

Machine Learning - Assignment 5

Due Thursday, November 4

In this assignment you will experiment with either SVMs or ensemble methods, There are 3 proposed tasks described below. Pick one of them and turn in the required report. These tasks are not specified in great detail - there is room to experiment according to your own interests.

1. Support vector machines

For this task, the main goal is to get a bit more familiar with SVMs and the different kinds of kernels. Several software packages are available from:

`www.kernel-machines.org`

I would recommend using SVMlight, which can be downloaded from:

`http://svmlight.joachims.org/`

There are several example problems with the software, or you can choose your own problem. In this case, you may need to write a little script to transform the data file into the format used by SVMlight.

Design an experiment to test one (or both) of the following aspects:

- The type of kernel used
- The trade-off between the training error and the margin

Examine the effects on the training/test error, and on the number of support vectors generated. Report your findings in a short write-up. Detail the package and data used, and the parameter settings.

2. Boosting

Less code is available in this case, but one nice applet is available at:

`http://www.cs.technion.ac.il/~rani/LocBoost/index.html`

It allows you to compare boosting, bagging and some basic classifiers on 2D data. You should design your own experimental procedure to compare and contrast some of these algorithms.

3. **Bagging** For this task, you are required to write a small bagging script (or program). Pick a basic classifier we learned about (e.g., C4.5). Pick five tasks from the UCI data repository. Compare the accuracy of the basic classifier with the bagged version on these tasks. Alternatively, you could create an artificial data set, then perturb it with different amounts of noise. Again, compare the basic and bagged classifiers in terms of accuracy and discuss the effect of the noise on each.

In your report you should describe the tasks you used, the empirical methodology and your results (presented either as a graph or as a table). Explain the behavior you observe, based on the properties of bagging we discussed in class. Include your bagging code as an annex to this little report.