Workshop
"Introduction and Developments on Support Vector Machines and Classification"

Econometric Institute
Erasmus School of Economics, Erasmus University Rotterdam

Friday, September 12, 2008

Organizers: Patrick Groenen, Georgi Nalbantov, Cor Bioch
Venue: Erasmus University Rotterdam, Burg. Oudlaan 50, room J1-41

Support vector machines (SVM) have become a valuable method for the prediction of two classes. Often, the SVM is the best method to classify the two groups. One of the aims of this workshop is to introduce the technique of support vector machines. Another aim is to discuss the use of kernels to allow for nonlinearity of the predictor variables. This workshop brings together several leading researchers active this area in the Netherlands and two foreign guests, Sarel Steel (Stellenbosch University, South Africa) and Thorsten Joachims (Cornell University NY, USA).

Program:

10:00 – 10:30	Welcome with coffee
10:30 – 11:10	**Ida Sprinkhuizen-Kuyper** (Radboud University Nijmegen) *Introduction to Support Vector Machines, a powerful tool for classification and prediction*
11:10 – 11:50	**Patrick Groenen** (Erasmus University Rotterdam) *Support Vector Machines using Majorization and Kernels*
11:50 – 12:00	Coffee
12:00 – 12:40	**Evgueni N. Smirnov** (Maastricht University) *Version Space Support Vector Machines*
12:40 – 13:45	Lunch
13:45 – 14:25	**Georgi Nalbantov** (Maastricht University): *Instance-based Penalization Methods for Classification*
14:25 – 15:05	**Sarel Steel** (Stellenbosch University, South Africa) *Variable selection for kernel methods*
15:05 – 15:30	Break
15:30 – 16:20	**Thorsten Joachims** (Cornell University NY, USA) *Support Vector Machines for Structured Output Prediction*
16.20	Drinks

Participation in the workshop is free. To know the number of participants in advance, please register by sending an e-mail to Elli Hoek van Dijke (hoekvandijke@few.eur.nl).

This workshop is supported by Erasmus Research Institute of Management (ERIM) and organized in cooperation with School for Information and Knowledge Systems (SIKS)

Abstracts

Presenter: Ida Sprinkhuizen-Kuyper (Radboud University Nijmegen)
Title: Introduction to Support Vector Machines, a powerful tool for classification and prediction
Abstract
I will give an introduction of the most important notions for understanding Support Vector Machines and how to use them.

Presenter: Patrick Groenen (Erasmus University Rotterdam)
Title: *Support Vector Machines using Majorization and Kernels*
Abstract
Support vector machines have become one of the main stream methods for two-group classification. In Groenen, Nalbantov, and Bioch (2007, 2008), we proposed SVM-Maj, a majorization algorithm that minimizes the SVM loss function.A big advantage of majorization is that in each iteration, the SVM-Maj algorithm is guaranteed to decrease the loss until the global minimum is reached. Nonlinearity was reached by replacing the predictor variables by their monotone spline bases and then doing a linear SVM. A disadvantage of the method so far is that if the number of predictor variables *m* is large, SVM-Maj becomes slow.

In this paper, we extend the SVM-Maj algorithm to handle efficiently cases where the number of observations *n* is (much) smaller than *m*. We show that the SVM-Maj algorithm can be adapted to handle this case of $n<<m$ s well. In addition, the use of kernels instead of splines for handling the nonlinearity becomes also possible while still maintaining the guaranteed descent properties of SVM-Maj.

Presenter: Evgueni N. Smirnov (Maastricht University)
Title: *Version Space Support Vector Machines*
Abstract
In this talk we consider version spaces as an approach to reliable classification. The key idea is to extend version spaces to contain the target hypothesis *t* or hypotheses similar to *t*. In this way, the unanimous-voting classification rule of version spaces does not misclassify; that is, instance classifications become reliable.

We propose to implement version spaces using support vector machines. The resulting combination is called version space support vector machines. Our experiments show that version space support vector machines are able to outperform the existing approaches to reliable classification.

Presenter: Georgi Nalbantov (Maastricht University)
Title: Instance-based Penalization Methods for Classification
Abstract
Three instance-based penalization methods for classification are presented: Support Hyperplanes, Nearest Convex Hull classifier, and Soft Nearest Neighbor. It is shown that the popular Support Vector Machine classifier can also be viewed as such a method. The relative merits between all of these classifiers are discussed, as well as their theoretical foundations. An emphasis is put on the role of penalization in general and why it is imperative to penalize (the coefficients of) unbiased methods to achieve better forecasting results.

Presenter: Sarel Steel (Stellenbosch University, South Africa)
Title: *Variable selection for kernel methods*
Abstract
Kernel methods such as the support vector machine (SVM) and kernel Fisher discriminant analysis (KFDA) are popular for analysing large data sets. The aim of variable selection is to identify important subsets of the input variables which yield better generalisation performance than the set consisting of all the input variables. In this presentation a brief overview of variable selection for kernel methods will be given.

Presenter: Thorsten Joachims (Cornell University)
Title: *Support Vector Machines for Structured Output Prediction*
Abstract
This talk explores large-margin approaches to predicting graph-based objects like trees, clusterings, or alignments. Such problems arise, for example, when a natural language parser needs to predict the correct parse tree for a given sentence, when one needs to determine the co-reference relationships of noun-phrases in a document, or when predicting the alignment between two proteins. In particular, the talk will show how structural SVMs can learn such complex prediction rules, using the problems of supervised clustering, protein sequence alignment, and diversification in search engines as application examples. Furthermore, the talk will present new cutting-plane algorithms that allows training of structural SVMs in time linear in the number of training examples.

Bio: Thorsten Joachims is an Associate Professor in the Department of Computer Science at Cornell University. In 2001, he finished his dissertation with the title "The Maximum-Margin Approach to Learning Text Classifiers: Methods, Theory, and Algorithms", advised by Prof. Katharina Morik at the University of Dortmund. From there he also received his Diplom in Computer Science in 1997 with a thesis on WebWatcher, a browsing assistant for the Web. From 1994 to 1996 he was a visiting scientist at Carnegie Mellon University with Prof. Tom Mitchell. His research interests center on a synthesis of theory and system building in the field of machine learning, with a focus on Support Vector Machines and machine learning with text. He authored the SVM-Light algorithm and software for support vector learning.

How to get to Erasmus University (Woudestein Campus)

By Car
Coming from the south (A16): First follow direction “Ring Rotterdam, Den Haag”. Near Ridderkerk, take directions “Feijenoord / Centrum / Kralingen / Capelle”. When crossing the Van Brienenoordbrug, take exit “Rotterdam Centrum / Capelle” and you will immediately arrive on a roundabout. On the roundabout take “Rotterdam Centrum”. Pass the Shell petrol station on your right side and take the first turn to the right after the Shell station. You have arrived at Burgemeester Oudlaan. Entrance of the Erasmus University main gate is 100 meters further on your right.

Coming from Utrecht and Den Haag: Follow directions “Ring Rotterdam / Dordrecht”, and then “Kralingen / Feijenoord / IJsselmonde” and take exit “Capelle / Centrum”. At the traffic lights follow “Rotterdam Centrum”. Pass the Shell petrol station on your right side and take the first turn to the right after the Shell station. You have arrived at Burgemeester Oudlaan. Entrance of the Erasmus main gate is 100 meters further on your right.

Parking is available on campus (parking fees will be charged).

By Public Transport
From Rotterdam Central Station “Centraal Station”: Take the underground metro direction “Spijkenisse”, change at “Beurs/Churchillplein”. Take direction “Ommoord / Nesselande / Capelle”, exit at “Kralingse Zoom”. When exiting the station, facing the main road, the Erasmus University is directly across the road.

You can also take tram number 7 in the direction “Woudestein” or tram number 21 in the direction “De Esch”, both of which leave at the front of the central station approximately every 10 minutes. The trip will take around 20 minutes. With tram 7, you should exit at the final stop “Woudestein”, just behind the J-building. With tram 21, you should exit at the stop “Woudestein” (in front of the football stadium; note that this is not the final stop!), from which it is a three-minute walk to the university campus.

Woudestein Campus Map