Machine Learning Theory (CS 6783)

Lecture 21: Oracle Efficient Contextual Bandits

1 Online Square Loss Regression Oracle

While the above approach does give an algorithm that is ERM oracle efficient, the above requires finding the ERM. However note that even if we had only two actions, the ERM optimization can be as bad as optimizing w.r.t. binary classification loss which in most cases is again computationally hard. The typical way out of this for classification is to replace the zero-one loss by some nicer losses like square loss or logistic loss etc. In this section, under the so called realizability assumption we can show that one can get contextual bandit algorithms that are online squared loss regression oracle efficient. To be able to achieve this, the algorithms require the following realizability assumption.

Assumption 1. *Assume that there is a class $\mathcal{L} \subset \mathbb{R}^{\mathcal{X}\times[N]}$ such that for some $g^* \in \mathcal{L}$, we have that for any $x \in \mathcal{X}$ and $a \in [N]$,*

$$\mathbb{E}\left[\ell_t(a)|x_t = x\right] = g^*(x, a)$$

The assumption tells us that the conditional expected loss of any action given a context is modeled will by a member g^* of some class $\mathcal{L}$. The rough idea then is to learn this model in some sense and take actions that are (close to) optimal w.r.t. this model given a context. More specifically, on every round we make a prediction of the loss given context for every action based on our ability to solve online squared loss regression w.r.t. class $\mathcal{L}$. Then we take a distribution that is skewed towards making the best decision based on this model for losses. Specifically we use the following algorithm.

SquareCB:

Set $\gamma = \sqrt{\frac{4N}{\text{RegSQ}_n(\mathcal{L})}}$

For $t = 1$ to n

Receive context $x_t \in \mathcal{X}$

For each action $a \in [N]$, compute $\hat{y}_t[a] = \hat{y}_t(x_t, a)$ by feeding x_t, a as input to the square loss regression algorithm for round t

Set $b_t = \underset{b\in[N]}{\text{argmin}}\, \hat{y}_t[b]$

$\forall a \neq b_t$, set $p_t(a) = \frac{1}{N+\gamma(\hat{y}_t[a]-\hat{y}_t[b_t])}$, set $p_t(b_t) = 1 - \sum_{a\neq b_t} p_t(a)$

Draw action $a_t \sim p_t$ and observe $\ell_t[a_t]$

Use online regression algorithm by feeding it input instance (x_t, a_t) and output $\ell_t[a_t]$

End For

Theorem 2. *Assume we have access to an online regression oracle that guarantees that for any sequence of context action pairs $(x_1, a_1), \ldots, (x_n, a_n)$ as input and any labels $y_1, \ldots, y_n$ produced possibly by an adversary, we have an online learning algorithm that guarantees that:*

$$\frac{1}{n}\sum_{t=1}^{n}((\hat{y}_t - y_t)^2 - \inf_{g\in\mathcal{L}}\frac{1}{n}\sum_{t=1}^{n}\left(g^*(a_t, x_t) - y_t\right)^2 \leq \mathrm{RegSQ}_n(\mathcal{L})$$

where $\mathrm{RegSQ}_n(\mathcal{L})$ *is the bound guaranteed for online squared loss regression against* $\mathcal{L}$*. Then, the Square CB algorithm enjoys the regret bound,*

$$\mathbb{E}\left[\mathrm{Reg}_n\right] \leq \sqrt{3N\ \mathrm{RegSQ}_n(\mathcal{L})}$$

where

Proof.

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Reg}_n\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\ell_t(a_t) - \frac{1}{n}\sum_{t=1}^{n}\ell_t(f^*(x_t))\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\ell_t(a_t)|x = x_t\right] - \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\ell_t(f^*(x_t))|x = x_t\right]\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}g^*(a_t, x_t) - \frac{1}{n}\sum_{t=1}^{n}g^*(f^*(x_t), x_t)\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(g^*(a_t, x_t) - g^*(f^*(x_t), x_t) - \frac{\gamma}{2}\left(\hat{y}_t(x_t, a_t) - g^*(a_t, x_t)\right)^2\right)\right] \\
&\qquad + \frac{\gamma}{2}\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\hat{y}_t(x_t, a_t) - g^*(a_t, x_t)\right)^2\right]
\end{aligned}$$

But due to realizability, $\left(\hat{y}_t(x_t, a_t) - g^*(a_t, x_t)\right)^2 = \mathbb{E}\left[\left(\hat{y}_t(x_t, a_t) - \ell_t[a_t]\right)^2 - \left(g^*(a_t, x_t) - \ell_t[a_t]\right)^2 |x_t = x\right]$

$$\begin{aligned}
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(g^*(a_t, x_t) - g^*(f^*(x_t), x_t) - \frac{\gamma}{2}\left(\hat{y}_t(x_t, a_t) - g^*(a_t, x_t)\right)^2\right)\right] \\
&\qquad + \frac{\gamma}{2}\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}((\hat{y}_t(x_t, a_t) - \ell_t[a_t])^2 - (g^*(a_t, x_t) - \ell_t[a_t])^2\right]
\end{aligned}$$

replacing $g^*(\cdot, x_t)$ by taking supremum over vector $g^* \in [0,1]^N$ for each round, and replacing $f^*(x_t)$ by macimum $a^* \in [N]$ we move to upper bound,

$$\begin{aligned}
&\leq \frac{1}{n}\sum_{t=1}^{n} \sup_{g^* \in [0,1]^N} \max_{a^* \in [N]} \mathbb{E}_{a_t \sim p_t}\left[g^*[a_t] - g^*[a^*] - \frac{\gamma}{2}\left(\hat{y}_t(x_t, a_t) - g^*[a_t]\right)^2\right] \\
&\qquad + \frac{\gamma}{2}\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}((\hat{y}_t(x_t, a_t) - \ell_t[a_t])^2 - (g^*(a_t, x_t) - \ell_t[a_t])^2\right] \\
&\leq \frac{1}{n}\sum_{t=1}^{n} \sup_{g^* \in [0,1]^N} \max_{a^* \in [N]} \mathbb{E}_{a_t \sim p_t}\left[g^*[a_t] - g^*[a^*] - \frac{\gamma}{2}\left(\hat{y}_t(x_t, a_t) - g^*[a_t]\right)^2\right] \\
&\qquad + \frac{\gamma}{2}\mathrm{RegSQ}_n(\mathcal{L})
\end{aligned}$$

where in the above, $\mathrm{RegSQ}_n(\mathcal{L})$ is the regret bound for online square loss regression w.r.t. loss class $\mathcal{L}$. It can be shown that for the choice of distribution p_t (shown in next lemma), we have that for any t:

$$\sup_{g^* \in [0,1]^N} \max_{a^* \in [N]} \mathbb{E}_{a_t \sim p_t}\left[g^*[a_t] - g^*[a^*] - \frac{\gamma}{2}\left(\hat{y}_t(x_t, a_t) - g^*[a_t]\right)^2\right] \leq \frac{3N}{2\gamma}$$

Using this we can conclude that:

$$\mathbb{E}\left[\mathrm{Reg}_n\right] \leq \frac{3N}{2\gamma} + \frac{\gamma}{2}\mathrm{RegSQ}_n(\mathcal{L})$$

using $\gamma = \sqrt{\frac{3N}{\mathrm{RegSQ}_n(\mathcal{L})}}$ we obtain that:

$$\mathbb{E}\left[\mathrm{Reg}_n\right] \leq \sqrt{3N\ \mathrm{RegSQ}_n(\mathcal{L})}$$

□

The point to note is that for a finite class $\mathcal{L}$, it turns out the exponential weights algorithm can actually ensure that $\mathrm{RegSQ}_n(\mathcal{L}) \leq \frac{\log|\mathcal{L}|}{n}$ and so for finite $\mathcal{L}$ class one has,

$$\mathbb{E}\left[\mathrm{Reg}_n\right] \leq \sqrt{\frac{2N\log|\mathcal{L}|}{n}}$$

Lemma 3. *For any vector $\hat{y} \in [0,1]^N$, let $b^* = \operatorname{argmin}_{a \in [N]} \hat{y}[a]$. Let distribution $p \in \Delta_N$ be given by, $\forall a \neq b^*, p(a) = \frac{1}{N + \gamma(\hat{y}[a] - \hat{y}[b^*])}$ and $p(b^*) = 1 - \sum_{a \neq b^*} p(a)$, then,*

$$\sup_{g \in [0,1]^N} \max_{a^* \in [N]} \mathbb{E}_{a \sim p}\left[g[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right] \leq \frac{3N}{2\gamma}$$

Proof. Now consider any $a^* \in [N]$ and any $g \in [0,1]^N$. Note that,

$$\begin{aligned}
&\mathbb{E}_{a\sim p}\left[g[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right] \\
&\quad= \sum_{a\in[A]} p(a)\left(g[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right) \\
&\quad= \sum_{a\neq a^*} p(a)\left(g[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right) - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad= \sum_{a\neq a^*} p(a)\left(g[a] - \hat{y}[a] + \hat{y}[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right) - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2
\end{aligned}$$

But now note that $g[a] - \hat{y}[a] \leq \frac{\gamma}{2}(\hat{y}[a] - g[a])^2 + \frac{1}{2\gamma}$ and so we have,

$$\begin{aligned}
&\mathbb{E}_{a\sim p}\left[g[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right] \\
&\quad= \sum_{a\neq a^*} p(a)\left(g[a] - \hat{y}[a] + \hat{y}[a] - g[a^*] - \frac{\gamma}{2}(\hat{y}[a] - g[a])^2\right) - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad\leq \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - g[a^*] + \frac{1}{2\gamma}\right) - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad= \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - g[a^*]\right) + \frac{1-p(a^*)}{2\gamma} - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad= \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - \hat{y}[a^*] + \hat{y}[a^*] - g[a^*]\right) + \frac{1-p(a^*)}{2\gamma} - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad= \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - \hat{y}[a^*]\right) + (1-p[a^*])\left(\hat{y}[a^*] - g[a^*]\right) + \frac{1-p(a^*)}{2\gamma} - \frac{p(a^*)\gamma}{2}(\hat{y}[a^*] - g[a^*])^2 \\
&\quad= \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - \hat{y}[a^*]\right) + (1-p[a^*])\left(\hat{y}[a^*] - g[a^*] - \frac{p(a^*)\gamma}{2(1-p[a^*])}(\hat{y}[a^*] - g[a^*])^2\right) + \frac{1-p(a^*)}{2\gamma}
\end{aligned}$$

Again using AM-GM to note that $\left(\hat{y}[a^*] - g[a^*] - \frac{p(a^*)\gamma}{2(1-p[a^*])}(\hat{y}[a^*] - g[a^*])^2\right) \leq \frac{1-p(a^*)}{2\gamma p(a^*)}$ we conclude that,

$$\begin{aligned}
&\leq \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - \hat{y}[a^*]\right) + \frac{(1-p(a^*))^2}{2\gamma p(a^*)} + \frac{1-p(a^*)}{2\gamma} \\
&= \sum_{a\neq a^*} p(a)\left(\hat{y}[a] - \hat{y}[a^*]\right) + \frac{(1-p(a^*))}{2p(a^*)\gamma}
\end{aligned}$$

Recall that $b^* = \underset{a \in [N]}{\operatorname{argmin}} \hat{y}[a]$,

$$
\begin{aligned}
&= \sum_{a \neq a^*} p(a) \left(\hat{y}[a] - \hat{y}[b^*] + \hat{y}[b^*] - \hat{y}[a^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma} \\
&= \sum_{a \neq a^*} p(a) \left(\hat{y}[a] - \hat{y}[b^*]\right) + (1 - p(a^*)) \left(\hat{y}[b^*] - \hat{y}[a^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma} \\
&= \sum_{a=1}^{N} p(a) \left(\hat{y}[a] - \hat{y}[b^*]\right) - \left(\hat{y}[a^*] - \hat{y}[b^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma} \\
&= \sum_{a : a \neq b^*} \frac{\left(\hat{y}[a] - \hat{y}[b^*]\right)}{N + \gamma((\hat{y}[a] - \hat{y}[b^*]))} - \left(\hat{y}[a^*] - \hat{y}[b^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma} \\
&= \sum_{a : a \neq b^*} \frac{1}{\frac{N}{(\hat{y}[a] - \hat{y}[b^*])} + \gamma} - \left(\hat{y}[a^*] - \hat{y}[b^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma}
\end{aligned}
$$

since each $(\hat{y}[a] - \hat{y}[b^*]) \leq 1$,

$$
\begin{aligned}
&\leq \frac{N-1}{N+\gamma} - \left(\hat{y}[a^*] - \hat{y}[b^*]\right) + \frac{(1 - p(a^*))}{2p(a^*)\gamma} \\
&\leq \frac{N-1}{N+\gamma} + \max\left\{ \frac{(1 - p(b^*))}{2p(b^*)\gamma}, \max_{a \neq b^*} \left\{ \frac{(1 - p(a))}{2p(a)\gamma} - \left(\hat{y}[a] - \hat{y}[b^*]\right) \right\} \right\} \\
&= \frac{N-1}{N+\gamma} + \max\left\{ \frac{(1 - p(b^*))}{2p(b^*)\gamma}, \max_{a \neq b^*} \left\{ \frac{N + \gamma\left(\hat{y}[a] - \hat{y}[b^*]\right)}{2\gamma} - \frac{1}{2\gamma} - \left(\hat{y}[a] - \hat{y}[b^*]\right) \right\} \right\} \\
&= \frac{N-1}{N+\gamma} + \max\left\{ \frac{(1 - p(b^*))}{2p(b^*)\gamma}, \max_{a \neq b^*} \left\{ \frac{N-1}{2\gamma} - \frac{1}{2}\left(\hat{y}[a] - \hat{y}[b^*]\right) \right\} \right\}
\end{aligned}
$$

Note that $p(b^*) \geq 1/N$ because we are picking b^* with highest probability, hence

$$
\begin{aligned}
&\leq \frac{N-1}{N+\gamma} + \max\left\{ \frac{N-1}{2\gamma}, \max_{a \neq b^*} \left\{ \frac{N-1}{2\gamma} - \frac{1}{2}\left(\hat{y}[a] - \hat{y}[b^*]\right) \right\} \right\} \\
&= \frac{N-1}{N+\gamma} + \frac{N-1}{2\gamma} \leq \frac{\frac{3}{2}(N-1)}{\gamma}
\end{aligned}
$$

□