CSE 522: Algorithms and Uncertainty Spring 2017

Lecture 12 — May 15, 2017

Lecturer: Nikhil R. Devanur

1 Bandit Linear Optimization

For this lecture, we assume we can pick points in S, rather than A. Given that in round t the algorithm picks $\mathbf{w}_t \in S$, the feedback that we observe in roung t is precisely the loss of this vector, $\ell_t \cdot \mathbf{w}_t$.

1.1 Estimating gradients

Suppose we could observe the loss at two points in one round, say $\mathbf{w}'_t$, and $\mathbf{w}''_t$. Then we could estimate the loss as

$$\tilde{\ell}_t(i) := \ell_t \cdot (\mathbf{w}'_t - \mathbf{w}''_t),$$

if

$$\mathbf{w}'_t - \mathbf{w}'_t = \mathbf{e}(i).$$

We also want $\mathbf{w}'_t$ and $\mathbf{w}''_t$ to be close to $\mathbf{w}_t$. Further, suppose we would like that

$$\frac{\mathbf{w}'_t + \mathbf{w}''_t}{2} = \mathbf{w}_t.$$

One easy way of ensuring the above conditions is to choose

$$\mathbf{w}'_t = \mathbf{w}_t + \mathbf{e}(i)/2, \quad \mathbf{w}''_t = \mathbf{w}_t - \mathbf{e}(i)/2.$$

But we are not allowed two point estimates; we are only allowed one. The cool thing is, randomization allows us to get the same effect with one point. Let

$$\sigma_t = \pm 1, \text{ with equal probability.}$$

Suppose that the point we pick is

$$\mathbf{w}_t + \sigma_t \mathbf{e}(i).$$

It is easy to see that in expectation, this is the same as $\mathbf{w}_t$, i.e.,

$$\mathbb{E}_{\sigma_t}[\mathbf{w}_t + \sigma_t \mathbf{e}(i)] = \mathbf{w}_t.$$

This is simply because $\mathbb{E}_{\sigma_t}[\sigma_t] = 0$. For defining the estimate we subtracted the two vectors. This is equivalent to the following.

$$\tilde{\ell}_t(i) := [\ell_t \cdot (\mathbf{w}_t + \sigma_t \mathbf{e}(i))]\sigma_t.$$

This is a valid estimator, since the feedback that we recieve is $[\ell_t \cdot (\mathbf{w}_t + \sigma_t \mathbf{e}(i))]$. Now note that

$$\mathbb{E}_{\sigma_t}[(\mathbf{w}_t + \sigma_t \mathbf{e}(i)) \cdot \sigma_t] = \mathbf{e}(i).$$

This is because, as before, $\mathbb{E}_{\sigma_t}[\sigma_t] = 0$, and $\mathbb{E}_{\sigma_t}[\sigma_t^2] = 1$. Therefore,

$$\mathbb{E}_{\sigma_t}[\tilde{\ell}_t(i)] = \ell_t(i).$$

We want to simultaneously estimate the loss of all the coordinates. For this, as we did before,

- we choose i_t from some probability distribution $\mathbf{p}_t$ over $[d]$.
- Pick $\sigma_t = \pm 1$ uniformly at random.
- Pick the vector $\mathbf{w}_t + \sigma_t \mathbf{e}(i_t)$.
- Observe the feedback $\ell_t \cdot (\mathbf{w}_t + \sigma_t \mathbf{e}(i_t))$.
- Define for all $i \in [d]$, $\tilde{\ell}_t(i) := \mathbf{1}(i_t = i)[\ell_t \cdot (\mathbf{w}_t + \sigma_t \mathbf{e}(i_t))]\sigma_t/\mathbf{p}_t(i)$.

Note that we still have that for all $i \in [d]$,

$$\mathbb{E}_{\sigma_t, i_t}[\tilde{\ell}_t(i)] = \ell_t(i).$$

One issue with this is that we need the vector that we pick, i.e., $\mathbf{w}_t + \sigma_t \mathbf{e}(i_t)$, to be in S. For this, we need one modification, which is to multiply the unit vectors by δ, for some small δ. In other words, pick the vector

$$\mathbf{w}_t + \delta \sigma_t \mathbf{e}(i_t),$$

and estimate

$$\tilde{\ell}_t(i) := \mathbf{1}(i_t = i)[\ell_t \cdot (\mathbf{w}_t + \delta \sigma_t \mathbf{e}(i_t))]\sigma_t/(\mathbf{p}(i)\delta).$$

We then restrict our initial search space to the following smaller set

$$S_\delta = \{\mathbf{u} : \mathbf{u} \pm \delta \mathbf{e}(i) \in S, \forall i \in [d]\}.$$

By this definition, if our initial $\mathbf{w}_t$ is in the set S_δ, then we are guaranteed that the final vector we pick is going to be in S. This introduces an additional error because we can now only compete against the best $\mathbf{u} \in S_\delta$. This error is

$$\min_{\mathbf{u} \in S} \{\ell_{1..T} \cdot \mathbf{u}\} - \min_{\mathbf{u} \in S_\delta} \{\ell_{1..T} \cdot \mathbf{u}\}. \tag{1}$$

Suppose that the set S is such that

$$\max_{\mathbf{u} \in S} \min_{\mathbf{v} \in S_\delta} \{\|\mathbf{u} - \mathbf{v}\|_1\} \leq O(\delta). \tag{2}$$

Then, given that the losses are such that $\|\ell_t\|_\infty \leq 1$ for all $t \in [T]$, the error in (1) is bounded above by

$$O(\delta T).$$

1.2 Combinatorial bandits

We now apply these ideas to combinatorial bandits, for the case that the set of actions is all subsets of size r. The convex hull S is the set of all non-negative vectors $\mathbf{u}$ such that $\|\mathbf{u}\|_1 = r$. We will use the FTRL/Mirror descent with the entropic regularizer. Recall that the regret of this algorithm is driven by the term

$$\mathbb{E}[\sum_{i=1}^{d} \mathbf{w}_t(i)\tilde{\ell}_i^{\,2}].$$

We will let $\mathbf{p}_t$ be the uniform distribution. Further, recall that $\ell_t \cdot \mathbf{u} \leq r$ for all $\mathbf{u} \in S$. These imply that

$$\|\tilde{\ell}_t\|_\infty \leq \frac{rd}{\delta}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^{d} \mathbf{w}_t(i)\tilde{\ell}_i^{\,2}] &\leq \sum_{i=1}^{d} \mathbf{p}_t(i)\mathbf{w}_t(i)\frac{r^2d^2}{\delta^2} \\
&= \frac{1}{d}\frac{r^2d^2}{\delta^2}\sum_{i=1}^{d}\mathbf{w}_t(i) \\
&= \frac{r^3d}{\delta^2}.
\end{aligned}$$

Also, recall that for this regularizer, and any $\mathbf{u} \in S$, the term $R(u) - R(\mathbf{w}_1) \leq r\log(d/r)$. Finally, for this set S, the condition in (2) holds. Putting this all together, the total regret of the algorithm, for some choice of partameters η and δ, ignoring constant terms, is

$$\frac{r\log(d/r)}{\eta} + \eta\frac{r^3d}{\delta^2}T + \delta T.$$

To minimize this, we want to choose δ such that the last two terms are equal, i.e.,

$$\delta^3 = \eta r^3 d.$$

The choice of η equalizes the first two terms, i.e.,

$$\begin{aligned}
\eta^2 &= \frac{\log(d/r)\delta^2}{r^2dT} \\
&= \frac{\log(d/r)\eta^{2/3}r^2d^{2/3}}{r^2dT} \\
\Rightarrow \eta^{4/3} &= \frac{\log(d/r)}{d^{1/3}T} \\
\Rightarrow \eta &= \frac{\log(d/r)^{3/4}}{d^{1/4}T^{3/4}} \\
\Rightarrow \delta &= \frac{rd^{1/4}\log(d/r)^{1/4}}{T^{1/4}} \\
\Rightarrow \delta T &= rd^{1/4}\log(d/r)^{1/4}T^{3/4}
\end{aligned}$$

This is not the optimal regret bound; the optimal bound gets a $\sqrt{T}$ dependence. It uses an estimator with a similar idea, but uses a spherical sampler. Note that the estimator we used (with $\mathbf{p}_t$ being the uniform distribution) can be written as

$$\tilde{\ell}_t := \frac{d}{\delta}[\ell_t \cdot (\mathbf{w}_t + \delta\sigma_t \mathbf{e}(i_t))]\sigma_t \mathbf{e}(i_t).$$

An alternate estimator is

$$\tilde{\ell}_t := \frac{d}{\delta}[\ell_t \cdot (\mathbf{w}_t + \delta \mathbf{v}_t)]\mathbf{v}_t,$$

where $\mathbf{v}_t$ is a random vector from the unit sphere, i.e., $\|\mathbf{v}_t\|_2 = 1$. This can be further extended to sample from an ellipsoid instead of a sphere.

Another difference is the usage of what are called self-concordant barrier functions as regularizers. These functions were first defined in the design of interior point methods in optimization, a very powerful technique. You will hear more about this in one of the guest lectures towards the end of the quarter.

2 Contextual Bandits

This is a variant, where there is a distinction between actions and experts. Actions are what the algorithm can pick, and suppose that there are d of them, which we identify with unit vectors in $\mathbb{R}^d$. Suppose that there are N experts, and in each round, each expert suggests her own probability distribution over actions. This is intended to capture the "context". The experts might have some side information about what the losses might be in round t. They themselves might be doing their own learning. The goal is to compete with the best fixed expert, not the best fixed action. Suppose that the probability distribution chosen by expert j in round t is $\mathbf{p}_{jt} \in \Delta^d$. As earier, we will assume that these are oblivious. Suppose algorithm picks $i_t \in [d]$ in round t. The regret is then

$$\text{REGRET} = \sum_{t=1}^{T} \ell_t(i_t) - \min_{j \in [N]} \sum_{t=1}^{T} \ell_t \cdot \mathbf{p}_{jt}.$$

As before, algorithm only observes $\ell_t(i_t)$.

The algorithm we will consider will

- simulate FTRL with the entropic regularizer, on the set Δ^N, with estimated losses.
- Let $\mathbf{w}_t \in \Delta^N$ be the choice of the algorithm in round t in the simulation.
- Pick $i_t = i$ with probability $\mathbf{q}_t(i) = \sum_{j=1}^{N} \mathbf{w}_t(j)\mathbf{p}_{jt}(i)$.
- Construct an estimate, for each $i \in [d]$, $\tilde{\ell}_t(i) = \mathbf{1}(i_t = i)\ell_t(i)/\mathbf{q}_t(i)$
- Define the estimated loss of expert j in round t as $\mathbf{p}_{jt} \cdot \tilde{\ell}_t$.

As throughout, it is easy to see that $\mathbb{E}[\tilde{\ell}_t(i)] = \ell_t(i)$. Once again the key term is

$$\begin{aligned}
\mathbb{E}_{i_t}[\sum_{j=1}^{N} \mathbf{w}_t(j)[\mathbf{p}_{jt} \cdot \tilde{\ell}_t]^2] &= \sum_{j=1}^{N} \mathbf{w}_t(j)\mathbb{E}_{i_t}[(\mathbf{p}_{jt} \cdot \tilde{\ell}_t)^2] \\
&\leq \sum_{j=1}^{N} \mathbf{w}_t(j)\mathbb{E}_{i_t}[\mathbf{p}_{jt} \cdot \tilde{\ell}_t^{\,2}] \\
&= \sum_{j=1}^{N} \mathbf{w}_t(j) \sum_{i=1}^{d} \mathbf{q}_t(i)\mathbf{p}_{jt}(i)\ell_t(i)^2/\mathbf{q}_t(i)^2 \\
&= \sum_{i=1}^{d} \ell_t(i)^2 \sum_{j=1}^{N} \mathbf{w}_t(j)\mathbf{p}_{jt}(i)/\mathbf{q}_t(i) \\
&= \sum_{i=1}^{d} \ell_t(i)^2 \\
&\leq d.
\end{aligned}$$

As previously, we have that $R(u) - R(\mathbf{w}_1) \leq \log(N)$. This gives a regret of

$$\frac{\log N}{\eta} + \eta dT.$$

As usual, by setting

$$\eta = \sqrt{\frac{\log N}{dT}},$$

we get a regret of

$$O(\sqrt{Td\log N}).$$