

CS 374 Assignment #5 Support Vector Machines

Due the week of March 3, 2008

This week, in which we study Support Vector Machines, continues our examination of linear classifiers – sort of. We consider a method that finds a linear separator in a high-dimensional feature space that is nonlinearly related to the input space.¹ That is, with this technique we can separate data that are, in fact, not linearly separable. Multilayer perceptrons were able to do this. There are several interesting new ideas here, though, including the fact that we don't explicitly have to consider the high-dimensional feature space into which we project the input data. Sound intriguing? Read on . . .

1 Support Vector Machines

1.1 Reading

Please read the following:

- Alpaydin, Sections 10.2 and 10.9,
- Alpaydin, Section 8.2 (especially 8.2.2),
- “Support vector machines”, from IEEE Intelligent Systems, July/August 1998.

1.2 Presentation of Support Vector Machines

There are many ways to gain real understanding of an algorithm – working through exercises, implementing the algorithm, etc. This week you will demonstrate your understanding by explaining Support Vector Machines (SVMs) in your tutorial session. You may work on your own or with your tutorial partner(s). Either way, I would like each group to be ready to present

- the rationale for maximizing the margin;
- the notion of a support vector;
- the derivation (at least at a high level) of SVMs;
- what the algorithm learns;
- how SVMs are used to classify new examples;
- kernel functions;
- (at a high level) how SMO works.

While you might find the readings above to be adequate, there are many other very good sources of information on SVMs. These include:

- A paper entitled “A Tutorial on Support Vector Machines for Pattern Recognition” by Christopher Burges, which appeared in the journal *Data Mining and Knowledge Discovery*, June 1998,

¹Description borrowed from Schölkopf.

- A paper by John Platt on the Sequential Minimal Optimization (SMO) algorithm,
- Lecture notes written by Andrew Ng at Stanford,
- Notes from a tutorial on “Support Vector and Kernel Machines” presented by Nello Cristianini at ICML 2001 (which was held here at Williams!).
- The book *Support Vector Machines and other kernel-based learning methods* by Nello Cristianini and John Shawe-Taylor.

Pointers to most of these can be found on the “Assignments” web page for this course. The book can be found on the “Machine Learning” shelf in the CS lounge.

There is no required length of presentation, but you might want to aim for 20-30 minutes. During your tutorial session I will also ask that you present solutions to the exercises below (which you also need to write up). You might find it useful to incorporate the exercises and their solutions into the presentation as examples of various ideas. (This will make your job of preparing the presentation a little easier and will also make our sessions more efficient.)

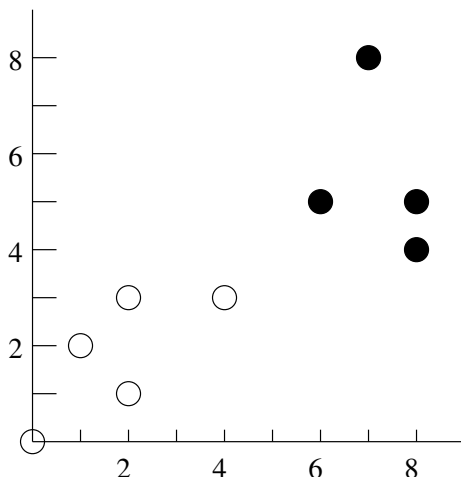
1.3 Exercises

1. The readings for this week told you that the quadratic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^2$ is equivalent to mapping each \mathbf{x} into a higher dimensional space where

$$\Phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$

for the case where $\mathbf{x} = (x_1, x_2)$. Now consider the cubic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^3$.

- (a) What is the corresponding Φ function for the case where $\mathbf{x} = (x_1, x_2)$?
 - (b) Compare the number of arithmetic operations involved in computing the cubic kernel to the number of operations involved in computing the corresponding Φ function followed by the dot product. Count additions and multiplications. You may assume that the square root of a number is given to you as a constant. Don't worry about optimizing – simply count the number of arithmetic operations assuming that the functions are computed in the most straightforward manner.
2. How many dot products need to be computed for M test points with N support vectors at test time?
 3. Say you are given the data set shown below. This is a binary classification task in which the instances are described by two real-valued attributes.



- (a) If the data in the graph are taken to be training data, which of the instances should be identified as support vectors?
- (b) Using Weka, train the SMO algorithm on the data above. The data can be found in

`~andrea/shared/cs374/svm-1.arff`

Recall from last week that typing

```
java -jar weka.jar
```

will start up a GUI. Click on the “Explorer” button. Then click on “Open file...” to select the data file. Now you’re almost ready to run SMO on the data set.

Click on “Classify” and then “Choose”. Now find SMO in the “functions” folder. Before you run SMO on the data, you’ll need to modify a couple of parameters. First, you want to be sure that SMO doesn’t normalize the input features. Next to the Classifier “Choose” button, you’ll see “SMO” along with a list of parameters. Click on the text. A window should pop up that gives you options for setting parameters. For “filterType”, choose “No normalization/standardization.” Click “OK”.

Next, select the “Use training set” as the “Test option”. This way the entire data set will be used for training.

Based on the results obtained from running the algorithm, what are the values of α_i for each of the nine instances (assuming that the algorithm identified the same support vectors that you did in part a).

4. Generate a two-dimensional example (assume two real-valued attributes and two possible classes) that cannot be linearly separated, but that can be separated by a polynomial kernel with degree two. You can test this with the SMO algorithm as implemented in Weka. (Just change the “exponent” parameter for the algorithm to 2.0.)
- (a) Give a graph representing your dataset.
- (b) Give the training error with the linear kernel and with the polynomial kernel of degree two. You should, of course, be able to achieve an error of 0 with the kernel of degree two. Depending on your choice of inputs, you might need to set “featureSpaceNormalization” to true.

2 Applications of SVMs

As the readings above indicate, SVMs have been successfully applied to many problems. This week you will have the opportunity to explore an application of SVMs.

Start by selecting one of the following papers:

- “Text Categorization with Support Vector Machines: Learning with Many Relevant Features”, T. Joachims. ECML-98, Tenth European Conference on Machine Learning, pp. 137–142.
- “Breast Cancer Survival and Chemotherapy: A Support Vector Machine Analysis”, Y.-J. Lee, O. L. Mangasarian, and W. H. Wolberg. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 55:1–20, 2000.
- “Classifying LEP Data with Support Vector Algorithms”, P. Vannerem, K.-R. Müller, B. Schölkopf, A. Smola, S. Söldner-Rembold. Procs. AIHENP99.

Links to all of these can be found on the “Assignments” web page.

Write a summary and critical analysis of the paper you’ve selected. (It should be 1.5-3 pages, double-spaced.) Be prepared to present the paper and your critique in the tutorial session.

The tutorial sessions might be most interesting if tutorial partners select different papers. While it isn’t necessary to do so, you might consider coordinating within each tutorial group.