Machine Learning *CS 726, Fall 2011*

HW04: Complex predictions

Hand in at: `http://www.cs.utah.edu/~hal/handin.pl?course=cs726`. 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`".

1. Define, in a manner analogous to the way Tasks are defined in Chapter 5, the regression problem under squared loss (refer back to Section 1.4 if you need to).

2. All of the theoretical results for complex classification say something like "if a binary classifier gets error at most ϵ, then the error on my more complex problem will be at most $g(\epsilon)$" (where g is whatever is appropriate for the particular algorithm). Hopefully you realize that there are multiple types of error that matter, for instance: training error and expected test error. To what type(s) of error do these theorems apply?

3. At the face of it, AVA seems more computationally intensive at training time than OVA because it trains $\mathcal{O}(K^2)$ classifiers rather than $\mathcal{O}(K)$ classifiers. However, all of the K-many OVA classifiers are on the full data set of N examples, while the $\mathcal{O}(K^2)$ AVA 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 OVA and AVA, as a function of N and K?

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

4. Define a ranking preference function ω that penalizes mispredictions *linearly* up to a threshold K. In other words, for $K = 20$, if I put the object that should be in position 5 in position 20, then I pay \$15; if I put it in position 30, I only pay \$20 because nothing costs more than $K = \$20$.