Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 4
Due: April 11, 2006

Problem 1: Perceptron

(1) Prove that the mistake bound given for the perceptron algorithm is tight. To do that, find a sequence of m examples $(x_1, y_1), \ldots, (x_m, y_m) \in X \times \{-1, +1\}$ linearly separable, such that $\|x_i\| \leq R$ for $i = 1, \ldots, m$, and such that the number of updates made by the perceptron algorithm is exactly $\frac{R^2}{\rho^2}$.

Problem 2: Boosting

(1) Let h_t be the base classifier selected by Adaboost at round t. What is the error rate of h_t with respect to the distribution D_{t+1}? Use this and the definition of weak learners to conclude that $h_{t+1} \neq h_t$.

(2) Adaboost may lead to significant overfitting in presence of noise. This could be because of the high penalization of misclassified examples. To reduce that effect, one could use instead the following objective function:

$$F = \sum_{i=1}^{m} G(-y_i f(x_i)), \tag{1}$$

where G is the function defined on $\mathbb{R}$ by

$$G(x) = \begin{cases} e^x & \text{if } x \leq 0 \\ x + 1 & \text{otherwise.} \end{cases} \tag{2}$$

(a) Show that the function G is convex and differentiable.

(b) Use F and greedy coordinate descent to derive an algorithm similar to Adaboost.

(c) Can you compare the reduction of the empirical error rate of this algorithm with that of Adaboost?