Advanced ML and AGT **Fall Semester, 2011/12**

Homework 2: Dec 12, 2011

Lecturer: Yishay Mansour

Homework number 2.

Question I: Consider a multi-arm bandit setting where the actions continuous in $[0, 1]$. At each time we have some unknown cost function $c_t(x)$. The only assumption is that $c_t(x) \in [0, 1]$ and that $c_t(x)$ is differential and its derivative is bounded by L. Show an online multi-arm bandit algorithm for this setting and bound its regret.

Question II: Derive a lower bound of $\Omega(1/\epsilon)$, using KL divergence, on the sample size for the case of realizable target function. (Namely, look at the case that the best function has zero error rate, and we like to select a function with error more than ϵ with probability at most δ.)

Question III: Consider the EXP3 algorithm when run on two actions ($K = 2$). Action 0 has always a reward of 0, while action 1 has always a reward of 1. Let the *sampling regret* at time t be $\frac{w_0(t)}{w_0(t)+w_1(t)}$. Upper bound the expected sampling regret in T steps, i.e., $E[\sum_{t=1}^{T} \frac{w_0(t)}{w_0(t)+w_1(t)}]$.

Question IV: Consider an adaptive adversary in the online gradient decent algorithm. Let $c_t(x)$ be a convex function over set S, and let $z_{t+1} = \Pi_S(z_t - \eta g_t)$, where $\eta > 0$ and $\Pi_S(\cdot)$ is the projection operator. Let $g_1, \ldots, g_T$ are random variables such that: (a) $E[g_t|z_1, \ldots z_t, c_1, \ldots c_t] = \bigtriangledown c_t(z_t)$, (b) $\|g_t\| \leq G$ and $\| \bigtriangledown c_t(x)\| \leq G$, and (c) $S \subset R\mathbb{B}$, where $\mathbb{B}$ is the unit ball and $R > 0$. . Let $h_t(x) = c_t(x) + x \cdot \rho_t$, where $\rho_t = g_t - \bigtriangledown c_t(z_t)$.

1. Show that $\|\rho_t\| \leq 2G$ and $E[\rho_t \cdot \rho_s] = 0$.
2. Show that for any $x \in S$ the following three inequalities hold:

$$E[| \sum_{t=1}^{T} h_t(x) - c_t(x)|] \leq_{(i)} R \cdot E[\| \sum_{t=1}^{T} \rho_t \|] \leq_{(ii)} R \sqrt{E[\| \sum_{t=1}^{T} \rho_t \|^2]} \leq_{(iii)} 2G\sqrt{T}$$

3. Assume that you already showed that $\sum_{t=1}^{T} h_t(z_t) - \min_{x\in S} \sum_{t=1}^{T} h_t(x) \leq RG\sqrt{T}$. Show an $O(RG\sqrt{T})$ regret bound for the **adaptive adversary**, under the above assumptions.

The homework is due in two weeks

Additional (fun) reading

Improved MAB: An $O(\sqrt{TN})$ regret bound is found in "Minimax policies for adversarial and stochastic bandits" by Audibert and Bubeck (COLT 2009)

Improved Online Linear Optimization in MAB: Improving the regret bound dependence on T to $O(\sqrt{T})$ is found in "Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization" by Abernethy, Hazan and Rakhlin (COLT 2008)