

KONONENKO
&
KUKAR

MACHINE LEARNING AND DATA MINING

MACHINE LEARNING AND DATA MINING

Introduction to Principles
and Algorithms

“Why do we not, since the phenomena are well known, build a ‘knowledge refinery’ as the basis of a new industry, comparable in some ways to the industry of petroleum refining, only more important in the long run? The product to be refined is codified human knowledge.”

– Donald Michie

This book describes the basics of machine learning principles and algorithms used in data mining. It is suitable for advanced undergraduate and postgraduate students of computer science, researchers who want to adapt algorithms for particular data mining tasks, and advanced users of machine learning and data mining tools.

“The main strength of this book is its breadth of coverage of ML and wealth of material presented, without loss of depth. In this respect, the book is rather unique among the books on machine learning. The book covers a large majority of key methods, concepts and issues in ML. The style of writing is terse and directly to the point.” – Ivan Bratko

“The authors deserve congratulations for very well organized and presented work. I also greatly appreciated the well chosen and stimulating quotations at the head of sections.” – Alan McConnell-Duff

IGOR KONONENKO
and
MATJAŽ KUKAR



HORWOOD

ISBN 978-1-904275-21-3



9 781904 275213

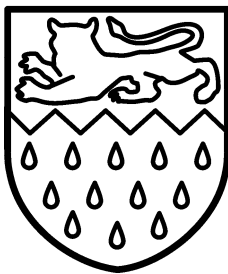


HORWOOD

MACHINE LEARNING AND DATA MINING:

Introduction to Principles and Algorithms

The ultimate goal of machine learning is knowledge
The ultimate goal of human learning is wisdom.



ABOUT THE AUTHORS

Igor Kononenko studied computer science at the University of Ljubljana, Slovenia, receiving his BSc in 1982, MSc in 1985 and PhD in 1990. He is now professor at the Faculty of Computer and Information Science there, teaching courses in Programming Languages, Algorithms and Data Structures; Introduction to Algorithms and Data Structures; Knowledge Engineering, Machine Learning and Knowledge Discovery in Databases. He is the head of the Laboratory for Cognitive Modelling and a member of the Artificial Intelligence Department at the same faculty. His research interests include artificial intelligence, machine learning, neural networks and cognitive modelling. He is the (co) author of 170 scientific papers in these fields and 10 textbooks. Professor Kononenko is a member of the editorial board of *Applied Intelligence* and *Informatica* journals and was also twice chair of the programme committee of the International Cognitive Conference in Ljubljana.

Matjaž Kukar studied computer science at the University of Ljubljana, Slovenia, receiving his BSc in 1993, MSc in 1996 and PhD in 2001. He is now the assistant professor at the Faculty of Computer and Information Science there and is also a member of the Artificial Intelligence Department at the same faculty. His research interests include knowledge discovery in databases, machine learning, artificial intelligence and statistics. Professor Kukar is the (co) author of over 50 scientific papers in these fields.

MACHINE LEARNING AND DATA MINING:

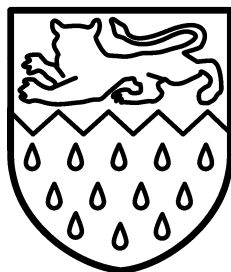
Introduction to Principles and Algorithms

Igor Kononenko

Professor of Computer and Information Science
University of Ljubljana, Slovenia

Matjaž Kukar

Assistant Professor of Computer and Information Science
University of Ljubljana, Slovenia



**Horwood Publishing
Chichester, UK**

HORWOOD PUBLISHING LIMITED

International Publishers in Science and Technology

Coll House, Westergate, Chichester, West Sussex, PO20 3QL, UK

First published in 2007

COPYRIGHT NOTICE

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the permission of Horwood Publishing, Coll House, Westergate, Chichester, West Sussex, PO20 3QL, UK.

© Horwood Publishing Limited, 2007.

British Library Cataloguing in Publication Data

A catalogue record of this book is available from the British Library

ISBN – 10: 1-904275-21-4

ISBN – 13: 978-1-904275-21-3

Cover design by Jim Wilkie.

Printed and bound in the UK by Antony Rowe Limited.

Table of Contents

Foreword	xv
Preface	xvii
1 Introduction	1
1.1 The name of the game	2
1.2 Overview of machine learning methods	5
1.2.1 Classification	5
1.2.2 Regression	9
1.2.3 Logical relations	11
1.2.4 Equations	13
1.2.5 Clustering	13
1.2.6 Reinforcement learning	14
1.3 History of machine learning	15
1.3.1 Symbolic rule learning	15
1.3.2 Neural networks	16
1.3.3 Reinforcement learning	17
1.3.4 Genetic algorithms	19
1.3.5 Statistical methods	20
1.3.6 Formal learning theory	20
1.4 Some early successes	21
1.4.1 Automatic Mathematician	21
1.4.2 EURISKO	22
1.4.3 DENDRAL and META-DENDRAL	22
1.4.4 Model Inference System	23
1.5 Applications of machine learning	24
1.5.1 Diagnosing the production process	24
1.5.2 Medical diagnosis	25
1.5.3 Risk evaluation of insurance and loan applicants	25
1.5.4 Image classification	26
1.5.5 Predicting the structure of chemical compounds	27
1.5.6 Game playing	28
1.6 Data mining tools and standards	29

1.6.1	Data Mining standards	29
1.6.2	Data Mining tools	31
1.7	Summary and further reading	34
2	Learning and Intelligence	37
2.1	What is learning	37
2.1.1	Definition of learning	37
2.1.2	Artificial intelligence research	38
2.2	Natural learning	39
2.2.1	Innate and learned	40
2.2.2	Memory	41
2.2.3	Types of natural learning	43
2.3	Learning, intelligence, consciousness	47
2.3.1	Amount of intelligence	47
2.3.2	Consciousness	48
2.3.3	Limits of symbolic computability	49
2.3.4	Possibility of artificial intelligence	50
2.3.5	(Im)possibility of artificial consciousness	51
2.3.6	Science and philosophy	53
2.4	Why machine learning	55
2.5	Summary and further reading	57
3	Machine Learning Basics	59
3.1	Basic principles	59
3.1.1	Learning as modeling	59
3.1.2	Minimum description length principle	62
3.1.3	Incremental (on-line) learning	64
3.1.4	Principle of multiple explanations	65
3.1.5	Estimating probabilities	66
3.2	Measures for performance evaluation	68
3.2.1	Classification accuracy and confusion matrix	70
3.2.2	Misclassification costs	71
3.2.3	Class probability estimation	72
3.2.4	Performance evaluation for two-class problems	74
3.2.5	Performance evaluation in regression	78
3.2.6	Bias and variance	79
3.3	Estimating performance	81
3.3.1	Reliability of quality estimations	81
3.3.2	Confidence interval	82
3.3.3	Cross validation	83
3.3.4	Bootstrap methods	84
3.4	*Comparing performance of ML algorithms	84
3.4.1	Two algorithms on a single domain	85
3.4.2	Two algorithms on several domains	89
3.4.3	Several algorithms on several domains	90

3.5	Combining several ML algorithms	91
3.5.1	Combining predictions of several hypotheses	92
3.5.2	Combining algorithms	93
3.5.3	Bagging, boosting and random forests	94
3.5.4	Transduction in machine learning	95
3.5.5	Cost-sensitive learning	97
3.5.6	Function imitation	99
3.5.7	Classifiers for regression and regressors for classification	99
3.5.8	Error-correcting output codes	100
3.6	Summary and further reading	102
4	Knowledge Representation	107
4.1	Propositional calculus	108
4.1.1	Representation with attributes	108
4.1.2	Attribute properties and dependencies	108
4.1.3	Classification, regression and association rules	110
4.1.4	Generality of rules	111
4.1.5	Operations on rules	112
4.1.6	Hypothesis space	112
4.1.7	Decision and regression trees	113
4.2	*First order predicate calculus	114
4.2.1	Prolog programming language	115
4.2.2	Inductive logic programming	117
4.2.3	Operations on predicates	118
4.3	Discriminant and regression functions	121
4.3.1	Linear functions	122
4.3.2	Square and Φ functions	123
4.3.3	Artificial neural networks	123
4.4	Probability distributions	126
4.4.1	Bayesian classifier	126
4.4.2	Learning examples as probability distribution	127
4.4.3	Naive Bayesian classifier	127
4.5	Summary and further reading	128
5	Learning as Search	131
5.1	Exhaustive search	133
5.1.1	Breadth-first search	133
5.1.2	Depth-first search	134
5.1.3	Iterative deepening	134
5.2	Bounded exhaustive search (branch and bound)	135
5.2.1	Bounded breadth-first search	136
5.2.2	Bounded depth-first search	138
5.3	Best-first search	138
5.4	Greedy search	139
5.5	Beam search	140

5.6	Local optimization	141
5.7	Gradient search	142
5.8	Simulated annealing	143
5.8.1	Basic algorithm	143
5.8.2	Markovian neural networks	144
5.9	Genetic algorithms	146
5.9.1	Basic algorithm	146
5.9.2	Genetic algorithms in machine learning	148
5.10	Summary and further reading	150
6	Attribute Quality Measures	153
6.1	Measures for classification	154
6.1.1	Impurity functions	154
6.1.2	Measures based on information content	155
6.1.3	Some other measures	162
6.1.4	ReliefF	164
6.2	Measures for regression	168
6.2.1	Change of variance	168
6.2.2	Regressional ReliefF	169
6.2.3	MDL in regression	171
6.3	**Formal derivations and proofs	172
6.3.1	Properties of the attribute quality	172
6.3.2	Entropy is an impurity measure	173
6.3.3	Distance measure	174
6.3.4	J-measure	175
6.3.5	Gini-index	177
6.3.6	RELIEF and Gini-index	177
6.3.7	Variance and Gini-index	178
6.4	Summary and further reading	179
7	Data Preprocessing	181
7.1	Representation of complex structures	181
7.1.1	Representation of text documents	182
7.1.2	Representation of images	182
7.1.3	Representation of graphs	183
7.2	Discretization of continuous attributes	184
7.2.1	Types of discretization	184
7.2.2	Controlling discretization	185
7.2.3	Fuzzy discretization	185
7.3	Attribute binarization	186
7.3.1	Two-class problems	187
7.3.2	Multi-class problems	188
7.4	Transforming discrete attributes into continuous	188
7.5	Dealing with missing values	189
7.6	Visualization	189

7.6.1	Visualizing a single attribute	190
7.6.2	Visualizing pairs of attributes	190
7.6.3	Visualizing several attributes	193
7.6.4	Visualizing results of machine learning	194
7.7	Dimensionality reduction	199
7.7.1	Feature subset selection	199
7.7.2	Feature extraction	200
7.7.3	Principal components analysis	201
7.7.4	Independent component analysis	203
7.7.5	Factor analysis	206
7.7.6	Random projections	207
7.8	**Formal derivations and proofs	208
7.9	Summary and further reading	210
8	*Constructive Induction	213
8.1	Dependence of attributes	215
8.1.1	Dependence of events	215
8.1.2	Bias, variance and attribute dependence	216
8.1.3	Measuring attribute dependence	216
8.1.4	Visualization of attribute dependence	218
8.2	Constructive induction with pre-defined operators	219
8.3	Constructive induction without pre-defined operators	222
8.3.1	Hypothesis-based constructive induction	222
8.3.2	Constructive induction based on Cartesian products	222
8.4	Summary and further reading	225
9	Symbolic Learning	227
9.1	Learning of decision trees	227
9.1.1	Building and using decision trees	227
9.1.2	Building binary decision trees	229
9.1.3	Pruning decision trees	229
9.1.4	Missing and “don’t care” values	231
9.2	Learning of decision rules	231
9.2.1	Generating a single rule	232
9.2.2	Generating a set of rules	232
9.3	Learning of association rules	233
9.3.1	Support and confidence	233
9.3.2	APriori algorithm	234
9.3.3	Improving APriori	235
9.4	Learning of regression trees	236
9.4.1	Building and using regression trees	236
9.4.2	Pruning of regression trees	238
9.5	*Inductive logic programming	239
9.5.1	Generating a set of rules in FOIL	239
9.5.2	ReliefF for ILP	241

9.6	Naive and semi-naive Bayesian classifier	242
9.6.1	Naive Bayesian classifier	242
9.6.2	Explaining decisions	243
9.6.3	Semi-naive Bayesian classifier	244
9.6.4	Practical considerations	248
9.7	Bayesian belief networks	249
9.7.1	Bayesian belief networks for classification	250
9.7.2	Classification with Bayesian belief networks	252
9.7.3	Learning the Bayesian network topology	253
9.7.4	Learning tree-augmented naive Bayesian classifiers	255
9.8	Summary and further reading	256
10	Statistical Learning	259
10.1	Nearest neighbors	259
10.1.1	K nearest neighbors	259
10.1.2	Weighted k -nearest neighbors	261
10.1.3	Locally weighted regression	261
10.2	Discriminant analysis	262
10.2.1	Optimal discriminant function	263
10.2.2	Classification with discriminant function	264
10.3	Linear regression	265
10.4	Logistic regression	266
10.5	*Support vector machines	267
10.5.1	Basic SVM principle	268
10.5.2	Inexact classification	271
10.5.3	SVM method	271
10.6	Summary and further reading	273
11	Artificial Neural Networks	275
11.1	Introduction	275
11.1.1	Simple example	275
11.1.2	Distributed memory	278
11.1.3	Properties of ANN	280
11.1.4	Applications	282
11.1.5	Analogy with brain	283
11.1.6	Relation to symbolic AI methods	284
11.2	Types of artificial neural networks	285
11.2.1	Topology	285
11.2.2	Application purpose	286
11.2.3	Learning rule	288
11.2.4	Combination function	289
11.3	*Hopfield's neural network	291
11.3.1	Discrete model	291
11.3.2	Continuous model	292
11.4	*Bayesian neural network	293

11.4.1	Topology and learning rule	293
11.4.2	Discrete models	294
11.4.3	Implementation of Bayesian classifiers	296
11.4.4	Continuous models	297
11.5	Perceptron	298
11.5.1	Two-layered perceptron	299
11.5.2	Multi-layered perceptron	302
11.5.3	Problems with backpropagation	305
11.6	Radial basis function networks	306
11.6.1	Topology of RBF networks	307
11.6.2	RBF network execution	308
11.6.3	RBF network training	308
11.6.4	Some properties of RBF networks	309
11.7	**Formal derivations and proofs	310
11.7.1	Stability of Hopfield's models	310
11.7.2	Stability of BNN models	312
11.7.3	Backpropagation of errors	315
11.8	Summary and further reading	317
12	Cluster Analysis	321
12.1	Introduction	321
12.1.1	Clustering and classification	322
12.1.2	Taxonomy of clustering algorithms	323
12.2	Measures of dissimilarity	324
12.2.1	Dissimilarity measures for discrete data	325
12.2.2	Dissimilarity measures for continuous data	326
12.2.3	Dissimilarity measures for mixed data	327
12.2.4	Dissimilarity measures for groups of examples	328
12.2.5	Construction and choice of measures	329
12.3	Hierarchical clustering	330
12.3.1	Agglomerative hierarchical clustering	330
12.3.2	Divisive hierarchical clustering	333
12.3.3	Practical considerations	336
12.4	Partitional clustering	338
12.4.1	Partitioning criteria	338
12.4.2	Partitional algorithms	340
12.4.3	Number of clusters	345
12.5	Model-based clustering	346
12.5.1	Finite mixture models	347
12.5.2	Maximum likelihood estimation and EM algorithm	347
12.6	Other clustering methods	350
12.6.1	Fuzzy clustering	350
12.6.2	Conceptual clustering	351
12.6.3	Evolutionary clustering	353
12.6.4	Knowledge-based clustering	354

12.7	Summary and further reading	356
13	** Learning Theory	359
13.1	Computability theory and recursive functions	361
13.1.1	Computability theory	361
13.1.2	Recursive functions	364
13.1.3	Undecidability of universal formalisms	365
13.1.4	Recursively enumerable sets	366
13.2	Formal learning theory	368
13.2.1	Identification	368
13.2.2	Basic theorems	370
13.3	Properties of learning functions	372
13.3.1	Learning strategies	372
13.3.2	Computational constraints	373
13.3.3	Infinite languages	376
13.3.4	Some sensible learning strategies	378
13.3.5	Reliability and confidence	381
13.4	Properties of input data	382
13.4.1	Evidential relations	382
13.4.2	Noisy and incomplete texts	382
13.4.3	Texts with positive and negative examples	384
13.4.4	Texts with oracles	385
13.5	Convergence criteria	386
13.5.1	Convergence	386
13.5.2	Extensional and bounded extensional convergence	387
13.5.3	Convergence towards simple hypotheses	388
13.5.4	Convergence towards recursive hypotheses	388
13.6	Implications for machine learning	389
13.7	Summary and further reading	390
14	** Computational Learning Theory	393
14.1	Introduction	393
14.2	General framework for concept learning	395
14.2.1	Basic notation	395
14.2.2	Some typical concept classes	396
14.2.3	Importance of representation	397
14.3	PAC Learning Model	398
14.3.1	Basic PAC learning model	398
14.3.2	Some variations of the PAC learning model	399
14.4	Vapnik-Chervonenkis dimension	400
14.4.1	Calculating the VC-Dimension	400
14.4.2	Calculating VC-dimension of complex concept classes	402
14.4.3	VC-dimension and PAC learning model	402
14.4.4	Occam learning	403

14.5	Learning in the presence of noise	406
14.5.1	Models of noise	406
14.5.2	Gaining noise tolerance with random classification noise . . .	407
14.5.3	Some results on more complicated models of noise	409
14.6	Exact and mistake bounded learning models	410
14.6.1	On-line learning model	410
14.6.2	Query learning model	414
14.7	Inherent unpredictability and PAC-reductions	415
14.7.1	Hard problems	415
14.7.2	Reducibility in PAC learning	416
14.8	Weak and strong learning	417
14.8.1	Weak and strong learnability	417
14.8.2	Boosting the confidence	418
14.8.3	Boosting the accuracy	418
14.9	Summary and further reading	420
A	*Definitions of some lesser known terms	423
A.1	Computational complexity classes	423
A.1.1	Complexity class P	424
A.1.2	Complexity class NP	424
A.1.3	Complexity class RP	424
A.2	Asymptotic notation	425
A.3	Some bounds for probabilistic analysis	427
A.3.1	Union bound	427
A.3.2	Chernoff bounds	427
A.4	Covariance matrix	428
	References	429
	Index	447

Foreword

In the past two decades, Machine Learning (ML) has made, among the many areas of Artificial Intelligence (AI), probably the most significant progress, and has grown into a field of its own. In this period, many of its methods found their way into practical applications, and in many application areas are now routinely used. The wide spread and visibility of ML was accelerated by powerful new techniques, and also by the appearance of effective and user-friendly implementations of ML tools, including some excellent freely available platforms, such as Weka and Orange. All this has resulted in ML, together with its related areas of Data Mining, Knowledge Discovery in Databases, and Intelligent Data Analysis, becoming an effective tool for many other disciplines as diverse as engineering, medicine, biology, ecology, biochemistry and finance. In science, it is becoming part of general scientific methodology and is now used by scientists in interpreting experimental data and in formulating new scientific theories. ML is beginning to be taught outside Computer Science and is on the way to attaining a similar status as some other subjects of general importance, in the same way as it is, for example, useful to students of medicine and biology to learn the basics of statistics.

This book is written by two scientists who have made their own original contributions to the methods and applications of ML, one of the co-authors (I. K.) having been involved since its early stages. In my view, the main strength of this book is its breadth of coverage and wealth of material presented, without loss of depth. In this respect, the book is rather unique among the books on machine learning. The book covers a large majority of key methods, concepts and issues. The style of writing is terse and directly to the point.

In comparison with other books on ML, the following chapters seem to be particularly strong. Chapter 3 systematically introduces a repertoire of basic elements that are universally used as components of ML methodology. Chapter 6 gives an outstanding review of measures for attribute selection. Chapter 8 introduces constructive induction, a topic that is rarely covered in books. In Chapters 13 and 14 on the theory of ML, the book addresses deep theoretical questions of learnability.

This book will be useful reading both for beginners to the field, and even more so to people who have already done some work in ML, but would like to broaden their knowledge and are looking for reference to learn about yet unfamiliar topics.

Ivan Bratko

Preface

In learning, each day something is gained. In following the Tao, each day something is lost.

— Lao Tse

The book describes the basics of machine learning principles and algorithms that are used in data mining and is intended for students of computer science, researchers who need to adapt various algorithms for particular data mining tasks, and for education of sophisticated users of machine learning and data mining tools.

Machine learning has exhibited tremendous progress in the last two decades. Numerous systems for machine learning are nowadays available and the number of their applications in various fields, such as industry, medicine, economics, ecology, finance, and many others, is rapidly growing. Machine learning is used for data analysis and knowledge discovery in databases, data mining, automatic generation of knowledge bases for expert systems, game playing, text classification and text mining, automatic recognition of speech, handwriting, images, etc.

The basic principle of machine learning is modeling of available data. Results of various machine learning systems may be rules, functions, relations, equation systems, probability distributions and other knowledge representations. The obtained “knowledge” aims to explain the data and can be used for forecasting, diagnostics, control, validation, and simulations.

Several excellent books have already been written on machine learning and data mining, most notably (Mitchell, 1997; Witten and Frank, 2000, 2005; Hand et al., 2001; Alpaydin, 2005). The purpose of our book is to provide a thorough and yet a comprehensible general overview of the field. Our approach differs from other textbooks by providing several different points of view to machine learning and data mining, as reflected from the structure of the book:

- Besides defining the basic notions and overviewing the machine learning methods, Chapter 1 provides the historical overview of the field and an overview of the state-of-the-art data mining tools as well as established and emerging standards.
- Chapter 2 is the least formal and introduces the psychological and philosophical issues related to machine and human learning, intelligence, and consciousness.
- Basic principles of machine learning, described in Chapter 3, are the central issue and can be considered as the core of the book.

- The next three chapters divide the machine learning process into three basic components: the representation of knowledge (Chapter 4) defines the hypothesis space, Chapter 5 describes basic search algorithms that can be used to search the hypothesis space, and the attribute quality measures (Chapter 6) are used to guide the search.
- Chapters 7 and 8 add to the whole story formal and practical guidelines for preparing, cleansing and transforming the input data in order to get the most out of it with the available machine learning algorithms. Constructive induction, described in Chapter 8, has great potentials, however, in the past not many researchers or implementations considered it worthwhile. We believe that in the future, when more research will be devoted to this area, the advantages of constructive induction will be revealed and implemented in useful practical tools.
- The following four chapters use the principles and notions, described in preceding chapters, as building blocks in order to describe particular learning algorithms. Four chapters divide the field into four main approaches. Symbolic learning (Chapter 9) includes all approaches that result in symbolic and transparent knowledge representations. Statistical learning (Chapter 10), on the other hand, results in knowledge representations that are typically hard to understand and interpret by the end user. Artificial neural networks (Chapter 11) are inspired by biological neural networks. With respect to transparency of knowledge representation, artificial neural networks are much closer to statistical learning than to symbolic learning. Cluster analysis deserves a separate chapter (Chapter 12) as the task of clustering algorithms differs a lot from other machine learning tasks.
- The last two chapters introduce formal approaches to machine learning. Chapter 13 describes a mathematically and philosophically inspired problem of identification in the limit. The results of this formal theory provide the ultimate answers to philosophical questions which, however, are of modest practical use. On the other hand, the computational learning theory, described in Chapter 14, aims to provide theoretical answers to questions of more practical value.

In order to improve readability, we provide all the references in separate sections on further reading, that also contain chapter summaries. Obvious exceptions that naturally include references are, besides this preface, also historical Sections 1.3 and 1.4. We try to keep the text free of formal derivations. Where needed, we provide separate sections for formal derivations and proofs (Sections 6.3, 7.8, and 11.7). The diamond symbol \diamond in the text indicates that the proof of the assertion is omitted from the text and can be found at the end of the chapter in the corresponding section on formal derivations and proofs. Obvious exceptions are the last two chapters, where proofs can be found in the literature indicated in further reading. Those sections and chapters are intended for advanced readers. In the table of contents, advanced topics are indicated with * (moderately advanced) or ** (advanced). Appendix A defines some less known formal notions, which are occasionally used in the text.

Both authors have experience in teaching various parts of the material presented in this book, to undergraduate and postgraduate students. For the undergraduate course we suggest excluding sections and chapters marked with *: Section 3.4 (comparing performance of machine learning algorithms), sections related to inductive logic programming (Sections 4.2 and 9.5), Chapter 8 (constructive induction), Section 10.5 (support vector machines), Sections 11.3 and 11.4 (Hopfield's and Bayesian neural networks), as well as all sections and chapters marked with **. Optionally, Sections 9.3 (association rules) and 9.7 (Bayesian belief networks), and/or Chapter 12 (cluster analysis) can also be kept for the postgraduate course.

The book web page can be found at the address: `mldmbook.fri.uni-lj.si`, where all information and post-press corrections will be available. The readers' comments, suggestions and indications about errors in the text are welcome, and should be sent to either author's e-mail address: `igor.kononenko@fri.uni-lj.si` or `matjaz.kukar@fri.uni-lj.si`.

Acknowledgements

The book is the result of numerous years of active research in the field of machine learning. We are grateful to Professor Ivan Bratko for his pioneering efforts in establishing an outstanding machine learning research group in Ljubljana. We are grateful to the Faculty of Computer and Information Science for support, both during research and teaching. For long term collaboration in machine learning research we are particularly grateful to Matjaž Bevk, Zoran Bosnić, Ivan Bratko, Bojan Cestnik, Janez Demšar, Aleks Jakulin, Matevž Kovačič, Nada Lavrač, Uroš Pompe, Marko Robnik Šikonja, Luka Šajn, and Blaž Zupan. We thank Luka Šajn and Alan McConnell-Duff for proofreading the text, Jožica Robnik Šikonja for proofreading Chapter 2, Ivan Bratko and the anonymous reviewer for their comments and suggestions, and Bogdan Filipič, Marko Robnik Šikonja, and Blaž Zupan for their comments on earlier drafts of the book. Useful comments on various parts of the book were provided by several people. In particular, we thank Janez Demšar for his comments on Chapter 3, Aleks Jakulin for his comments on Chapter 8, Aristidis Likas and Eduardo Hruschka for their comments on Chapter 12, and France Dacar, Matevž Kovačič and Boštjan Vilfan for their comments on Chapter 13. The author of the “spider net” graphics from the book cover is a young artist and Igor's daughter Nina Kononenko. We thank also Francesca Bonner and Ellis Horwood from Horwood Publishing for their kind assistance and patience. Finally, we are grateful to our wives and children for their loving support and patience and to whom with all our love we dedicate this book.

Ljubljana, December 2006

Igor Kononenko and Matjaž Kukar