SPRINGER BRIEFS IN COMPUTER SCIENCE

M.N. Murty
Rashmi Raghava

Support Vector Machines and Perceptrons

Learning, Optimization, Classification, and Application to Social Networks

Springer

SpringerBriefs in Computer Science

More information about this series at http://www.springer.com/series/10028

M.N. Murty · Rashmi Raghava

Support Vector Machines and Perceptrons

Learning, Optimization, Classification, and Application to Social Networks

Springer

M.N. Murty
Department of Computer Science
and Automation
Indian Institute of Science
Bangalore, Karnataka
India

Rashmi Raghava
IBM India Private Limited
Bangalore, Karnataka
India

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-319-41062-3 ISBN 978-3-319-41063-0 (eBook)
DOI 10.1007/978-3-319-41063-0

Library of Congress Control Number: 2016943387

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Overview

Support Vector Machines (SVMs) have been widely used in *Classification, Clustering and Regression*. In this book, we deal primarily with classification. Classifiers can be either *linear* or *nonlinear*. The linear classifiers typically are learnt based on a *linear discriminant function* that separates the feature space into two half-spaces, where one half-space corresponds to one of the two classes and the other half-space corresponds to the remaining class. So, these half-space classifiers are ideally suited to solve *binary classification* or two-class classification problems. There are a variety of schemes to build multiclass classifiers based on combinations of several binary classifiers.

Linear discriminant functions are characterized by a *weight vector* and a *threshold* weight that is a scalar. These two are learnt from the training data. Once these entities are obtained we can use them to classify patterns into any one of the two classes. It is possible to extend the notion of linear discriminant functions (LDFs) to deal with even nonlinearly separable data with the help of a suitable mapping of the data points from the low-dimensional *input* space to a possibly higher dimensional *feature space*.

Perceptron is an early classifier that successfully dealt with linearly separable classes. Perceptron could be viewed as the simplest form of *artificial neural network*. An excellent theory to characterize parallel and distributed computing was put forth by *Misky and Papert* in the form of a book on perceptrons. They use logic, geometry, and group theory to provide a computational framework for perceptrons. This can be used to show that any computable function can be characterized as a linear discriminant function possibly in a high-dimensional space based on *minterms* corresponding to the input Boolean variables. However, for some types of problems one needs to use all the minterms which correspond to using an exponential number of minterms that could be realized from the primitive variables.

SVMs have revolutionized the research in the areas of *machine learning* and *pattern recognition*, specifically classification, so much that for a period of more

than two decades they are used as state-of-the-art classifiers. Two distinct properties of SVMs are:

1. The problem of learning the LDF corresponding to SVM is posed as a convex optimization problem. This is based on the intuition that the hyperplane separating the two classes is learnt so that it corresponds to maximizing the *margin* or some kind of separation between the two classes. So, they are also called as *maximum-margin classifiers*.
2. Another important notion associated with SVMs is the *kernel trick* which permits us to perform all the computations in the low-dimensional input space rather than in a higher dimensional feature space.

These two ideas become so popular that the first one lead to the increase of interest in the area of *convex optimization*, whereas the second idea was exploited to deal with a variety of other classifiers and clustering algorithms using an appropriate kernel/similarity function.

The current popularity of SVMs can be attributed to excellent and popular software packages like *LIBSVM*. Even though SVMs can be used in nonlinear classification scenarios based on the kernel trick, the linear SVMs are more popular in the real-world applications that are high-dimensional. Further learning the parameters could be time-consuming. There is a renewal of energy, in the recent times, to examine other linear classifiers like perceptrons. Keeping this in mind, we have dealt with both perceptron and SVM classifiers in this book.

Audience

This book is intended for senior undergraduate and graduate students and researchers working in *machine learning, data mining, and pattern recognition*. Even though SVMs and perceptrons are popular, people find it difficult to understand the underlying theory. We present material in this book so that it is accessible to a wide variety of readers with some basic exposure to undergraduate level mathematics. The presentation is intentionally made simpler to make the reader feel comfortable.

Organization

This book is organized as follows:

1. *Literature and Background*: Chapter 1 presents literature and state-of-the-art techniques in SVM-based classification. Further, we also discuss relevant background required for pattern classification. We define some of the important terms that are used in the rest of the book. Some of the concepts are explained with the help of easy to understand examples.

2. *Linear Discriminant Function*: In Chap. 2 we introduce the notion of *linear discriminant function* that forms the basis for the linear classifiers described in the text. The role of *weight vector W* and the *threshold b* are explained in describing linear classifiers. We also describe other linear classifiers including the *minimal distance classifier* and the *Naïve* Bayes classifier. It also explains how nonlinear discriminant functions could be viewed as linear discriminant functions in higher dimensional spaces.
3. *Perceptron*: In Chap. 3 we describe perceptron and how it can be used for classification. We deal with *perceptron learning algorithm* and explain how it can be used to learn Boolean functions. We provide a simple proof to show how the algorithm converges. We explain the notion of *order of a perceptron* that has bearing on the computational complexity. We illustrate it on two different classification datasets.
4. *Linear SVM*: In this Chap. 4, we start with the similarity between SVM and perceptron as both of them are used for *linear classification*. We discuss the difference between them in terms of the form of computation of w, the optimization problem underlying each, and the *kernel trick*. We introduce the linear SVM which possibly is the most popular classifier in machine learning. We introduce the notion of *maximum margin* and the geometric and semantic interpretation of the same. We explain how a binary classifier could be used in building a multiclass classifier. We provide experimental results on two datasets.
5. *Kernel Based SVM*: In Chap. 5, we discuss the notion of *kernel* or similarity function. We discuss how the optimization problem changes when the classes are not linearly separable or when there are some data points on the margin. We explain in simple terms the *kernel trick* and explain how it is used in classification. We illustrate using two practical datasets.
6. *Application to Social Networks*: In Chap. 6 we consider *social networks*. Specifically, issues related to representation of social networks using *graphs*; these graphs are in turn represented as matrices or lists. We consider the problem of community detection in social networks and *link prediction*. We examine several existing schemes for link prediction including the one based on SVM classifier. We illustrate its working based on some network datasets.
7. *Conclusion*: We conclude in Chap. 7 and also present potential future directions.

Bangalore, India

M.N. Murty
Rashmi Raghava

Contents