Introduction to Machine Learning

Second Edition

Adaptive Computation and Machine Learning

Thomas Dietterich, Editor
Christopher Bishop, David Heckerman, Michael Jordan, and Michael Kearns, Associate Editors

A complete list of books published in The Adaptive Computation and Machine Learning series appears at the back of this book.

Introduction to Machine Learning

Second Edition

Ethem Alpaydın

The MIT Press
Cambridge, Massachusetts
London, England

Typeset in 10/13 Lucida Bright by the author using LaTeX2ε.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Information

Alpaydin, Ethem.
Introduction to machine learning / Ethem Alpaydin. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-01243-0 (hardcover : alk. paper)
1. Machine learning. I. Title
Q325.5.A46 2010
006.3'1—dc22 2009013169
CIP

10 9 8 7 6 5 4 3 2 1

Preface

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We need learning in cases where we cannot directly write a computer program to solve a given problem, but need example data or experience. One case where learning is necessary is when human expertise does not exist, or when humans are unable to explain their expertise. Consider the recognition of spoken speech—that is, converting the acoustic speech signal to an ASCII text; we can do this task seemingly without any difficulty, but we are unable to explain how we do it. Different people utter the same word differently due to differences in age, gender, or accent. In machine learning, the approach is to collect a large collection of sample utterances from different people and learn to map these to words.

Another case is when the problem to be solved changes in time, or depends on the particular environment. We would like to have general-purpose systems that can adapt to their circumstances, rather than explicitly writing a different program for each special circumstance. Consider routing packets over a computer network. The path maximizing the quality of service from a source to destination changes continuously as the network traffic changes. A learning routing program is able to adapt to the best path by monitoring the network traffic. Another example is an intelligent user interface that can adapt to the biometrics of its user—namely, his or her accent, handwriting, working habits, and so forth.

Already, there are many successful applications of machine learning in various domains: There are commercially available systems for recognizing speech and handwriting. Retail companies analyze their past sales data to learn their customers' behavior to improve customer rela-

tionship management. Financial institutions analyze past transactions to predict customers' credit risks. Robots learn to optimize their behavior to complete a task using minimum resources. In bioinformatics, the huge amount of data can only be analyzed and knowledge extracted using computers. These are only some of the applications that we—that is, you and I—will discuss throughout this book. We can only imagine what future applications can be realized using machine learning: Cars that can drive themselves under different road and weather conditions, phones that can translate in real time to and from a foreign language, autonomous robots that can navigate in a new environment, for example, on the surface of another planet. Machine learning is certainly an exciting field to be working in!

The book discusses many methods that have their bases in different fields: statistics, pattern recognition, neural networks, artificial intelligence, signal processing, control, and data mining. In the past, research in these different communities followed different paths with different emphases. In this book, the aim is to incorporate them together to give a unified treatment of the problems and the proposed solutions to them.

This is an introductory textbook, intended for senior undergraduate and graduate-level courses on machine learning, as well as engineers working in the industry who are interested in the application of these methods. The prerequisites are courses on computer programming, probability, calculus, and linear algebra. The aim is to have all learning algorithms sufficiently explained so it will be a small step from the equations given in the book to a computer program. For some cases, pseudocode of algorithms are also included to make this task easier.

The book can be used for a one-semester course by sampling from the chapters, or it can be used for a two-semester course, possibly by discussing extra research papers; in such a case, I hope that the references at the end of each chapter are useful.

The Web page is http://www.cmpe.boun.edu.tr/~ethem/i2ml/ where I will post information related to the book that becomes available after the book goes to press, for example, errata. I welcome your feedback via email to alpaydin@boun.edu.tr.

I very much enjoyed writing this book; I hope you will enjoy reading it.