Advanced topics in ML and AGT **Fall Semester, 2017/18**

Homework 1: Nov 6,2017

Lecturer: Yishay Mansour

Homework number 1.

Switching hypothesis:
Given a class of hypothesis $H = \{h : X \to \{0,1\}\}$ we call a sequence of T pairs (x_i, y_i), such that $x_i \in X$ and $y_i \in \{0,1\}$, k-realizable, if there are k hypotheses $h_1, \ldots, h_k \in H$, and times $t_0 = 0 \leq t_1 \leq \ldots, t_{k-1} \leq t_k = T$, such that $h_j(x_i) = y_i$ for $i \in (t_{j-1}, t_j]$. Show an algorithm that makes at most $O(k \log |H|)$ mistakes on sequences which are k-realizable.

Be the Leader
Consider what would be the regret if the online algorithm updates: $w_t = \arg\min_{w \in S} \sum_{i=1}^{t} f_i(w)$.
(Note that the sum is until t and not $t-1$!)
Why would this algorithm be impossible to implement?

Online gradient descent:
~~Consider online gradient descent with a changing regularization. Specifically, at time t the update is $w_{t+1} = \arg\min_{w \in S} R_0(w) + \sum_{i=1}^{t-1} f_i(w) + R_i(w)$, where $R_i(w) \geq 0$ for any $w \in S$.~~

This problem has an error, please ignore it!

Entropic Regularization and RWM
Consider the experts experts problem with entropic regularization. Specifically, $R(w) = \frac{1}{\eta} \sum w_i \ln w_i$ for w which is a distribution and ∞ otherwise. The losses at time t is a vector $\ell_t \in [0,1]^d$. Let $L_T[i] = \sum_{t=1}^{T} \ell_t[i]$. Show that the minimizing vector $w_{T+1}[i] = e^{-\eta L_T[i]} / (\sum_{j=1}^{d} e^{-\eta L_T[j]})$.
(Solve the optimization problem: $\min_w \sum_{t=1}^{T} w^\top \ell_t + R(w)$ such that $\|w\|_1 = 1$ and $w > 0$.)

Convexity

1. **Jensen inequality:** Show that if f is a convex function, then $\mathbb{E}[f(x)] \geq f[\mathbb{E}(x)]$. (It is enough to show it for a finite support.)

2. **Algebraic vs. Geometric mean:** Given n values $x_i \geq 0$, show that $(\prod_{i=1}^{n} x_i)^{1/n} \leq \frac{1}{n} \sum_{i=1}^{n} x_i$. (Hint: show first that $g(x) = -\log(x)$ is a convex function.)

3. **Holder inequality:** Show that for $x, y \in \mathbb{R}^n$ we have $x^\top y \leq \|x\|_p \|y\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$. (Hint: First show that $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$.)

The homework is due in two weeks