Thresholded Lasso Bandit

Kaito Ariu [1,2], Kenshi Abe [2], Alexandre Proutière [1]

[1] KTH Royal Institute of Technology

[2] CyberAgent, Inc.

Background

Linear contextual bandits have been applied in online services such as

- Online advertisement
- Recommendation systems
- Personalized medicine, etc.

Joint information about a user, ad image... are encoded in a context vector.

- Typically, only a few features are significant

Model and Objective

In each round $t \geq 1$:

(1) The decision-maker receives a set of feature vectors $\mathcal{A}_t$:

$$\mathcal{A}_t = \{A_{t,k} \in \mathbb{R}^d : k \in [K]\}$$

(2) She selects a vector $A_t \in \mathcal{A}_t$

(3) Observes a random reward with sub-Gaussian noise:

$$r_t = \langle A_t, \theta \rangle + \varepsilon_t$$

The high-dimensional parameter vector $\theta \in \mathbb{R}^d$ is fixed but unknown.

Sparsity. Assume that θ has at most s_0 non-zero components and $s_0 \ll d$.

Objective. Devise an algorithm with minimal regret, where regret is defined as:

$$R(T) := \mathbb{E}\left[\sum_{t=1}^{T} \max_{A \in \mathcal{A}_t} \langle A, \theta \rangle - r_t\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} \max_{A \in \mathcal{A}_t} \langle A - A_t, \theta \rangle\right].$$

Proposed Algorithm: Thresholded Lasso Bandit

In each round $t \geq 1$:

1. Receive feature vectors $\mathcal{A}_t = \{A_{t,k} \in \mathbb{R}^d : k \in [K]\}$
2. Greedily select arm $A_t = \arg\max_{A \in \mathcal{A}_t} \langle A, \hat{\theta}_t \rangle$
3. Obtain support estimate $\hat{S}$ by applying the two-step thresholding procedure to the Lasso estimate
4. Obtain new estimate $\hat{\theta}_{t+1}$ using OLS **only** on $\hat{S}$

Fewer parameters:

- Does not require the prior knowledge of s_0
- There is only one hyper-parameter λ_0
  - Can be even parameter-free when d is large enough

Regret Upper Bounds for Thresholded Lasso Bandit

Regret bounds with/without the margin condition (a probabilistic condition on the separation of the arm rewards) under some symmetric assumptions in Oh et al., 2021, etc.

Theorem. *Regret of Thresholded Lasso Bandit satisfies,*
(a) Under the margin condition,

$$R(T) = \mathcal{O}(\log d + \log T).$$

(b) Without the margin condition,

$$R(T) = \mathcal{O}(\log d + \sqrt{T}).$$

Previously, the regret bounds were $\mathcal{O}(\log d \log T)$ and $\mathcal{O}(\log d + \sqrt{T \log(dT)})$. Match the minimax lower bound.

Numerical Experiments