| | |
|---|---|
| **CS281B/Stat241B (Spring 2008) Statistical Learning Theory** | **Lecture: 25** |

Follow the perturbed leader, online shortest path

*Lecturer: Sasha Rakhlin* *Scribes: Barlas Oğuz, David Burkett*

# 1 Oracle Regret

We begin by a lemma

**Lemma 1.1.** If

$$x_{t+1} = \arg\min_{x \in K} \eta \sum_{s=1}^{t} l_s(x) + R(x)$$

then for any $u \in K$

$$\sum_{t=1}^{T} l_t(x_{t+1}) \leq \sum_{t=1}^{T} l_t(u) + \eta^{-1}(R(u) - R(x_1)) \tag{1}$$

where $x_1 = \arg\min R(x)$.

This says that the hypothetical forecaster who knows $l_t$ before having to predict $x_t$ is almost perfect.

Proof. Use induction.

$T = 0$ true, because $R(x_1) \leq R(u), \forall u \in K$.

Suppose statement holds for $T - 1$.

$$\sum_{t=1}^{T-1} l_t(x_{t+1}) + \eta^{-1} R(x_1) \leq \sum_{t=1}^{T-1} l_t(u) + \eta^{-1} R(u), \forall u \tag{2}$$

in particular, this holds for $u = x_{T+1}$.

$$\sum_{t=1}^{T-1} l_t(x_{t+1}) + \eta^{-1} R(x_1) \leq \sum_{t=1}^{T-1} l_t(x_{T+1}) + \eta^{-1} R(x_{T+1}) \tag{3}$$

Add $l_T(x_{T+1})$ to both sides

$$\sum_{t=1}^{T} l_t(x_{t+1}) + \eta^{-1} R(x_1) \leq \sum_{t=1}^{T} l_t(x_{T+1}) + \eta^{-1} R(x_{T+1}) \tag{4}$$

$$\sum_{t=1}^{T} l_t(x_{t+1}) + \eta^{-1} R(x_1) \leq \sum_{t=1}^{T} l_t(u) + \eta^{-1} R(u), \forall u \tag{5}$$

□

As a consequence, for $x^* = \arg\min_{x \in K} \sum_{t=1}^{T} l_t(x)$,

$$\sum_{t=1}^{T} l_t(x_t) - \sum_{t=1}^{T} l_t(x^*) \leq \sum_{t=1}^{T} (l_t(x_t) - l_t(x_{t+1})) + \eta^{-1}(R(x^*) - R(x_1)) \tag{6}$$

## 2 Regret Bounds for Follow the Perturbed Leader

Now we focus on linear $l_t(.)$

Solve $x_{t+1} = \arg\min_{x\in K} \eta \sum_{s=1}^{t} l_s x + rx$, where $r$ is drawn at the beginning of the game from distribution $f$.

Assume *oblivious adversary*: $l_1 \dots l_T$ are chosen by the adversary before the game.

**Theorem 2.1** (A)**.** Suppose $l_t \in \mathbb{R}_+^n, K \subset \mathbb{R}_+^n, f(r)$ has support on $\mathbb{R}_+^n$.

$$\forall u \in K, \mathbb{E}\sum_{t=1}^{T} l_t x_t \leq \sum_{t=1}^{T} l_t u + \sum_{t=1}^{T} l_t \int_{\{r: f(r)\geq f(r-\eta l_t)\}} x_t f(r)\, \mathrm{d}r + \eta^{-1}\mathbb{E}\sup_{x\in K} rx \tag{7}$$

Proof. Let $x^* = \arg\min \sum_{t=1}^{T} l_t x$.

$$\sum_{t=1}^{T} l_t x_t - \sum_{t=1}^{T} l_t x^* \leq \sum_{t=1}^{T} l_t(x_t - x_{t+1}) + \eta^{-1}(rx^* - rx_1) \tag{8}$$

and since $rx \geq 0$,

$$\sum_{t=1}^{T} l_t x_t - \sum_{t=1}^{T} l_t x^* \leq \sum_{t=1}^{T} l_t(x_t - x_{t+1}) + \eta^{-1} rx^* \tag{9}$$

Taking expectations, we have

$$\mathbb{E}(\sum_{t=1}^{T} l_t x_t - \sum_{t=1}^{T} l_t x^*) \leq \mathbb{E}(\sum_{t=1}^{T}(l_t x_t - l_t x_{t+1})) + \eta^{-1}(rx^* - rx_1) \tag{10}$$

We want to have $\mathbb{E}(\sum_{t=1}^{T}(l_t x_t - l_t x_{t+1}))$ small.

$$\mathbb{E} l_t x_t = \int l_t \arg\min_x ((\eta\sum_{s=1}^{t-1} l_s + r)x) f(r)\, \mathrm{d}r \tag{11}$$

$$\mathbb{E} l_t x_{t+1} = \int l_t \arg\min_x ((\eta\sum_{s=1}^{t} l_s + r)x) f(r)\, \mathrm{d}r \tag{12}$$

Let $r' = r + l_t$

$$\mathbb{E} l_t x_t = \int l_t \arg\min_x ((\eta\sum_{s=1}^{T-1} l_s + r')x) f(r' - \eta l_t)\, \mathrm{d}r' \tag{13}$$

$$\mathbb{E} l_t(x_t - x_{t+1}) = \int l_t x_t (f(r) - f(r - \eta l_t))\mathrm{d}r = l_t \int x_t (f(r) - f(r-\eta l_t))\mathrm{d}r \tag{14}$$

$$\leq l_t \int_{\{r: f(r) > f(r-\eta l_t)\}} x_t f(r)\mathrm{d}r = l_t \mathbb{E} x_t 1_{\{r: f(r)>f(r-\eta l_t)\}} \tag{15}$$

□

Now, we prove an analogous theorem where we relax the restriction to the positive orthant.

**Theorem 2.2** (B).

$$\mathbb{E}\sum_{t=1}^{T} l_t x_t \leq \sup r, t \frac{f(r)}{f(r-\eta l_t)} \left[\sum l_t + \eta^{-1}\mathbb{E}\sup_{x\in K} r\cdot x + \eta^{-1}\mathbb{E}\sup_{x\in K} -r\cdot x\right] \tag{16}$$

for any $u \in K$

PROOF.

$$\begin{aligned}\mathbb{E}l_t x_t &= \int l_t \operatorname{argmin}_{x\in K}\left[\left(\eta\sum_{s=1}^{t-1} l_s + r\right)x\right] f(r)dr && (17)\\ &\leq \sup_{r,t}\frac{f(r)}{f(r-\eta l_t)}\int l_t \operatorname{argmin}_{x\in K}\left[\left(\eta\sum_{s=1}^{t} l_s + r\right)x\right] f(r)dr && (18)\\ &= \sup_{r,t}\frac{f(r)}{f(r-\eta l_t)}\mathbb{E}l_t x_{t+1} && (19)\end{aligned}$$

By lemma 1.1,

$$\forall u\in K, \sum_{t=1}^{T} l_t x_{t+1} \leq \sum_{t=1}^{T} l_t\cdot u + \eta^{-1}\sup_{x\in K}(r\cdot x) + \eta^{-1}\sup_{x\in K}(-r\cdot x) \tag{20}$$

So,

$$\mathbb{E}\sum_{t=1}^{T} l_t x_t = \sup_{r,t}\frac{f(r)}{f(r-\eta l_t)}\mathbb{E}\left(\sum_{t=1}^{T} l_t\cdot u + \eta^{-1}\sup_{x\in K}(r\cdot x) + \eta^{-1}\sup_{x\in K}(-r\cdot x)\right) \tag{21}$$

□

# 3 Examples

## 3.1 Expert Setting

Here, $K$ is the $n-$simplex. We will draw $r \sim \text{Unif}\left([0,1]^N\right)$ and $l_t \in [0,1]^N$

Applying Theorem A,

$$\begin{aligned}\mathbb{E}R_T &\leq \sum_{t=1}^{T} l_t \int_{\{r:f(r)>f(r-\eta l_t)\}} x_t f(r)dr + \eta^{-1}\underbrace{\mathbb{E}\sup_{x\in K} r\cdot x}_{\leq 1} && (22)\\ &\leq \sum_{t=1}^{T}\int_{\{r:f(r)>f(r-\eta l_t)\}} f(r)dr + \eta^{-1} && (23)\\ &\leq \sum_{t=1}^{T} Vol(\{r:\exists i, r_i - \eta l_t(i) < 0\}) + \eta^{-1} && (24)\\ &\leq \sum_{t=1}^{T}\eta\sum_{i=1}^{N} l_t(i) + \eta^{-1} && (25)\\ &\leq \eta^{-1} + T\eta N && (26)\\ &= 2\sqrt{TN} \text{ (with } \eta = \tfrac{1}{\sqrt{TN}}\text{)} && (27)\end{aligned}$$

By using Theorem B, this result can be improved to replace the $N$ term with $\log(N)$.

## 3.2 Online Shortest Path

In this setting, there is a fixed DAG with labeled vertices $u$ and $v$ such that $v$ is reachable from $u$. At each time step, the player picks a path from $u$ to $v$, and then the opponent reveals the cost of each edge. The loss is the cost of the chosen path.

Each path can be associated with some $x \in \{0,1\}^{|E|}$, and the set of paths is $P \subseteq \{0,1\}^{|E|}$.

The adversary picks $l_t \in \mathbb{R}_+^n$, so the loss can be written as $l_t \cdot x_t$ as usual.

To use the Follow the Perturbed Leader methodology, we draw $r \sim \text{Unif}([0,1]^{|E|})$. Suppose, $l_t \in [0,1]^{|E|}$ and the length of the longest path (number of edges) is $\xi$.

Then, applying Theorem A,

$$\begin{aligned} l_t \int_{\{r:f(r)>f(r-\eta l_t)\}} x_t f(r) dr &\leq ||l_t||_\infty \left|\left| \int_{\{r:f(r)>f(r-\eta l_t)\}} x_t f(r) dr \right|\right|_1 && (28) \\ &\leq \xi \int_{\{r:f(r)>f(r-\eta l_t)\}} f(r) dr && (29) \\ &\leq \xi |E| \eta && (30) \end{aligned}$$

So,

$$\begin{aligned} \mathbb{E} R_T &\leq \eta^{-1}\xi + \xi|E|\eta T && (31) \\ &= 2\xi\sqrt{|E|T} \text{ (with } \eta = \tfrac{1}{\sqrt{|E|T}}) && (32) \end{aligned}$$

By using Theorem B, this result can be improved to replace the $|E|$ term with $\log(|E|)$.