Theoretical Foundations of Machine Learning - Homework #5

Jacob Abernethy and Chansoo Lee Due: 12/10/2015

Homework Policy: Working in groups is fine, but *every student* must submit their own writeup. Please write the members of your group on your solutions. There is no strict limit to the size of the group but we may find it a bit suspicious if there are more than 4 to a team. Questions labelled with **(Challenge)** are not strictly required, but you'll get some participation credit if you have something interesting to add, even if it's only a partial answer.

1) **Generalized Minimax Theorem.** Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be convex compact sets. Let $f : X \times Y \to \mathbb{R}$ be some differentiable function with bounded gradients, where $f(\cdot, \mathbf{y})$ is convex in its first argument for all fixed $\mathbf{y}$, and $f(\mathbf{x}, \cdot)$ is concave in its second argument for all fixed $\mathbf{x}$.

Prove that

$$\inf_{\mathbf{x}\in X} \sup_{\mathbf{y}\in Y} f(\mathbf{x}, \mathbf{y}) = \sup_{\mathbf{y}\in Y} \inf_{\mathbf{x}\in X} f(\mathbf{x}, \mathbf{y}).$$

Furthermore, give an efficient algorithm for finding an ϵ-optimal pair $(\mathbf{x}^*, \mathbf{y}^*)$ for any parameter $\epsilon > 0$.

2) **Regret of Follow the Perturbed Leader.** We will observe a sequence of loss vectors $\boldsymbol{\ell}^1, \boldsymbol{\ell}^2, \ldots, \boldsymbol{\ell}^T \in [0,1]^n$. We need an algorithm for picking a sequence of distributions $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^T \in \Delta_n$ with the goal of minimizing regret. For the rest of this problem we shall define regret relative to some $\mathbf{p}$ as

$$\text{Regret}_T(\text{Alg}; \mathbf{p}) := \sum_{t=1}^{T} (\mathbf{p}^t \cdot \boldsymbol{\ell}^t - \mathbf{p} \cdot \boldsymbol{\ell}^t).$$

Note that this differs slightly than our usual notion where $\mathbf{p}$ is chosen to be the best distribution (or expert) in hindsight.

I have already mentioned an algorithm often called Follow The Leader (FTL) defined as

$$\text{FTL} \quad := \quad \mathbf{p}^t \leftarrow \arg\min_{\mathbf{p}\in\Delta_n} \mathbf{p} \cdot \left(\sum_{s=1}^{t-1} \boldsymbol{\ell}^s \right)$$

There's an easy lower bound that shows that this algorithm can achieve $\Theta(T)$ regret which is bad! But what if we just perturb this algorithm slightly? Here's an alternative approach which involves playing FTL on the cumulative loss vector with some added noise.

$$\text{FTPL} \quad := \quad X \overset{\text{u.a.r.}}{\sim} [0, b]^n; \quad \text{then } \forall t \quad \mathbf{p}^t \leftarrow \arg\min_{\mathbf{p}\in\Delta_n} \mathbf{p} \cdot \left(X + \sum_{s=1}^{t-1} \boldsymbol{\ell}^s \right)$$

Note that the perturbation X is only sampled *once* in this algorithm. X is sampled uniformly at random from a cube, and note that the sidelength of the cube $b > 0$ is a parameter which we can tune.

For analysis purposes, it is convenient to define two *fictitious* algorithms, known as Be The Leader (BTL), and Be The Perturbed Leader (BTPL).

$$\begin{aligned}
\text{BTL} &:= \quad \mathbf{p}^t \leftarrow \arg\min_{\mathbf{p}\in\Delta_n} \mathbf{p}\cdot\left(\sum_{s=1}^{t}\boldsymbol{\ell}^s\right) \\
\text{BTPL} &:= \quad X \overset{\text{u.a.r.}}{\sim} [0,b]^n; \quad \text{then } \forall t \quad \mathbf{p}^t \leftarrow \arg\min_{\mathbf{p}\in\Delta_n} \mathbf{p}\cdot\left(X+\sum_{s=1}^{t}\boldsymbol{\ell}^s\right).
\end{aligned}$$

What is different here? Notice I changed the sum to end at $s = t$ rather than $s = t - 1$ – that's why these algorithms are fictitious, they get to see one datapoint in the future! Here we are "being" the leader rather than "following" the leader because we actually can compute the leader up to *and including* the loss vector that will arrive today.

(a) BTL, while not a realistic algorithm, kicks ass! Prove, for any $\mathbf{p} \in \Delta_n$, that

$$\text{Regret}_T(\text{BTL}; \mathbf{p}) \leq 0$$

Hint: Induction.

(b) BTPL is really not that much worse than BTL. Prove that

$$\text{Regret}_T(\text{BTPL}; \mathbf{p}) \leq b.$$

Note that this is a deterministic statement, doesn't depend on the sample of X.

(c) Assume that BTPL and FTPL were run using the same perturbation X sampled before round 1. Let $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^T$ be the distributions played by FTPL throughout the sequence. Prove that for any perturbation X,

$$\text{Regret}_T(\text{FTPL}; \mathbf{p}) = \text{Regret}_T(\text{BTPL}; \mathbf{p}) + \sum_{t=1}^{T}(\mathbf{p}^t - \mathbf{p}^{t+1}) \cdot \boldsymbol{\ell}^t.$$

Again this is for a fixed, arbitrary X.

(d) It turns out that by perturbing the loss by X, we are much less likely to switch from round to round. If $\mathbf{p}^t, \mathbf{p}^{t+1}$ are the distributions played by FTPL on rounds t and $t+1$ (respectively), then show that for any t,

$$\mathop{\mathbb{E}}_{X \overset{\text{u.a.r.}}{\sim} [0,b]^n}[(\mathbf{p}^t - \mathbf{p}^{t+1}) \cdot \boldsymbol{\ell}^t] \leq \frac{n}{b}.$$

Hint: Define the random variables $Z_t := X + \sum_{s=1}^{t-1} \boldsymbol{\ell}^s$ and $Z_{t+1} := X + \sum_{s=1}^{t} \boldsymbol{\ell}^s$. Notice that the distributions of Z_t and Z_{t+1} overlap significantly.

(e) Let's put it all together! Prove that for a particular choice of b we can achieve:

$$\mathop{\mathbb{E}}_{X \overset{\text{u.a.r.}}{\sim} [0,b]^n}[\text{Regret}(\text{FTPL}; \mathbf{p})] \leq \sqrt{nT}.$$

It's ok if you didn't solve all of the above, you may use the conclusions from each subproblem.

(f) **(Challenge)** It is important for the analysis that X is sampled once and fixed throughout the sequence. But in terms of expected regret, would it matter if we sampled X separately for each round? Why or why not?

(g) (**Challenge**) It's too bad the above bound isn't as tight as the $O(\sqrt{T \log n})$ bound we can get with EWA. Can FTPL be improved using a better choice of perturbation X? I might suggest a Laplace distribution or a Gaussian.

(h) (**Challenge**) Is there a way to implement EWA (exponential weights algorithm) in the action setting (i.e. the "hedge" setting) using FTPL? In other words, can you choose a perturbation random variable X such that the maximizing action on round t is chosen with the same probabilities as the EWA distribution?

(i) (**Challenge**) Show that any FTPL can be formulated as FTRL, although the regularizer for FTRL may not be efficiently computable. *Hint*: Given FTPL with distribution $\mathcal{D}$, consider the Fenchel conjugate of the function

$$g(\mathbf{L}) := \mathbb{E}_{X \sim \mathcal{D}} \left[\max_{\mathbf{p} \in \Delta_n} -\mathbf{p} \cdot (X + \mathbf{L}) \right].$$

3) **Online Non-Convex Optimization.** Sometimes our nice assumptions don't always hold. AWWW SHUCKS!! But maybe things will still work out just fine. For the rest of this problem assume that $X \subset \mathbb{R}^n$ is the learner's decision set, and the learner observes a sequence of functions $f_1, f_2, \ldots, f_T$ mapping $X \to \mathbb{R}$. The regret of an algorithm choosing a sequence of $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is defined in the usual way:

$$\text{Regret}_T := \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in X} \sum_{t=1}^{T} f_t(\mathbf{x})$$

Wouldn't it ruin your lovely day if the functions f_t were not convex? Maybe the only two conditions you can guarantee is that the functions f_t are bounded (say in $[0,1]$) and are 1-*Lipschitz*: they satisfy that $|f_t(\mathbf{x}) - f_t(\mathbf{x}')| \leq \|\mathbf{x} - \mathbf{x}'\|_2$. Prove that, assuming X is convex and bounded, there exists a randomized algorithm with a reasonable expected-regret bound. Something like $\mathbb{E}[\text{Regret}_T] \leq \sqrt{nT \log T}$ would be admirable. (Hint: Always good to ask the *experts* for ideas. And you needn't worry about efficiency.)

4) **Information-theoretic Lower Bound for EXP3.** Someone hands you N coins and tells you that all BUT ONE are symmetrically weighted: each will come up heads exactly half the time. But there's one odd coin $I \in [N]$ that lands heads with probability $\frac{1}{2} + \epsilon$ for some $\epsilon > 0$. Your task is to sequentially select coins to flip, in any manner you choose, and then to make a guess $\hat{I} \in [N]$ which is the odd coin.

CLAIM(*): If you perform fewer than $\frac{cN}{\epsilon^2}$ coin flips, where $c > 0$ is some constant, then $\Pr(\hat{I} = I) \leq 3/4$. In other words, no algorithm will have a very good chance to guess the odd coin with so few coin tosses.

It turns out that the above claim is true, but we will not concern ourselves with proving it here. Instead we will use it to construct a lower bound.

Recall that the EXP3 algorithm has a regret bound on the order of $\sqrt{TN \log N}$ where N is the number of arms and T is the length of the sequence. Prove that if there exists an algorithm with a slightly better regret bound, say $O(T^{\frac{1}{2}} N^{\frac{1}{2} - \delta})$ for some $\delta > 0$, then we would violate **CLAIM(*)**.

Remark: The problem shows precisely that the $\sqrt{N}$ dependence is fundamental to the bandits problem. Plus, this lower bound is *information-theoretic* as opposed to *computational*. That is, any algorithm, whether polynomial or exponential time, would lead to the contradiction.

Partial Credit: Show that any algorithm with regret bound $O((TN)^{\frac{1}{2} - \delta})$ would violate **CLAIM(*)**.