Stat 260/CS 294-102. Learning in Sequential Decision Problems.

Peter Bartlett

1. Linear bandits.
   - Exponential weights with unbiased loss estimates.
   - Controlling loss estimates and their variance.

Recall: Linear bandits

At round t,

- Strategy chooses $a_t \in \mathcal{A} \subset \mathbb{R}^d$.
- Adversary chooses *linear* loss $\ell_t \in \mathcal{L} \subseteq [-1,1]^{\mathcal{A}}$.
- Strategy sees loss $\ell_t(a_t)$.

Loss is *linear* in action.

Aim to minimize pseudo-regret:

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^{n} \ell_t(a_t) - \inf_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{n} \ell_t(a).$$

Exponential weights for linear bandits

Given $\mathcal{A}$, distribution μ on $\mathcal{A}$, mixing coefficient $\gamma > 0$, learning rate $\eta > 0$,

set q_1 uniform on $\mathcal{A}$.

for $t = 1, 2, \ldots, n$,

1. $p_t = (1-\gamma)q_t + \gamma\mu$
2. choose $a_t \sim p_t$
3. observe $\ell_t^T a_t$
4. update $q_{t+1}(a) \propto q_t(a) \exp(-\eta \tilde{\ell}_t^T a)$,

where $$\tilde{\ell}_t = \Sigma_t^{-1} a_t a_t^T \ell_t,$$
$$\Sigma_t = \mathbb{E}_{a \sim p_t} a a^T.$$

Unbiased loss estimates

- Assume $\mathrm{span}(\mathcal{A}) = \mathbb{R}^d$ (otherwise, we can project to a lower dimension) and that μ has support on a d-dimensional set. So $\mathbb{E}_{a \sim p_t} a a^T$ has rank d.
- Strategy observes $a_t^T \ell_t$ and a_t, so it can compute

$$\tilde{\ell}_t = \Sigma_t^{-1} a_t \left(a_t^T \ell_t \right).$$

- $\tilde{\ell}_t$ is unbiased:

$$\mathbb{E}\left[\tilde{\ell}_t | \mathcal{F}_{t-1} \right] = \left(\mathbb{E}_{a \sim p_t} a a^T \right)^{-1} \left(\mathbb{E}_{a_t \sim p_t} a_t a_t^T \right) \ell_t = \ell_t.$$

Regret bound

Theorem: For $\eta \sup_{a \in \mathcal{A}} \left| \tilde{\ell}_t^T a \right| \le 1$,

$$\overline{R}_n \le \gamma n + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \mathbb{E}\mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2.$$

So we need to control η times the magnitude of the loss estimates,

$$\eta \sup_{a \in \mathcal{A}} \left| \tilde{\ell}_t^T a \right|$$

and the variance term,

$$\mathbb{E}\mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2.$$

Proof

The regret is

$$\mathbb{E}\left[\sum_{t=1}^{n}\left(\ell_t^T a_t - \ell_t^T a^*\right)\right].$$

We've seen that, given history $\mathcal{F}_{t-1}$,

$$\mathbb{E}\left[\tilde{\ell}_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\Sigma_t^{-1} a_t a_t^T \ell_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\ell_t|\mathcal{F}_{t-1}\right].$$

Lemma: Some unbiased estimates involving $\tilde{\ell}_t$:

$$\mathbb{E}\left[\ell_t^T a\right] = \mathbb{E}\left[\tilde{\ell}_t^T a\right],$$

$$\mathbb{E}\left[\ell_t^T a_t\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} p_t(a)\mathbb{E}\left[\tilde{\ell}_t\middle|\mathcal{F}_{t-1}\right]^T a\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a\right].$$

Proof

So we can write the strategy's expected cumulative loss as

$$\mathbb{E}\sum_{t=1}^{n} \ell_t^T a_t = \mathbb{E}\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a.$$

We'll give up on the loss incurred in the exploration trials:

$$\begin{aligned}\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a &= \sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \left((1-\gamma)q_t(a)+\gamma\mu(a)\right)\tilde{\ell}_t^T a \\ &= (1-\gamma)\left(\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} q_t(a)\tilde{\ell}_t^T a\right) + \underbrace{\gamma\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mu(a)\tilde{\ell}_t^T a}_{\text{exploration}}.\end{aligned}$$

Proof

For q_t, we follow the standard analysis (see Adversarial Bandits), but instead of using non-negativity of the $\tilde{\ell}$s, we use a lower bound:

$$\begin{aligned}\log \mathbb{E} \exp\left(-\eta(X - \mathbb{E}X)\right) &\leq \mathbb{E}\left(\exp(-\eta X) - 1 + \eta X\right) \\ &\leq (e-2)\eta^2 \mathbb{E}X^2,\end{aligned}$$

where the last inequality uses $\exp(-x) \leq 1 - x + (e-2)x^2$ for $x \geq -1$.

So if $\eta \tilde{\ell}_t^T a \geq -1$ for all $a \in \mathcal{A}$, the previous analysis shows that, for any $a^* \in \mathcal{A}$, the first term above satisfies

$$\sum_{t=1}^{n} \sum_{a \in \mathcal{A}} q_t(a) \tilde{\ell}_t^T a \leq \sum_{t=1}^{n} \tilde{\ell}_t^T a^* + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \sum_{a \in \mathcal{A}} q_t(a) \left(\tilde{\ell}_t^T a\right)^2 .$$

Proof

Combining, and using the fact that $(1-\gamma)q_t(a) \leq p_t(a)$,

$$\sum_{t=1}^{n} \sum_{a \in \mathcal{A}} p_t(a) \tilde{\ell}_t^T a \leq \sum_{t=1}^{n} \tilde{\ell}_t^T a^*$$

$$+ \text{(exploration)} + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \sum_{a \in \mathcal{A}} p_t(a) \left(\tilde{\ell}_t^T a \right)^2 .$$

The unbiasedness lemma gives

$$\overline{R}_n \leq \gamma n + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2 .$$

Controlling variance

Lemma: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$, the variance term is bounded:

$$\mathbb{E}\mathbb{E}_{a\sim p_t}\left(\tilde{\ell}_t^T a\right)^2 \le d.$$

$$\begin{aligned}
\mathbb{E}\left(\tilde{\ell}_t^T a\right)^2 &= a^T \mathbb{E}\left(\tilde{\ell}_t \tilde{\ell}_t^T\right) a \\
&= a^T \mathbb{E}\left(\left(\ell_t^T a_t\right)^2 \Sigma_t^{-1} a_t a_t^T \Sigma_t^{-1}\right) a \\
&\le a^T \Sigma_t^{-1} \mathbb{E}\left(a_t a_t^T\right) \Sigma_t^{-1} a \\
&= a^T \Sigma_t^{-1} a.
\end{aligned}$$

$$\mathbb{E}_{a\sim p_t}\mathbb{E}\left(\tilde{\ell}_t^T a\right)^2 \le \mathbb{E}\,\mathrm{tr}\left(a^T \Sigma_t^{-1} a\right) = \mathrm{tr}\left(\Sigma_t^{-1}\mathbb{E}\left(a a^T\right)\right) = \mathrm{tr}\left(I\right) = d.$$

Controlling the magnitude of the estimator

Lemma: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$,

$$\left|\tilde{\ell}_t^T a\right| \leq \sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b.$$

$$\begin{aligned}\left|\tilde{\ell}_t^T a\right| &= \left|a_t^T \ell_t \left(\Sigma_t^{-1} a_t\right)^T a\right| \\ &\leq \left|a_t^T \ell_t\right| \left|a_t^T \Sigma_t^{-1} a\right| \\ &\leq \sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b.\end{aligned}$$

We'll see that typically $\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq c_d/\gamma$.

Regret bound

Theorem: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$, if

$$\sup_{a,b \in \mathcal{A}} a^T \Sigma_t^{-1} b \le \frac{c_d}{\gamma},$$

$$\text{setting } \eta = \sqrt{\frac{\log |\mathcal{A}|}{n\left((e-2)d + c_d\right)}}$$

$$\gamma = c_d \eta$$

$$\text{gives } \overline{R}_n \le 2\sqrt{n(d + c_d) \log |\mathcal{A}|}.$$

Barycentric spanner

(Suppose that $\mathcal{A} \subseteq \mathbb{R}^d$ spans $\mathbb{R}^d$.)

A *barycentric spanner* of $\mathcal{A}$ is a set $\{b_1, \ldots, b_d\}$ that spans $\mathbb{R}^d$ and satisfies:
for all $a \in \mathcal{A}$ there is an $\alpha \in [-1, 1]^d$ such that $a = B\alpha$, where
$B = \begin{pmatrix} b_1 & \cdots & b_d \end{pmatrix}$.

- Every compact $\mathcal{A}$ has a barycentric spanner.
- If linear functions can be efficiently optimized over $\mathcal{A}$, then there is an efficient algorithm for finding an approximate barycentric spanner (that is, $|\alpha_i| \leq 1 + \delta$; $O(d^2 \log d/\delta)$ linear optimizations).

Barycentric spanner

Lemma: If $\{b_1, \ldots, b_d\} \subset \mathcal{A}$ maximizes $\det(B)$, then it is a barycentric spanner.

Proof. For $a = B\alpha$,

$$\begin{aligned} |\det(B)| &\geq \left|\det\begin{pmatrix} a & b_2 & \cdots & b_d \end{pmatrix}\right| \\ &= \left|\sum_i \alpha_i \det\begin{pmatrix} b_i & b_2 & \cdots & b_d \end{pmatrix}\right| \\ &= |\alpha_1| \, |\det(B)| \,. \end{aligned}$$

□

Barycentric spanner

Theorem: For $\mathcal{A} \subseteq [-1,1]^d$ and μ uniform on a barycentric spanner of $\mathcal{A}$,

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq \frac{d^2}{\gamma}$$

(that is, $c_d \leq d^2$). Hence,

$$\overline{R}_n \leq 2d\sqrt{2n \log |\mathcal{A}|}.$$

$$\Sigma_t = \frac{\gamma}{d} B B^T + \underbrace{(1-\gamma) \sum_{a\in\mathcal{A}} q_t(a) a a^T}_{M}.$$

Barycentric spanner: Proof

$$
\begin{aligned}
\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b &\le \sup_{\alpha,\beta\in[-1,1]^d} \alpha^T B^T \Sigma_t^{-1} B\beta \\
&\le \sup_{\|\alpha\|=\|\beta\|=\sqrt{d}} \alpha^T B^T \Sigma_t^{-1} B\beta \\
&= d\lambda_{\max}\left(B^T \Sigma_t^{-1} B\right) \\
&= d\lambda_{\max}\left(B^{-1} \Sigma_t B^{-T}\right)^{-1} \\
&= \frac{d}{\lambda_{\min}\left(B^{-1}\left(\frac{\gamma}{d} B B^T + M\right) B^{-T}\right)} \\
&\le \frac{d^2}{\gamma\lambda_{\min}\left(B^{-1} B B^T B^{-T}\right)} = \frac{d^2}{\gamma},
\end{aligned}
$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the largest and smallest eigenvalues.

Other exploration distributions

Lemma:

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \le \frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\gamma \lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right)}.$$

$$\begin{aligned}
\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b &\le \sup_{a\in\mathcal{A}} \|a\|_2^2 \lambda_{\max}\left(\Sigma_t^{-1}\right) \\
&= \frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\lambda_{\min}\left(\Sigma_t\right)}. \\
\lambda_{\min}\left(\Sigma_t\right) &= \min_{\|v\|=1} \sum_{a\in\mathcal{A}} p_t(a) v^T a a^T v \\
&\ge \gamma \min_{\|v\|=1} \sum_{a\in\mathcal{A}} \mu(a) v^T a a^T v = \gamma \lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right).
\end{aligned}$$

John's distribution

Theorem: [John's Theorem] For any convex set $\mathcal{A} \subset \mathbb{R}^d$, denote the ellipsoid of minimal volume containing it as

$$E = \left\{ x \in \mathbb{R}^d : (x - c)^T M (x - c) \leq 1 \right\}.$$

Then there is a set $\{u_1, \ldots, u_m\} \subseteq E \cap \mathcal{A}$ of $m \leq d(d+1)/2 + 1$ contact points and a distribution p on this set such that any $x \in \mathbb{R}^d$ can be written

$$x = c + d \sum_{i=1}^{m} p_i \langle x - c, u_i - c \rangle (u_i - c),$$

where $\langle \cdot, \cdot \rangle$ is the inner product for which the minimal ellipsoid is the unit ball about its center c: $\langle x, y \rangle = x^T M y$.

John's distribution

This shows that

$$
\begin{aligned}
x - c &= d \sum_i p_i (u_i - c)(u_i - c)^T M (x - c) \\
\Leftrightarrow \qquad \tilde{x} &= d \sum_i p_i \tilde{u}_i \tilde{u}_i^T \tilde{x} \\
\Leftrightarrow \qquad \frac{1}{d} I &= \sum_i p_i \tilde{u}_i \tilde{u}_i^T,
\end{aligned}
$$

where $\tilde{u}_i = M^{1/2}(u_i - c)$, and similarly for $\tilde{x}$. Setting the exploration distribution μ to be the distribution p over the set of transformed contact points $\tilde{u}_i$, we see that, for $a, b \in \mathcal{A}$,

$$\tilde{a}^T \mathbb{E}_{u \sim \mu} u u^T \tilde{b} = \frac{1}{d} \tilde{a}^T \tilde{b}.$$

John's distribution

So if we shift the origin of the set $\mathcal{A}$ and of the u_i (and the corresponding introduction of a constant component in the losses), we have

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq \frac{d}{\gamma},$$

that is, $c_d \leq d$. Hence,

$$\overline{R}_n \leq 2\sqrt{2nd \log|\mathcal{A}|}.$$

Exploration distributions

- (Dani, Hayes, Kakade, 2008):
  For μ uniform over *barycentric spanner*,
  $$\overline{R}_n = O\left(d\sqrt{n \log |\mathcal{A}|}\right) = \tilde{O}\left(d^{3/2}\sqrt{n}\right).$$
- (Cesa-Bianchi and Lugosi, 2009):
  For several combinatorial problems, $\mathcal{A} \subseteq \{0,1\}^d$, μ uniform over $\mathcal{A}$ gives
  $$\frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right)} = O(d),$$
  so
  $$\overline{R}_n = O\left(\sqrt{dn \log |\mathcal{A}|}\right) = \tilde{O}\left(d\sqrt{n}\right).$$
- (Bubeck, Cesa-Bianchi and Kakade, 2009): *John's Theorem*:
  $\tilde{O}\left(d\sqrt{n}\right)$.