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

Peter Bartlett

1. Linear bandits.
   - Lower bounds.

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 $\ell_t(a_t)$.

Recall: Regret bound for exponential weights

Use: $p_t = (1-\gamma)q_t + \gamma\mu$ where μ is an exploration distribution and q_t is the exponential weights distribution based on loss estimates

$$\tilde{\ell}_t = \Sigma_t^{-1} a_t a_t^T \ell_t,$$

$$\Sigma_t = \mathbb{E}_{a \sim p_t} a a^T.$$

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

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

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

Recall: Barycentric spanner

For μ uniform on a *barycentric spanner*:

$$\arg\max_{b_1,\ldots,b_d} \det \begin{pmatrix} b_1 & b_2 & \cdots & b_d \end{pmatrix}$$

we have

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

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

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

Recall: John's distribution

For any convex set $\mathcal{A} \subset \mathbb{R}^d$, there is a set of m contact points $u_1, \ldots, u_m$ between $\mathcal{A}$ and the ellipsoid of minimal volume containing it, and a distribution p on this set such that any $x \in \mathbb{R}^d$ can be written

$$x = d \sum_{i=1}^{m} p_i \langle x, u_i \rangle u_i,$$

where $\langle \cdot, \cdot \rangle$ is the inner product for which the minimal ellipsoid is the unit ball. Setting the exploration distribution μ to be the distribution p over the set of contact points, we see that

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

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

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

Lower bounds

Again, lower bounds from the stochastic setting suffice.

Theorem: Consider $\mathcal{A} = \{\pm 1\}^d$, $\mathcal{L} \supseteq \{\pm e_i : 1 \leq i \leq d\}$. There is a constant c such that, for any strategy and any n, there is an i.i.d. adversary for which

$$\overline{R}_n \geq cd\sqrt{n}.$$

(Here, $\sqrt{nd \log |\mathcal{A}|} = O(d\sqrt{n})$.)

Lower bounds: proof

Probabilistic method: Fix $\epsilon \in (0, 1/2)$ and, for each $b \in \{\pm 1\}^d$, define P_b on $\mathcal{L}$ as

$$P_b(e_i) = \frac{1 - b_i \epsilon}{2d},$$
$$P_b(-e_i) = \frac{1 + b_i \epsilon}{2d}.$$

(so that the optimal $a^* = b$). We'll choose b uniformly, and show that the expected regret under this choice is large.

Lower bounds: proof

$$
\begin{aligned}
\overline{R}_n(P_b) &= \sum_{t=1}^{n}\sum_{i=1}^{d}\mathbb{E}\left[\ell_{t,i}\left(a_{t,i}-b_i\right)\right] \\
&= \sum_{t=1}^{n}\sum_{i=1}^{d}(a_{t,i}-b_i)\left(\frac{1-2b_i\epsilon}{2d}-\frac{1+2b_i\epsilon}{2d}\right) \\
&= \sum_{t=1}^{n}\sum_{i=1}^{d}(b_i-a_{t,i})\frac{b_i\epsilon}{d} \\
&= \sum_{i=1}^{d}\underbrace{\frac{2\epsilon}{d}\sum_{t=1}^{n}1[a_{t,i}\neq b_i]}_{\overline{R}_n^i(b_i)}\,.
\end{aligned}
$$

Lower bounds: proof

The regret of sub-game i, $\overline{R}_n^i(b_i)$, is at least the regret that would be incurred if the strategy knew that the adversary was using one of the P_b distributions, and also knew $\{b_j : j \neq i\}$. In that case, it would know

$$\theta := \mathbb{E} \sum_{j \neq i} l_{t,j} a_{t,j},$$

and so at each round, it would see a (± 1) Bernoulli random variable $\ell_t^T a_t$, with mean

$$\theta - b_i a_{t,i} \frac{\epsilon}{d}.$$

Notice that the $1/d$ here is crucial: because information about the ith component only arrives once every d rounds on average, the range of values of the unknown Bernoulli mean has shrunk. If the strategy saw the components of ℓ_i (even in the semi-bandit setting, with $\mathcal{A} = \{0,1\}^d$ and feedback $(\ell_{t,1} a_{t,1}, \ldots, \ell_{t,d} a_{t,d})$), it would not suffer this disadvantage.

Lower bounds: proof

Using the same argument as we saw for the stochastic multi-armed bandit case (with a little extra work to show that θ is unlikely to be too close to 0 or 1, so that the variance of the Bernoulli is not too small), we see that

$$\mathbb{E}\overline{R}_n^i(b_i) \geq \frac{2\epsilon n}{d}\left(\frac{1}{2} - c\frac{\epsilon\sqrt{n}}{d}\right).$$

Choosing $\epsilon = d/(4c\sqrt{n})$ gives $\mathbb{E}\overline{R}_n^i(b_i) = \Omega(\sqrt{n})$, and so $\mathbb{E}\overline{R}_n(P_b) = \Omega(d\sqrt{n})$.