Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 3
Due: May 08, 2009

**A. Perceptron algorithm**

Let $S$ be a labeled sample of $N$ points in $\mathbb{R}^N$ with

$$x_i = (\underbrace{(-1)^i, \ldots, (-1)^i, (-1)^{i+1}}_{i \text{ first components}}, 0, \ldots, 0) \quad \text{and} \quad y_i = (-1)^{i+1}. \tag{1}$$

- Show that the perceptron algorithm makes $\Omega(2^N)$ updates before finding a separating hyperplane, regardless of the order in which it receives the points.

**B. Boosting**

This problem considers an algorithm similar to AdaBoost but with a different objective function. Assume that the training data is given as $m$ labeled examples $(x_1, y_1), \ldots, (x_m, y_m) \in X \times \{-1, +1\}$. Let $\Phi \colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$\Phi(u) = \begin{cases} (1+u)^2 & \text{if } u \geq -1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- Consider the objective function $F$ defined by $F(\alpha) = \sum_{i=1}^m \Phi(-y_i f(x_i))$ where $f$ is a linear combination of base classifiers: $f = \sum_{t=1}^T \alpha_t h_t$ as for AdaBoost. Show that $F$ is convex and differentiable.

- Derive a new boosting algorithm using the objective function $F$. Characterize the best base classifier $h_u$ to select at each round of boosting if we use coordinate descent.