Advanced topics in ML and AGT **Fall Semester, 2018/19**

Homework 1: Oct 29, 2018

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?

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 $a^\theta b^{1-\theta} \leq \theta a + (1 - \theta) b$.)

Changing regularizers :

1. Consider FoReL with changing regularizers, meaning that 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 $w \in S$. Express an OGD with adaptive learning rates η_t using FoReL with changing regularizers. (Recall that in OGD we have $w_{t+1} = w_t - \eta_t \nabla f_t(w_t)$. You can consider a modified

OGD where $w_{t+1} = -\eta_t \sum_{i=1}^{t} \nabla f_i(w_i)$, which should significantly simplify. You need to find a sequence of regularizers R_t, that depends on η_t and w_t, such that the OGD actions are identical to the actions of FoReL with the sequence of regularizers. You can assume that the f_ts have been linearized, i.e., $f_t(x) = g_t^\top x$.)

2. Assuming that each R_t is σ_t-strongly-convex, how that $h_{0:t}(w) = \sum_{i=1}^{t-1} f_i(w) + \sum_{i=0}^{t} R_i(w)$ is 1-strongly-convex w.r.t. the norm $\|\cdot\|_{(t)} = \sqrt{\sum_{i=0}^{t} \sigma_i}\, \|\cdot\|_2$.

3. For $\|\cdot\|_{(t)} = \sqrt{\sum_{i=0}^{t} \sigma_i}\, \|\cdot\|_2$ show that $\|\cdot\|_{(t),\star} = \left(\sqrt{\sum_{i=0}^{t} \sigma_i}\right)^{-1} \|\cdot\|_2$

4. Bound the regret of adaptive OGD. You may use the following theorem:

   Theorem 0.1 *Suppose $R_t(x) \geq 0$ are chosen such that $\sum_{i=0}^{t} R_i + \sum_{i=1}^{t-1} f_i$ is 1-strongly-convex w.r.t. some norm $\|\cdot\|_{(t)}$. Then for any x^* we have*

$$Regret(x^*) \leq \sum_{t=0}^{T-1} R_t(x^*) + \frac{1}{2} \sum_{t=1}^{T} \|g_t\|^2_{(t-1),\star}$$

5. For $\|x^*\| \leq L$ and $\|g_t\| \leq G$, give adaptive learning rates η_t such that the regret is $O(LG\sqrt{T})$. You may use the inequality $\sum_{i=1}^{n} 1/\sqrt{i} \leq 2\sqrt{n}$.

The homework is due in two weeks