Machine Learning *CS 689, Spring 2015*

HW05: Beyond binary classification

Hand in via moodle at: `https://moodle.umass.edu/course/view.php?id=20836`. Remember that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running "`pdflatex hw00.tex`". You'll need mydefs.sty and notes.sty which can be downloaded from the course page.

1. At the face of it, OVO seems more computationally intensive at training time than OVA because it trains $O(K^2)$ classifiers rather than $O(K)$ classifiers. However, all of the K-many OVA classifiers are on the full data set of N examples, while the $O(K^2)$ OVO classifiers are only on subsets of the data. Suppose that you have N data points, divided evenly into K classes (so that there are N/K examples per class).

   (a) Suppose that the training time for your binary classifier is linear in the number of examples it receives. What is the complexity of training OVO and OVA, as a function of N and K?

   (b) Suppose the training time is quadratic; then what is the complexity of OVO and OVA?

2. We are given a binary classifier f that can predict which of the two items a or b is ranked higher. From this we can obtain a ranking of N items by considering all pairs and ordering items them based on how many times they were ranked higher. This method requires $O(N^2)$ evaluations of f. Describe how one might rank these items using $O(N \log N)$ evaluations of f?

3. Explain why stacked classifiers are more susceptible to overfitting than regular classifiers.