**MCS 548 – Mathematical Theory of Artificial Intelligence**
**Fall 2016**
**Problem Set 3**

Lev Reyzin

**Due:** 11/23/16 at the beginning of class

**Instructions:** Atop your problem set, please write your name and list your collaborators.

**Problems**

1. We proved a margin bound (Theorem 7.8 of Mohri et al.) on the number of mistakes for the Perceptron algorithm for the update rule \( w_{t+1} \leftarrow w_t + y_t x_t \). Consider the general update rule \( w_{t+1} \leftarrow w_t + \eta y_t x_t \), where \( \eta > 0 \). Prove a bound on the maximum number of mistakes for this rule. How does \( \eta \) affect the bound?

2. Imagine that a bandit or an online learning algorithm \( A \) that runs in \( T \) rounds and has an expected regret bound of \( \epsilon + T/\epsilon \), where \( \epsilon \) is set by the algorithm. Clearly the optimal setting is \( \epsilon = \sqrt{T} \). The problem is that sometimes \( T \) is not known in advance. How do we fix this issue? We can have an algorithm \( A' \) that does the following: \( A' \) starts with a parameter \( \epsilon_1 \) and runs \( A \) for \( T_1 \) rounds, then adjusts the parameter to \( \epsilon_2 \) and runs \( A \) for \( T_2 \) rounds, and so on. Construct a schedule of \((\epsilon_i, T_i)\) that asymptotically achieves the \( \sqrt{T} \) expected regret bound without knowing \( T \) in advance.

3. Suppose you have two coins, one perfectly fair, and one with bias toward H of \( 1/2 + \epsilon \) for some \( \epsilon > 0 \). It is known that to tell which coin is biased (with probability \( > 3/4 \)) one needs to perform at least \( c/\epsilon^2 \) coin flips (\( c > 0 \) is some constant). Show that this implies that EXP3’s asymptotic regret dependence of \( T^{1/2} \) cannot be improved to \( T^{1/2-\delta} \) for any constant \( \delta > 0 \).

4. In unregularized least squares regression, we solve \( W = (XX^T)^{-1}XY \), where \( x_i \in \mathcal{R}^N \) (for \( 1 \leq i \leq m \)) and

\[
X = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}, \quad W = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix},
\]

when \( XX^T \) is invertible. What conditions are required on \( x_1, \ldots, x_m \) (and thereby \( X \)) for \( XX^T \) to be invertible?