Machine Learning Lab5

SVM classification

In this lab, you are asked to implement the SVM algorithm firstly on a toy dataset, and then on the MNIST dataset. For the SVM classifier, you are suggested to use libsvm which can be downloaded from:

http://www.csie.ntu.edu.tw/~cjlin/libsvm/

or a local copy (already compiled) from:

http://www.viplab.cs.nott.ac.uk/course/G53MLE/libsvm-3.11.zip

2D Toy data: Suppose there are 7 points belong to two classes: (3, 3), (1, 3) and (2, 2.5) are positive data, (1, 1), (3, 1), (3, 2.5) and (4, 3) are negative data.

1. Use SVM to classify these 7 points and try to visualize the results. The desired visualized results of SVM classifier should be like this:
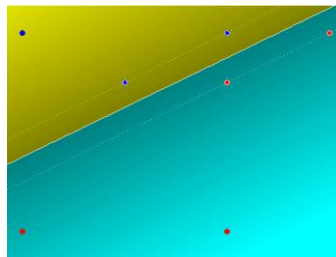
2. Suppose (1.5, 1.5) is another positive data and (1, 2) is another negative data. Try to select a suitable kernel function to obtain a nonlinear classifier to classify these 9 points. The desired visualized results should be like this.
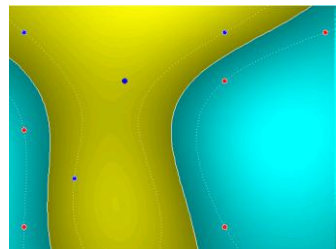
3. Change the value of the slack variable and the parameters of the kernel function and try to understand their influences on the final results.
4. Try to use linear SVM and different kinds of kernel SVM on the MNIST dataset, see what accuracy you could get.

Random Forest classification

1. Try to implement a Random Forest (RF) classifier on the MNIST dataset. Please analyze by yourself how the parameters in the RF, (including the number of trees, the maximum depth of each tree, the minimum samples contained in each leaf node, etc) influence the RF performance.

2. You can also perform Principle Component Analysis (PCA) on the samples first. Will this procedure help boosting the RF performance?