Machine Learning *CS 726, Fall 2011*

HW03: Perceptrons and evaluation

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. The perceptron will only converge if the data is linearly separable. For linearly separable data, if $||\boldsymbol{x}|| \leq R$ for all data points, then it will converge in at most R^2/γ^2 iterations. It is possible to *force* your data to be linearly separable as follows. If you have N data points in D dimensions, map data point $\boldsymbol{x}_n$ to the $(D+N)$-dimensional point $\langle \boldsymbol{x}_n, e_n \rangle$, where e_n is a N-dimensional vector of zeros, except for the nth position, which is 1. (Eg., $e_4 = \langle 0, 0, 0, 1, 0, \ldots \rangle$.)

   (a) Show that if you apply this mapping the data becomes linearly separable (you may wish to do so by providing a weight vector $\boldsymbol{w}$ in $(D+N)$-dimensional space that successfully separates the data).

   (b) How long will it take the peceptron algorithm to converge on this augmented data?

   (c) How does this mapping affect generalization?

2. Why is averaging favored over voting?

3. For each of { centering, variance scaling } and each of { decision trees, KNN, perceptron }, state whether the given preprocessing will affect the classifier or not.