EECS598: Prediction and Learning: It's Only a Game Fall 2013

Lecture 20: EXP3 Algorithm

Prof. Jacob Abernethy *Scribe: Zhihao Chen*

Announcements

- None

20.1 EWA with ϵ-exploration (recap)

This algorithm generates an unbiased estimator for $\underline{l}^t$.

Strategy With probability $1-\epsilon$

- Choose $\underline{p}^t = EWA(\underline{\tilde{l}}^1, \ldots, \underline{\tilde{l}}^{t-1})$ and let $\underline{\tilde{l}}^t = \vec{0}$

With probability ϵ

- Choose $\underline{p}^t = \langle \frac{1}{n}, \ldots, \frac{1}{n} \rangle$, sample $I_t \sim \underline{p}^t$, and let $\underline{\tilde{l}}^t = \langle 0, \ldots, 0, \frac{n}{\epsilon} l^t_{I_t}, 0 \ldots, 0 \rangle = \frac{n}{\epsilon} l^t_{I_t} \underline{e}_{I_t}$ ($\underline{e}_{I_t}$ is the I_tth unit vector).

Proof that this algorithm works

$$\mathbb{E}[\underline{\tilde{l}}^t] = (1-\epsilon)\vec{0} + \epsilon \left(\sum_{i=1}^{n} p^t_i \left(\frac{n}{\epsilon} l^t_i \underline{e}_i \right) \right) = \sum_{i=1}^{n} l^t_i \underline{e}_i = \underline{l}^t$$

Naive expected regret bound

$$O\left(\epsilon T + \frac{n^2 T \eta}{\epsilon^2} + \frac{\log n}{\eta} \right) = O\left(T^{\frac{3}{4}} \sqrt{n} (\log n)^{\frac{1}{4}} \right) \quad \text{(with tuning)}$$

Better expected regret bound

$$O\left(\epsilon T + \frac{nT\eta}{\epsilon} + \frac{\log n}{\eta} \right) = O\left(T^{\frac{2}{3}} \sqrt{n} \right) \quad \text{(with tuning)}$$

Question from last lecture: Is there some algorithm that is better than EWA with ϵ-exploration? In particular, is it possible to reduce the power on T to $\frac{1}{2}$ in the expected regret?

20.2 EXP3

We claim that the EXP3 algorithm is a better algorithm than EWA with ϵ-exploration, and, in fact, has an expected regret bound of $\sqrt{2Tn\log n}$. Let us begin by stating the algorithm.

20.2.1 Algorithm

Let $\underline{\tilde{L}}^t$ be the cumulative losses up to period t.

for t=1,...,T **do**
  Sample $I_t \sim \underline{p}^t$
  Observe $l_{I_t}^t$
  Set $\underline{\tilde{l}}^t = \left\langle 0, ..., 0, \frac{l_{I_t}^t}{p_{I_t}^t}, 0..., 0 \right\rangle$
  Set $\underline{\tilde{L}}^t = \underline{\tilde{L}}^{t-1} + \underline{\tilde{l}}^t$
  for i=1,...,n **do**
    Set $p_i^{t+1} = \dfrac{e^{-\eta \tilde{L}_i^t}}{\sum_{j=1}^{n} e^{-\eta \tilde{L}_i^t}}$
  end for
end for

20.2.2 Comments on EXP3

The EXP3 algorithm looks very similar to that of EWA with ϵ-exploration. Indeed in both cases, the chosen loss vectors are divided by the probability of obtaining that vector. The key difference between the algorithms is that EXP3 does not drop observations in any round (as opposed to EWA with ϵ-exploration dropping observations with a probability of $1 - \epsilon$).

Intuitively, it seems that EXP3 might be a pretty bad algorithm, given that p_i^t's could get exponentially small, meaning that we could be dividing by a very small number in the algorithm. However, this works out in the end, as we will see in the analysis of the expected regret.

20.2.3 Analysis of the expected regret for EXP3

We analyze the regret of EXP3 by looking at the potential function

$$\Phi_t = -\frac{1}{\eta} \log\left(\sum_{i=1}^{n} e^{-\eta \tilde{L}_i^{t-1}} \right)$$

and taking the *expected* increase in potential in every period.

The increase in potential from period t to $t+1$ is

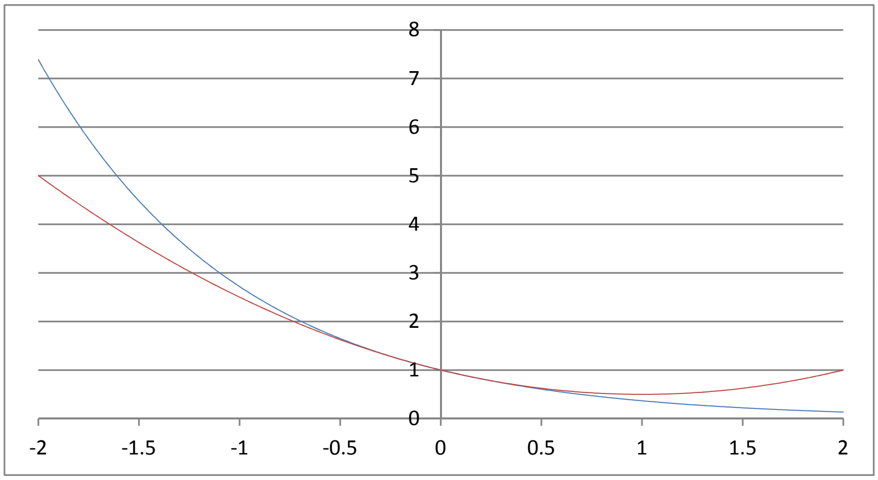
$$\Phi_{t+1} - \Phi_t = -\frac{1}{\eta} \log\left(\frac{\sum_{i=1}^n e^{-\eta \tilde{L}_i^t}}{\sum_{i=1}^n e^{-\eta \tilde{L}_i^{t-1}}} \right) = -\frac{1}{\eta} \log\left(\frac{\sum_{i=1}^n e^{-\eta \tilde{L}_i^{t-1} - \eta \tilde{l}_i^t}}{\sum_{i=1}^n e^{-\eta \tilde{L}_i^{t-1}}} \right) = -\frac{1}{\eta} \log\left(\mathbb{E}_{i \sim p^t} \left[e^{-\eta \tilde{l}_i^t} \right] \right)$$

To proceed, we need the following lemma.

Lemma 20.1. *For all $x \geq 0$,*

$$e^{-x} \leq 1 - x + \frac{1}{2}x^2$$

You can see this by plotting the two graphs e^{-x} and $1 - x + \frac{1}{2}x^2$. The blue line is e^{-x} and the red line is $1 - x + \frac{1}{2}x^2$ in the plot below.



Using the lemma, we get

$$\begin{aligned}
\Phi_{t+1} - \Phi_t &\geq -\frac{1}{\eta}\log\left(\mathbb{E}_{i\sim p^t}\left[1 - \eta\tilde{l}_i^t + \frac{1}{2}\eta^2(\tilde{l}_i^t)^2\right]\right) \\
&= -\frac{1}{\eta}\log\left(1 - \mathbb{E}_{i\sim p^t}\left[\eta\tilde{l}_i^t + \frac{1}{2}\eta^2(\tilde{l}_i^t)^2\right]\right) \\
&\geq \frac{1}{\eta}\mathbb{E}_{i\sim p^t}\left[\eta\tilde{l}_i^t + \frac{1}{2}\eta^2(\tilde{l}_i^t)^2\right] \qquad \text{(because } log(1-x) \leq -x\text{)} \\
&= \sum_{i=1}^{n} p_i^t\tilde{l}_i^t - \frac{\eta}{2}\sum_{i=1}^{n} p_i^t(\tilde{l}_i^t)^2
\end{aligned}$$

Taking the expectation on both sides,

$$\begin{aligned}
\mathbb{E}[\Phi_{t+1} - \Phi_t] &\geq \mathbb{E}\left[\sum_{i=1}^{n} p_i^t\tilde{l}_i^t - \frac{\eta}{2}\sum_{i=1}^{n} p_i^t(\tilde{l}_i^t)^2\right] \\
&= \sum_{i=1}^{n} p_i^t l_i^t - \frac{\eta}{2}\mathbb{E}\left[p_{I_t}^t\left(\frac{l_{I_t}^t}{p_{I_t}^t}\right)^2\right] \\
&= \underline{p}^t \cdot \underline{l}^t - \frac{\eta}{2}\mathbb{E}\left[\frac{(l_{I_t}^t)^2}{p_{I_t}^t}\right] \\
&= \underline{p}^t \cdot \underline{l}^t - \frac{\eta}{2}\sum_{i=1}^{n}(l_i^t)^2 \\
&\geq \underline{p}^t \cdot \underline{l}^t - \frac{\eta n}{2}
\end{aligned}$$

Now, we sum the differences in potential to get

$$\mathbb{E}[\Phi_{T+1} - \Phi_1] = \mathbb{E}\left[\sum_{t=1}^{T}(\Phi_{t+1} - \Phi_t)\right] \geq \sum_{t=1}^{T} \underline{p}^t \cdot \underline{l}^t - \frac{T\eta n}{2}$$

Furthermore,

$$\mathbb{E}[\Phi_{T+1} - \Phi_1] \leq \mathbb{E}\left[\tilde{L}_{i^*}^T - \left(-\frac{1}{\eta}\log n\right)\right] = L_{i^*}^T + \frac{1}{\eta}\log n$$

Combining the two inequalities, we get

$$\mathbb{E}\text{-regret}_T(EXP3) = \sum_{t=1}^{T} \underline{p}^t \cdot \underline{l}^t - L_{i^*}^T \leq \frac{1}{\eta}\log n + \frac{T\eta n}{2} \qquad -(*)$$

Theorem 20.2.

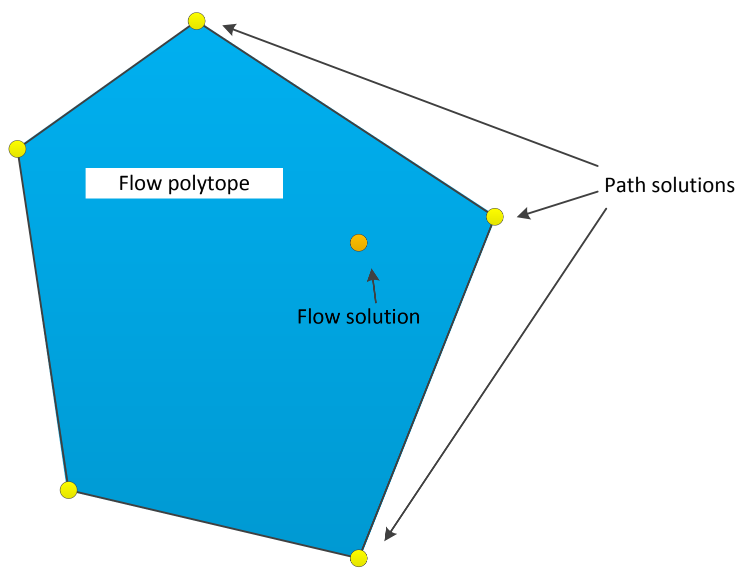
$$\mathbb{E}\text{-}regret_T(EXP3) \leq \sqrt{2Tn\log n}$$

Proof. Tune $\eta = \sqrt{\frac{2\log n}{Tn}}$ in $(*)$. □

20.3 Bandit problem in stochastic shortest path setting (in-class slide presentation)

Problem Want to find the shortest path from source to sink in a network with stochastic costs. Only flow costs on the selected path are known at the end of a period (bandit setting).

Flows ⇔ paths The number of paths is exponential, so we would want to work with flow solutions instead of arc solutions. It is easily seen that a path solution can be reduced to a flow solution.



Conversely, assuming that the decision maker can choose paths with some distribution, it can be shown that a flow solution is equivalent to some convex combination of path solutions.

FTRL in the bandit setting In every round,

$$x_t = \arg\min_{x \in \mathcal{K}} \sum_{s=1}^{t-1} f_s \cdot x + \lambda R(x)$$

Can we use estimated loss functions instead? i.e.

$$x_t = \arg\min_{x \in \mathcal{K}} \sum_{s=1}^{t-1} \tilde{f}_s \cdot x + \lambda R(x)$$

As it turns out, yes! Due of the convexity of regret in f_t's,

$$\begin{aligned} \mathbb{E}[\text{Regret}(\tilde{f}_1, \dots, \tilde{f}_T)] &\geq \text{Regret}(\mathbb{E}[\tilde{f}_1], \dots, \mathbb{E}[\tilde{f}_T]) \\ &= \text{Regret}(f_1, \dots, f_T) \end{aligned}$$

It is sufficient to compete with an unbiased estimate of the loss functions (instead of the actual loss functions).