

Homework 4

(Due: Mar 11th 2007)

SVM

Write programs in R to solve the following problems. You can use SVM functions such as `ksvm` (in package ```kernlab``` <http://cran.r-project.org/src/contrib/Descriptions/kernlab.html>), `svm` (in package ```e1071``` <http://cran.r-project.org/src/contrib/Descriptions/e1071.html>), or others of your own choice. Please email the source code to the TAs (Jiang Du: jiang.du@yale.edu; Edo Liberty: edo.liberty@yale.edu).

Problem 1 Different kernels

Predict the classes of (X, Y) pairs. First, build a *LINEAR SVM* based on the whole dataset, and measure its error rate. Plot the SVM with the data using corresponding plot functions. Second, *choose A SUITABLE KERNEL* for this problem and build another SVM based on the whole dataset, measure the error rate of the new model, and plot the new model with the data. (For this problem, you can turn off the ```scaling``` option when training the SVM, so that no preprocessing will take place.)

X	Y	Class
1	1	0
2	3	0
3	5	1
5	5	1
6	7	1
7	5	1
8	8	0
10	8	0
9	9	0

Problem 2 Feature selection

A In dataset ```bc_data.txt``` (http://www.gersteinlab.org/courses/545/07-spr/hw/bc_data.txt), there are 16 attributes X_1, X_2, \dots, X_{16} that can be incorporated to predict the corresponding Class. Instead of using all these attributes, develop your own method to choose a subset with *at most 6* of these attributes to build a *LINEAR SVM*. Compute the error rate of your SVM on the whole dataset. (*Hint*: obviously you can try all possible combinations of these attributes to find out a ```best``` subset, but that will be computationally intensive for large problems... http://en.wikipedia.org/wiki/Feature_selection should serve as a good starting point for you to develop a simple feature selection method. Actually, you can even use SVM to do feature selection if you really want to...)

B Suppose there is a dataset with K features that can be linearly separated, is it possible to linearly separate the data using only a subset of these K features? On the other hand, if a dataset can be linearly separated using only a subset of its K features, would the dataset still be linearly separable with all its features? (You can write the answers to this question as comments in your program for problem 2.A, so that you don't have to hand in a separate document.)