anthony Doyle and Grace Quinn for their enthusiastic and professional support throughout the period of manuscript preparation.

K.-L. Du also wishes to thank Jiabin Lu (Guangdong University of Technology, China), Jie Zeng (Richcon MC, Inc., China), Biaobiao Zhang and Hui Wang (Enjoyor, Inc., China), and many of his graduate students, including Na Shou, Shengfeng Yu, Lusha Han, Xiaolan Shen, Yuanyuan Chen, and Xiaoling Wang (Zhejiang University of Technology, China) for their consistent assistance.

In addition, we should mention at least the following names for their help: Omer Morgul (Bilkent University, Turkey), Yanwu Zhang (Monterey Bay Aquarium Research Institute, USA), Chi Sing Leung (City University of Hong Kong, Hong Kong), M. Omair Ahmad and Jianfeng Gu (Concordia University, Canada), Li Yu, Limin Meng, Jingyu Hua, Zhijiang Xu, and Luping Fang (Zhejiang University of Technology, China), Yuxing Dai (Wenzhou University, China), and Renwang Li (Zhejiang Sci-Tech University, China). Last, but not least, we would like to thank our families for their support and understanding during the course of writing this book.

A book of this length is certain to have some errors and omissions. Feedback is welcome via email at kldu@ieee.org or swamy@encs.concordia.ca. Due to restriction on the length of this book, we have placed two appendices, namely, Mathematical preliminaries, and Benchmarks and resources, on the website of this book. MATLAB code for the worked examples is also downloadable from the website of this book.

Hangzhou, China — Ke-Lin Du
Montreal, Canada — M. N. S. Swamy
April 2013

Contents

Abbreviations

A/D	Analog-to-digital
adaline	Adaptive linear element
AI	Artificial intelligence
AIC	Akaike information criterion
ALA	Adaptive learning algorithm
ANFIS	Adaptive-network-based fuzzy inference system
AOSVR	Accurate online SVR
APCA	Asymmetric PCA
APEX	Adaptive principal components extraction
API	Application programming interface
ART	Adaptive resonance theory
ASIC	Application-specific integrated circuit
ASSOM	Adaptive-subspace SOM
BAM	Bidirectional associative memory
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BIC	Bayesian information criterion
BIRCH	Balanced iterative reducing and clustering using hierarchies
BP	Backpropagation
BPTT	Backpropagation through time
BSB	Brain-states-in-a-box
BSS	Blind source separation
CBIR	Content-based image retrieval
CCA	Canonical correlation analysis
CCCP	Constrained concave-convex procedure
cdf	Cumulative distribution function
CEM	Classification EM
CG	Conjugate gradient
CMAC	Cerebellar model articulation controller
COP	Combinatorial optimization problem
CORDIC	Coordinate rotation digital computer

CoSaMP	Compressive sampling matching pursuit
CPT	Conditional probability table
CPU	Central processing units
CURE	Clustering using representation
DBSCAN	Density-based spatial clustering of applications with noise
DCS	Dynamic cell structures
DCT	Discrete cosine transform
DFP	Davidon–Fletcher–Powell
DFT	Discrete Fourier Transform
ECG	Electrocardiogram
ECOC	Error-correcting output code
EEG	Electroencephalogram
EKF	Extended Kalman filtering
ELM	Extreme learning machine
EM	Expectation–maximization
ERM	Empirical risk minimization
E-step	Expectation step
ETF	Elementary transcendental function
EVD	Eigenvalue decomposition
FCM	Fuzzy *C*-means
FFT	Fast Fourier Transform
FIR	Finite impulse response
fMRI	Functional magnetic resonance imaging
FPGA	Field-programmable gate array
FSCL	Frequency-sensitive competitive learning
GAP-RBF	Growing and pruning algorithm for RBF
GCS	Growing cell structures
GHA	Generalized Hebbian algorithm
GLVQ-F	Generalized LVQ family algorithms
GNG	Growing neural gas
GSO	Gram–Schmidt orthonormal
HWO	Hidden weight optimization
HyFIS	Hybrid neural fuzzy inference system
ICA	Independent component analysis
IHT	Iterative hard thresholding
iid	Independently drawn and identically distributed
IoT	Internet of Things
KKT	Karush–Kuhn–Tucker
LASSO	Least absolute selection and shrinkage operator
LBG	Linde–Buzo–Gray
LDA	Linear discriminant analysis
LM	Levenberg–Marquardt
LMAM	LM with adaptive momentum
LMI	Linear matrix inequality
LMS	Least mean squares

LMSE	Least mean squared error
LMSER	Least mean square error reconstruction
LP	Linear programming
LS	Least squares
LSI	Latent semantic indexing
LTG	Linear threshold gate
LVQ	Learning vector quantization
MAD	Median of the absolute deviation
MAP	Maximum a posteriori
MCA	Minor component analysis
MDL	Minimum description length
MEG	Magnetoencephalogram
MFCC	Mel frequency cepstral coefficient
MIMD	Multiple instruction multiple data
MKL	Multiple kernel learning
ML	Maximum likelihood
MLP	Multilayer perceptron
MSA	Minor subspace analysis
MSE	Mean squared error
MST	Minimum spanning tree
M-step	Maximization step
NARX	Nonlinear autoregressive with exogenous input
NEFCLASS	Neurofuzzy classification
NEFCON	Neurofuzzy controller
NEFLVQ	Non-Euclidean FLVQ
NEFPROX	Neuronfuzzy function approximation
NIC	Novel information criterion
k-NN	*k*-nearest neighbor
NOVEL	Nonlinear optimization via external lead
OBD	Optimal brain damage
OBS	Optimal brain surgeon
OLAP	Online analytical processing
OLS	Orthogonal least squares
OMP	Orthogonal matching pursuit
OWO	Output weight optimization
PAC	Probably approximately correct
PAST	Projection approximation subspace tracking
PASTd	PAST with deflation
PCA	Principal component analysis
PCM	Possibilistic *C*-means
pdf	Probability density function
PSA	Principal subspace analysis
QP	Quadratic programming
QR-cp	QR with column pivoting
RAN	Resource-allocating network

RBF	Radial basis function
RBM	Restricted Boltzmann machine
ReLU	Rectified linear unit
RIC	Restricted isometry constant
RIP	Restricted isometry property
RLS	Recursive least squares
RPCCL	Rival penalized controlled competitive learning
RPCL	Rival penalized competitive learning
Rprop	Resilient propagation
RTRL	Real-time recurrent learning
RVM	Relevance vector machine
SDP	Semidefinite programs
SIMD	Single instruction, multiple data
SLA	Subspace learning algorithm
SMO	Sequential minimal optimization
SOM	Self-organization maps
SPMD	Single program multiple data
SRM	Structural risk minimization
SVD	Singular value decomposition
SVDD	Support vector data description
SVM	Support vector machine
SVR	Support vector regression
TDNN	Time-delay neural network
TDRL	Time-dependent recurrent learning
TLMS	Total least mean squares
TLS	Total least squares
TREAT	Trust-region-based error aggregated training
TRUST	Terminal repeller unconstrained subenergy tunneling
TSK	Takagi–Sugeno–Kang
TSP	Traveling salesman problem
VC	Vapnik–Chervonenkis
VLSI	Very large-scale integrated
WINC	Weighted information criterion
k-WTA	k-winners-take-all
WTA	Winner-takes-all
XML	eXtensible markup language