Linear Bandits

Sham M. Kakade

Outline

1. Linear Bandits
   - Setting
   - LinUCB and An Optimal Regret Bound
2. Analysis
   - Regret Analysis
   - Confidence Analysis

Handling Large Actions Spaces

- On each round, we must choose a decision $x_t \in D \subset R^d$.
- Obtain a reward $r_t \in [-1, 1]$, where

$$\mathbb{E}[r_t | x_t = x] = \mu^\star \cdot x \in [-1, 1],$$

  - so the the conditional expectation of r_t is linear)
  - Also, we have the *noise sequence*,

$$\eta_t = r_t - \mu^\star \cdot x_t$$

  is i.i.d noise.

model due to Abe & Long '99

Our Objective

If $x_0, \ldots x_{T-1}$ are our decisions, then our cumulative regret is

$$R_T = \mu^\star \cdot x^\star - \sum_{t=0}^{T-1} \mu^\star \cdot x_t$$

where $x^\star \in D$ is an optimal decision for $\mu^\star$, i.e.

$$x^\star \in \mathrm{argmax}_{x \in D} \, \mu^\star \cdot x$$

Outline

1. Linear Bandits
   - Setting
   - LinUCB and An Optimal Regret Bound
2. Analysis
   - Regret Analysis
   - Confidence Analysis

LinUCB & The "Confidence Ball"

- After t rounds, define our uncertainty region $\textsc{Ball}_t$: with center, $\widehat{\mu}_t$, and shape, Σ_t, using the λ-regularized least squares solution:

$$\widehat{\mu}_t = \arg\min_{\mu} \sum_{\tau=0}^{t-1} \|\mu \cdot x_\tau - r_\tau\|_2^2 + \lambda \|\mu\|_2^2$$

$$\Sigma_t = \lambda I + \sum_{\tau=0}^{t-1} x_\tau x_\tau^\top, \text{ with } \Sigma_0 = \lambda I$$

$$\textsc{Ball}_t = \left\{ \mu \mid (\widehat{\mu}_t - \mu)^\top \Sigma_t^{-1} (\widehat{\mu}_t - \mu) \leq \beta_t \right\},$$

  where β_t is a parameter of the algorithm.
- LinUCB: For $t = 0, 1, \ldots$
  1. Execute $x_t = \operatorname{argmax}_{x \in D} \max_{\mu \in \textsc{Ball}_t} \mu \cdot x$
  2. Observe the reward r_t and update $\textsc{Ball}_{t+1}$.

LinUCB Regret Bound

Sublinear regret: $R_T \leq O^\star(d\sqrt{T})$
poly dependence on d , no dependence on the cardinality $|D|$.

Theorem (Dani, Hayes, K. '09)

Suppose: bounded noise $|\eta_t|$; $\|\mu^\star\| \leq W$; and $\|x\| \leq B$, for $x \in D$.
Set $\lambda = \sigma^2/W^2$ and $\beta_t := c_1\sigma^2\left(d\log\left(1+\frac{TB^2W^2}{d}\right)+\log(1/\delta)\right)$.
With probability greater than $1-\delta$, that for all $t \geq 0$,

$$R_T \leq c_2\sigma\sqrt{T}\left(d\log\left(1+\frac{TB^2W^2}{d\sigma^2}\right)+\log(4/\delta)\right)$$

where c_1, c_2 are absolute constants.

Outline

1. Linear Bandits
   - Setting
   - LinUCB and An Optimal Regret Bound

2. Analysis
   - Regret Analysis
   - Confidence Analysis

Two Key Lemma in the Proof

Lemma

(Confidence) Let $\delta > 0$*. We have that* $\Pr(\forall t,\ \mu^\star \in \text{BALL}_t) \geq 1 - \delta$*.*

Lemma

(Sum of Squares Regret Bound) Define:

$$\text{regret}_t = \mu^\star \cdot x^* - \mu^\star \cdot x_t$$

Suppose $\beta_t \geq 1$ *and* β_t *is increasing; and* $\mu^\star \in \text{BALL}_t$ *for all* t*. Then*

$$\sum_{t=0}^{T-1} \text{regret}_t^2 \leq 4\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Completing the Proof

Proof:[Proof of Theorem 1] With the two previous Lemmas, along with the Cauchy-Schwarz inequality, we have, with probability at least $1-\delta$,

$$R_T = \sum_{t=0}^{T-1} \mathrm{regret}_t \le \sqrt{T \sum_{t=0}^{T-1} \mathrm{regret}_t^2} \le \sqrt{4T\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)}.$$

The remainder of the proof follows from our chosen value of β_T. ■

Outline

1 Linear Bandits
- Setting
- LinUCB and An Optimal Regret Bound

2 Analysis
- Regret Analysis
- Confidence Analysis

Lemma

Let $x \in D$. If $\mu \in \textsc{Ball}_t$ and $x \in D$. Then

$$|(\mu - \widehat{\mu}_t)^\top x| \le \sqrt{\beta_t x^\top \Sigma_t^{-1} x}$$

Proof: Triangle ineq. + def of $\textsc{Ball}_t$

By Cauchy-Schwarz, we have:

$$|(\mu - \widehat{\mu}_t)^\top x| = |(\mu - \widehat{\mu}_t)^\top \Sigma_t^{1/2} \Sigma_t^{-1/2} x| = |(\Sigma_t^{1/2}(\mu - \widehat{\mu}_t))^\top \Sigma_t^{-1/2} x|$$

$$\le \|\Sigma_t^{1/2}(\mu - \widehat{\mu}_t)\| \|\Sigma_t^{-1/2} x\| = \|\Sigma_t^{1/2}(\mu - \widehat{\mu}_t)\| \sqrt{x^\top \Sigma_t^{-1} x} \le \sqrt{\beta_t x^\top \Sigma_t^{-1} x}$$

where the last inequality holds since $\mu \in \textsc{Ball}_t$. ∎

Instantaneous Regret Lemma

Define

$$w_t := \sqrt{x_t^\top \Sigma_t^{-1} x_t}$$

which is the "normalized width" at time t in the direction of our decision.

Lemma

Fix $t \le T$. If $\mu^\star \in \textsc{Ball}_t$, then

$$\text{regret}_t \le 2 \min\left(\sqrt{\beta_t} w_t, 1\right) \le 2\sqrt{\beta_T} \min\left(w_t, 1\right)$$

Proof: Due to"optimism".
Let $\widetilde{\mu} \in \textsc{Ball}_t$ denote the vector which minimizes the dot product $\widetilde{\mu}^\top x_t$.
By choice of x_t, $\widetilde{\mu}^\top x_t = \max_{\mu \in \textsc{Ball}_t} \max_{x \in D} \mu^\top x \ge (\mu^\star)^\top x^*$, where the inequality used the hypothesis $\mu^\star \in \textsc{Ball}_t$. Hence,

$$\begin{aligned}\text{regret}_t &= (\mu^\star)^\top x^* - (\mu^\star)^\top x_t \le (\widetilde{\mu} - \mu^\star)^\top x_t \\ &= (\widetilde{\mu} - \widehat{\mu}_t)^\top x_t + (\widehat{\mu}_t - \mu^\star)^\top x_t \le 2\sqrt{\beta_t} w_t\end{aligned}$$

where the last step follows from the "width" Lemmasince $\widetilde{\mu}$ and $\mu^\star$ are

Geometric Argument: Part 1

The next two lemmas give us 'geometric' potential function argument, where can bound the sum of widths independently of the choices made by the algorithm.

Lemma

We have:

$$\det \Sigma_T = \det \Sigma_0 \prod_{t=0}^{T-1} (1 + w_t^2).$$

Proof: By the definition of Σ_{t+1}, we have

$$\begin{aligned} \det \Sigma_{t+1} &= \det(\Sigma_t + x_t x_t^\top) = \det(\Sigma_t^{1/2}(I + \Sigma_t^{-1/2} x_t x_t^\top \Sigma_t^{-1/2})\Sigma_t^{1/2}) \\ &= \det(\Sigma_t)\det(I + \Sigma_t^{-1/2} x_t (\Sigma_t^{-1/2} x_t)^\top) = \det(\Sigma_t)\det(I + v_t v_t^\top), \end{aligned}$$

where $v_t := \Sigma_t^{-1/2} x_t$. Now observe that $v_t^\top v_t = w_t^2$ and ... ■

Geometric Argument: Part 2

Lemma

For any sequence $x_0, \dots x_{T-1}$ such that, for $t < T$, $\|x_t\|_2 \le B$, we have:

$$\log\left(\det \Sigma_{T-1} / \det \Sigma_0\right) = \log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) \le d \log\left(1 + \frac{TB^2}{d\lambda}\right).$$

Proof: Denote the eigenvalues of $\sum_{t=0}^{T-1} x_t x_t^\top$ as $\sigma_1, \dots \sigma_d$, and note:

$$\sum_{i=1}^{d} \sigma_i = \mathrm{Trace}\left(\sum_{t=0}^{T-1} x_t x_t^\top\right) = \sum_{t=0}^{T-1} \|x_t\|^2 \le TB^2.$$

Using the AM-GM inequality,

$$\log\det\left(I + \frac{1}{\lambda}\sum_{t=0}^{T-1} x_t x_t^\top\right) = \log\left(\prod_{i=1}^{d}(1 + \sigma_i/\lambda)\right)$$

$$= d\log\left(\prod_{i=1}^{d}\left(1 + \sigma_i/\lambda\right)\right)^{1/d} \le d\log\left(\frac{1}{d}\sum_{i=1}^{d}(1 + \sigma_i/\lambda)\right) \le d\log\left(1 + \frac{TB^2}{d\lambda}\right)$$

Proving "sum of squares regret" Lemma

Proof: Assume $\mu^\star \in \textsc{Ball}_t$ for all t. We have:

$$\sum_{t=0}^{T-1} \text{regret}_t^2 \leq \sum_{t=0}^{T-1} 4\beta_t \min(w_t^2, 1) \leq 4\beta_T \sum_{t=0}^{T-1} \min(w_t^2, 1)$$
$$\leq 4\beta_T \sum_{t=0}^{T-1} \ln(1 + w_t^2) \leq 4\beta_T \log\left(\det \Sigma_{T-1} / \det \Sigma_0\right)$$
$$= 4\beta_T d \log\left(1 + \frac{TB^2}{d\lambda}\right)$$

where the first inequality follow from by Lemma 5; the second from that β_t is an increasing function of t; the third uses that for $0 \leq y \leq 1$, $\ln(1 + y) \geq y/2$; the final two inequalities follow by Lemmas 6 and 7. ■

Outline

1. Linear Bandits
   - Setting
   - LinUCB and An Optimal Regret Bound
2. Analysis
   - Regret Analysis
   - Confidence Analysis

Self-Normalizing Sum

Lemma (Self-Normalized Bound for Vector-Valued Martingales)

(Abassi et. al '11) Suppose $\{\varepsilon_i\}_{i=1}^{\infty}$ *are mean zero random variables (can be generalized to martingales), and* ε_i *is bounded by* σ*. Let* $\{X_i\}_{i=1}^{\infty}$ *be a stochastic process. Define* $\Sigma_t = \Sigma_0 + \sum_{i=1}^{t} X_i X_i^{\top}$*. With probability at least* $1 - \delta$*, we have for all* $t \geq 1$*:*

$$\left\| \sum_{i=1}^{t} X_i \varepsilon_i \right\|_{\Sigma_t^{-1}}^{2} \leq \sigma^2 \log \left(\frac{\det(\Sigma_t) \det(\Sigma_0)^{-1}}{\delta^2} \right).$$

(This is a general version of the Self-Normalized Sum argument in [Dani, Hayes, K. '09]).

Confidence [Proof of Lemma 2]

Proof: Since $r_\tau = x_\tau \cdot \mu^\star + \eta_\tau$, we have:

$$\widehat{\mu}_t - \mu^\star = \Sigma_t^{-1} \sum_{\tau=0}^{t-1} r_\tau x_\tau - \mu^\star = \Sigma_t^{-1} \sum_{\tau=0}^{t-1} x_\tau (x_\tau \cdot \mu^\star + \eta_\tau) - \mu^\star$$

$$= \Sigma_t^{-1} \left(\sum_{\tau=0}^{t-1} x_\tau (x_\tau)^\top \right) \mu^\star - \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau$$

$$= \lambda \Sigma_t^{-1} \mu^\star + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau$$

By the triangle inequality,

$$\sqrt{(\widehat{\mu}_t - \mu^\star)^\top \Sigma_t (\widehat{\mu}_t - \mu^\star)} \le \left\| \lambda \Sigma_t^{-1/2} \mu^\star \right\| + \left\| \Sigma_t^{-1/2} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \right\|$$

$$\le \sqrt{\lambda} \|\mu^\star\| \qquad + \qquad ??.$$

How can we bound "??" To be continued... ■

Continued... [Proof of Lemma 2]

Proof:

$$(\widehat{\mu}_t - \mu^\star)^\top \Sigma_t (\widehat{\mu}_t - \mu^\star) \leq \left\| \lambda \Sigma_t^{-1/2} \mu^\star \right\| + \left\| \Sigma_t^{-1/2} \sum_{\tau=0}^{t-1} \eta_\tau x_\tau \right\|$$
$$\leq \sqrt{\lambda} \|\mu^\star\| + \sqrt{2\sigma^2 \log\left(\det(\Sigma_t)\det(\Sigma^0)^{-1}/\delta_t\right)}.$$

We seek to lower bound $\Pr(\forall t,\ \mu^\star \in \textsc{Ball}_t)$. Assign failure probability $\delta_t = (3/\pi^2)/t^2$ for the t-th event, which gives us:

$$1 - \Pr(\forall t,\ \mu^\star \in \textsc{Ball}_t) = \Pr(\exists t,\ \mu^\star \notin \textsc{Ball}_t) \leq \sum_{t=1}^{\infty} \Pr(\mu^\star \notin \textsc{Ball}_t)$$
$$< \sum_{t=1}^{\infty} (1/t^2)(3/\pi^2) = 1/2.$$

This along with Lemma 7 completes the proof. ∎