CPSC 531H Machine Learning Theory (Term 2, 2013-14)
Assignment 4

Due: Thursday April 10th.

Question 1: Consider the problem of learning with expert advice when one of the experts gives perfect predictions. On some round t, let q be the fraction of surviving experts that predict 1. (A surviving expert is one that has not made any mistakes so far.) In class, we talked about the halving algorithm which predicts with the majority vote of the expert predictions.

In general, we can predict 1 with probability $F(q)$ and 0 with probability $1 - F(q)$ for some function F. For instance, for the halving algorithm, $F(q)$ is 1 if $q > 1/2$ and 0 if $q < 1/2$ (and arbitrary if $q = 1/2$).

Consider now a function $F : [0, 1] \to [0, 1]$ satisfying the following property:

$$1 + \frac{\lg q}{2} \le F(q) \le -\frac{\lg(1-q)}{2}. \tag{1}$$

- **(a):** Suppose we run an on-line learning algorithm that uses a function F satisfying (1) as described above. Show that the expected number of mistakes made by the learning algorithm is at most $(\lg N)/2$, where N is the number of experts.
- **(b):** Show that the function

  $$F(q) = \frac{\lg(1-q)}{\lg q + \lg(1-q)}$$

  has range $[0, 1]$ and satisfies (1). (At the endpoints, we define $F(0) = 0$ and $F(1) = 1$ to make F continuous, but you don't need to worry about these.)

Question 2: Bad Shooting.

- **(a):** Prove that for any point $p_t \in \mathbb{R}^d$ there exists a point $y_t \in \mathbb{R}^d$ such that $\|y_t\| \le 1$ and $\|p_t - y_t\| \ge 1$.
- **(b):** Conclude that for any deterministic prediction strategy for the Shooting Game there exists a sequence $y_1, y_2, \ldots, y_n$ such that the total loss $\sum_{t=1}^{n} \|p_t - y_t\|^2$ after n rounds is at least n.

Question 3: Lower Bounds for Follow The Leader.

Consider the Follow The Leader Algorithm for the Shooting Game.

- **(a):** Prove that for any sequence $y_1, y_2, \ldots, y_n$ and any number of rounds n the regret of the algorithm is non-negative.
- **(b):** According our theorem of lecture 18, the regret of FTL on any sequence $y_1, y_2, \ldots, y_n$ in the unit ball of R^d is $O(\log n)$. Show that this bound is tight. More precisely, for any $n \ge 1$, construct a sequence $y_1, y_2, \ldots, y_n$ in the unit ball such that FTL has $\Omega(\log n)$ regret on this sequence.

  Hint: Consider the sequence $y_t = (-1)^t v$ where v is an arbitrary unit vector. First solve the case when n is even. Harmonic sums may be useful.

Question 4: OPTIONAL BONUS PROBLEM: Analysis of PSim.

In this question, we will complete the proof of the analysis of the PSim algorithm. You must prove the following lemma.

Lemma 1. *For all $x \in [n]$,*

$$\frac{1}{T} \cdot \mathrm{E}\left[\sum_{t=1}^{T} (c_t(x_t) - c_t(x))\right] \;\leq\; \epsilon + \frac{2\log n}{\epsilon P} + \frac{nP}{T}.$$

Hint: In Week 8 of Kleinberg's lecture notes, the purpose of Lemma 3 and 4 is to establish this lemma. However, I think there is a small bug in the first line of the proof of Lemma 3. I think it is possible that $\mathrm{E}\,[\,c_t(\hat{x}_j) - c_t(x)\,] < 0$ for some $x \in$ Explore, in which case that first inequality might not be correct. Nevertheless, a fairly direct argument can prove our desired lemma, by appealing to the analysis of Randomized Weighted Majority, and an additional error term.