Prediction and Learning: It's Only a Game - Homework #2

Jacob Abernethy Due: 10/28/2013

Homework Policy: Working in groups is fine. 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) **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

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

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

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

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.

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

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).

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

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

What is different here? Notice I changed the sum to end at $s = t$ rather than $s = t - 1$ – that's why they are fictitious, they get to see one datapoint in the future! In these algorithms 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) Let $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^T$ be the distributions played by FTPL throughout the sequence. Assume that BTPL and FTPL were run using the same perturbation X. Prove that

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

Again this is for fixed (nonrandom) 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) 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. I know a Laplace will work, but I have reason to believe that the Gaussian is actually roughly the minimax optimal choice, although I've not seen a bound proven for this case.

(h) (**Challenge**) Is there a way to implement EWA in the action 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?

2) **Solving Feasibility Problems with Perceptron.** We can define a special class of LP feasibility (LFP) problems as follows. Given a set of m vectors $\mathbf{a}^1, \ldots, \mathbf{a}^m \in \mathbb{R}^d$, we want to determine if there exists an $\mathbf{x} \in \mathbb{R}^d$ such that $\mathbf{a}^i \cdot \mathbf{x} > 0$ for $i = 1, \ldots, m$. For any $\mathbf{x}$ we can define $\nu(\mathbf{x}) := \min_i \frac{\mathbf{a}^i \cdot \mathbf{x}}{\|\mathbf{a}^i\|_2 \|\mathbf{x}\|_2}$ which is positive if $\mathbf{x}$ is feasible. Define the *wiggle room* of the feasibility problems as $\nu^* := \sup_{\mathbf{x} \in \mathbb{R}^d} \nu(\mathbf{x})$.

- **(a)** Assume we are given an LFP defined by $\mathbf{a}^1, \ldots, \mathbf{a}^m \in \mathbb{R}^d$ which we know to be feasible with wiggle room $\nu^* > 0$. Using Perceptron as a subroutine, give an algorithm that finds a feasible solution requiring no more than $\left(\frac{1}{\nu^*}\right)^2$ updates to Perceptron. Prove that your solution works.

- **(b)** Implement your solution and find a solution to a constraint problem I generated! You will find a CSV file on the course website `constraints.csv` which defines a 10-dimensional LFP problem with 439 constraints. Each row of the file represents a vector $\mathbf{a}^i = (a^i_1, \ldots, a^i_{10})$. Please provide your final solution *normalized*, i.e. give a vector $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$. Include a printout of your code.

3) **Computing Equilibria via Learning.** Let's find an ϵ-Nash solution to the following zero sum game:

$$M = \begin{pmatrix} 5 & 8 & 3 & 1 & 6 \\ 4 & 2 & 6 & 3 & 5 \\ 2 & 4 & 6 & 4 & 1 \\ 1 & 3 & 2 & 5 & 3 \end{pmatrix}.$$

Recall that a pair $(\mathbf{p}, \mathbf{q}) \in \Delta_4 \times \Delta_5$ is an ϵ-Nash if

$$\max_{j \in \{1,2,3,4,5\}} \mathbf{p}^\top M \mathbf{e}_j - \min_{i \in \{1,2,3,4\}} \mathbf{e}_i^\top M \mathbf{q} \le \epsilon.$$

We will use the term *optimality gap* to refer to the quantity on the left hand side of the above expression.

(a) Find an ϵ-Nash solution to the above game by implementing, in code, the following methods:

(1) Have both players "learn" from the other using EWA. (IMPORTANT: How do we actually get an ϵ-Nash using a learning strategy?)

(2) Have player 1 use EWA to compute $\mathbf{p}^t$ but let player 2 select a best response to $\mathbf{p}^t$; i.e. $\mathbf{q}^t := \arg\max_{\mathbf{q}} (\mathbf{p}^t)^\top M \mathbf{q}$.

(3) Have both players use FTPL. Have both players sample their perturbation independently on each round.

(4) Have both players use FTL, i.e. no perturbation. This is known as *fictitious play*.

For each of the above, run the algorithms for at least $T > 10,000$ rounds, and report the ϵ-Nash pair you obtain. Make a single plot showing the optimality gap as a function of T for all methods. Include printout of your code.

(b) (**Challenge**) We don't have a regret bound for FTL, so we can not use existing techniques to show an optimality gap that shrinks at a rate of $O(T^{-1/2})$. But empirically fictitious play appears to converge at such a rate. Can you prove why this is? (Note: this is the Karlin Conjecture, and it has been open for a long time. So good luck :-))