

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 $x_t \in D \subset \mathbb{R}^d$.

Handling Large Actions Spaces

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

$$\begin{aligned}\mathbb{E}[r_t | x_t = x] &= \mu^* \cdot x \in [-1, 1], \\ &= \mu^* \cdot \tilde{\phi}(x)\end{aligned}$$

) on. μ^* - unknown weight vector.

Handling Large Actions Spaces

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

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

$$r_t = \mu^* \cdot x + \eta_t$$

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

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

is i.i.d noise.

model due to Abe & Long '99

Our Objective

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

$$R_T = \left(\mu^* \cdot x^* \right) - \sum_{t=0}^{T-1} \mu^* \cdot x_t \quad \text{in expectation}$$

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

we get

$$x^* \in \operatorname{argmax}_{x \in D} \mu^* \cdot x$$

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 BALL_t : with center, $\hat{\mu}_t$, and shape, Σ_t , using the λ -regularized least squares solution:

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

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

$\text{Ball}_t = \{ \vec{\mu} \mid \text{BALL}_t = \{ (\hat{\mu}_t - \vec{\mu})^{\top} \Sigma_t^{-1} (\hat{\mu}_t - \vec{\mu}) \leq \beta_t \}$

*posterior
cont.
 $\vec{\mu} \cdot x - \hat{\mu}_t \cdot x$*

where β_t is a parameter of the algorithm.

LinUCB & The “Confidence Ball”

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

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

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

$\text{Ball}_t = \{ \mu \mid \text{BALL}_t \cap \{ (\hat{\mu}_t - \mu^*)^{\top} \Sigma_t^{-1} (\hat{\mu}_t - \mu^*) \leq \beta_t \} \}$, $\text{Ball}_0 = \{ \mu \mid \lambda \| \mu \|_2^2 \leq \beta_0 \}$.

where β_t is a parameter of the algorithm.

- LinUCB: For $t = 0, 1, \dots$

- Execute $x_t = \operatorname{argmax}_{x \in D} \max_{\mu \in \text{BALL}_t} \mu \cdot x$
- Observe the reward r_t and update BALL_{t+1} .

LinUCB Regret Bound

Sublinear regret: $R_T \leq O^*(d\sqrt{T})$

poly dependence on d , no dependence on the cardinality $|D|$.

LinUCB Regret Bound

Sublinear regret: $R_T \leq O^*(d\sqrt{T})$

poly dependence on d , no dependence on the cardinality $|D|$.

Theorem (Dani, Hayes, K. '09)

Suppose: bounded noise $|\eta_t|; \|\mu^*\| \leq W$; 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 $t \geq 0$,

$$R_T \leq O(d\sqrt{T})$$

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

where c_1, c_2 are absolute constants.

$d\sqrt{T}$ is optimal

for bounded rewards.

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 $\Pr(\forall t, \mu^* \in \text{BALL}_t) \geq 1 - \delta$.

Two Key Lemma in the Proof

Lemma

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

Lemma

(*Sum of Squares Regret Bound*) Define:

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

think of
 $\beta_t = O(d \log T)$

Suppose $\beta_t \geq 1$ and β_t is increasing; and $\mu^* \in \text{BALL}_t$ 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 $1 - \delta$,

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

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

“Width” of Confidence Ball

poinwise confidence

Lemma

Let $x \in D$. If $\mu \in \text{BALL}_t$ and $x \in D$. Then

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

“Width” of Confidence Ball

Lemma

Let $x \in D$. If $\mu \in \text{BALL}_t$ and $x \in D$. Then

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

Proof: Triangle ineq. + def of BALL_t

“Width” of Confidence Ball

Lemma

Let $x \in D$. If $\mu \in \text{BALL}_t$ and $x \in D$. Then

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

Proof: Triangle ineq. + def of BALL_t

By Cauchy-Schwarz, we have:

$$\begin{aligned} |(\mu - \hat{\mu}_t)^\top x| &= |(\mu - \hat{\mu}_t)^\top \Sigma_t^{1/2} \Sigma_t^{-1/2} x| = |(\Sigma_t^{1/2}(\mu - \hat{\mu}_t))^\top \Sigma_t^{-1/2} x| \\ &\leq \|\Sigma_t^{1/2}(\mu - \hat{\mu}_t)\| \|\Sigma_t^{-1/2} x\| = \|\Sigma_t^{1/2}(\mu - \hat{\mu}_t)\| \sqrt{x^\top \Sigma_t^{-1} x} \leq \sqrt{\beta_t x^\top \Sigma_t^{-1} x} \end{aligned}$$

where the last inequality holds since $\mu \in \text{BALL}_t$.

■

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.

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 $t \leq T$. If $\mu^* \in \text{BALL}_t$, then

$$\lesssim \sqrt{\beta_T} w_t$$

for

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

by VCB

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 $t \leq T$. If $\mu^ \in \text{BALL}_t$, then*

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

Proof: Due to “optimism”.

Instantaneous Regret Lemma

Define

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

$\Sigma_t = I + \sum_{z \in \mathcal{Z}} x_z x_z^\top$

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

Lemma

Fix $t \leq T$. If $\mu^* \in \text{BALL}_t$, then

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

Proof: Due to “optimism”.

Let $\tilde{\mu} \in \text{BALL}_t$ denote the vector which minimizes the dot product $\tilde{\mu}^\top x_t$.

By choice of x_t , $\tilde{\mu}^\top x_t = \max_{\mu \in \text{BALL}_t} \max_{x \in D} \mu^\top x \geq (\mu^*)^\top x^*$, where the inequality used the hypothesis $\mu^* \in \text{BALL}_t$. Hence,

$$\begin{aligned} \text{regret}_t &= (\mu^*)^\top x^* - (\mu^*)^\top x_t \leq (\tilde{\mu} - \mu^*)^\top x_t \\ &= (\tilde{\mu} - \hat{\mu}_t)^\top x_t + (\hat{\mu}_t - \mu^*)^\top x_t \leq 2\sqrt{\beta_t} w_t \end{aligned}$$

where the last step follows from the “width” Lemmas since $\tilde{\mu}$ and μ^* are

Geometric Argument: Part 1

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

Geometric Argument: Part 1

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

Lemma

We have:

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

Geometric Argument: Part 1

The next two lemmas give us 'geometric' potential function argument, where we 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 $t < T$, $\|x_t\|_2 \leq B$, we have:

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

Geometric Argument: Part 2

Lemma

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

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

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

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

Using the AM-GM inequality,

$$\begin{aligned} \log \det \left(I + \frac{1}{\lambda} \sum_{t=0}^{T-1} x_t x_t^\top \right) &= \log \left(\prod_{i=1}^d (1 + \sigma_i/\lambda) \right) \\ &= d \log \left(\prod_{i=1}^d (1 + \sigma_i/\lambda) \right)^{1/d} \leq d \log \left(\frac{1}{d} \sum_{i=1}^d (1 + \sigma_i/\lambda) \right) \leq d \log \left(1 + \frac{TB^2}{d\lambda} \right) \end{aligned}$$

App G
AM-GM

Proving “sum of squares regret” Lemma

Proof: Assume $\mu^* \in \text{BALL}_t$ for all t . We have:

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

where the first inequality follows from 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 $r_\tau = \mathbf{x}_\tau \cdot \mu^* + \eta_\tau$, we have:

$$\begin{aligned}\hat{\mu}_t - \mu^* &= \Sigma_t^{-1} \sum_{\tau=0}^{t-1} r_\tau \mathbf{x}_\tau - \mu^* = \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \mathbf{x}_\tau (\mathbf{x}_\tau \cdot \mu^* + \eta_\tau) - \mu^* \\ &= \Sigma_t^{-1} \left(\sum_{\tau=0}^{t-1} \mathbf{x}_\tau (\mathbf{x}_\tau)^\top \right) \mu^* - \mu^* + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau \mathbf{x}_\tau \\ &= \lambda \Sigma_t^{-1} \mu^* + \Sigma_t^{-1} \sum_{\tau=0}^{t-1} \eta_\tau \mathbf{x}_\tau\end{aligned}$$

By the triangle inequality,

$$\begin{aligned}\sqrt{(\hat{\mu}_t - \mu^*)^\top \Sigma_t (\hat{\mu}_t - \mu^*)} &\leq \left\| \lambda \Sigma_t^{-1/2} \mu^* \right\| + \left\| \Sigma_t^{-1/2} \sum_{\tau=0}^{t-1} \eta_\tau \mathbf{x}_\tau \right\| \\ &\leq \sqrt{\lambda} \|\mu^*\| + ??.\end{aligned}$$

How can we bound “??” To be continued... ■

Continued... [Proof of Lemma 2]

Proof:

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

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

$$\begin{aligned} 1 - \Pr(\forall t, \mu^* \in \text{BALL}_t) &= \Pr(\exists t, \mu^* \notin \text{BALL}_t) \leq \sum_{t=1}^{\infty} \Pr(\mu^* \notin \text{BALL}_t) \\ &< \sum_{t=1}^{\infty} (1/t^2)(3/\pi^2) = 1/2. \end{aligned}$$

This along with Lemma 7 completes the proof. ■