Model Selection for Contextual Bandits

Dylan Foster
MIT

Akshay Krishnamurthy
MSR NYC

Haipeng Luo
USC

Poster #5, Wednesday @ 5:00

Model Selection in Statistical Learning

Setup

- Data $\{(x_i, y_i)\}_{i=1}^n \sim D$
- Nested function classes $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_M$
- (Assume Bayes optimal predictor $f^\star \in \mathcal{F}_{m^\star}$)

Model Selection in Statistical Learning

Setup

- Data $\{(x_i, y_i)\}_{i=1}^{n} \sim D$
- Nested function classes $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_M$
- (Assume Bayes optimal predictor $f^\star \in \mathcal{F}_{m^\star}$)

Model selection guarantee: Learner $\hat{f}_n$ satisfies

$$R(\hat{f}_n) \leq R(f^\star) + \sqrt{\frac{\text{comp}(\mathcal{F}_{m^\star})}{n} \cdot \log(m^\star/\delta)}.$$

- Adapts to complexity of Bayes predictor $f^\star$!
- Algorithmic principle: Structural risk minimization **[Vapnik'92]**.

Model Selection in Statistical Learning

Setup

- Data $\{(x_i, y_i)\}_{i=1}^{n} \sim D$
- Nested function classes $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_M$
- (Assume Bayes optimal predictor $f^\star \in \mathcal{F}_{m^\star}$)

Model selection guarantee: Learner $\hat{f}_n$ satisfies

$$R(\hat{f}_n) \leq R(f^\star) + \sqrt{\frac{\text{comp}(\mathcal{F}_{m^\star})}{n} \cdot \log(m^\star/\delta)}.$$

- Adapts to complexity of Bayes predictor $f^\star$!
- Algorithmic principle: Structural risk minimization [Vapnik'92].

Goal: Achieve similar guarantee in online learning with partial information (contextual bandits).

Is this even possible?

Is this even possible?

- **Statistical learning**: Structural risk minimization.

Is this even possible?

- **Statistical learning**: Structural risk minimization.
- **Online learning**: Exponential weights w/ prior, LR tuning.

Is this even possible?

- **Statistical learning**: Structural risk minimization.
- **Online learning**: Exponential weights w/ prior, LR tuning.
- **Contextual bandits**:
    - No positive results known:
        - Even for specific function classes.
        - Even if we're fine with, e.g., $T^{2/3}$-type rates.
    - Standard algorithmic approaches fail (cf. paper for details).

Is this even possible?

- **Statistical learning**: Structural risk minimization.
- **Online learning**: Exponential weights w/ prior, LR tuning.
- **Contextual bandits**:
  - No positive results known:
    - Even for specific function classes.
    - Even if we're fine with, e.g., $T^{2/3}$-type rates.
  - Standard algorithmic approaches fail (cf. paper for details).

Our result

Model selection for linear contextual bandits.

(Linear) Contextual Bandits

For $t = 1, \ldots, T$:

1. Observe $x_t \in \mathcal{X}$
2. Take action $a_t \in [K]$
3. Incur loss $\ell_t(a_t) \in [0, 1]$

$$\text{Regret}(T) = \sum_{t=1}^{T} \ell_t(a_t) - \sum_{t=1}^{T} \ell_t(\pi^\star(x_t))$$

(Linear) Contextual Bandits

For $t = 1, \ldots, T$:

1. Observe $x_t \in \mathcal{X}$
2. Take action $a_t \in [K]$
3. Incur loss $\ell_t(a_t) \in [0, 1]$

$$\text{Regret}(T) = \sum_{t=1}^{T} \ell_t(a_t) - \sum_{t=1}^{T} \ell_t(\pi^\star(x_t))$$

Linear setup:

Feature maps: $\{\phi_m\}_{m \in [M]}$, $\phi_m(x, a) \in \mathbb{R}^{d_m}$.

Realizability: $\exists \theta^\star \in \mathbb{R}^{d_{m^\star}}$, s.t. $\mathbb{E}[\ell(a) \mid x] = \left\langle \theta^\star, \phi_{m^\star}(x, a) \right\rangle$.

(Optimal policy is $\pi^\star(x_t) = \arg\max_a \left\langle \theta^\star, \phi_{m^\star}(x, a) \right\rangle$.)

(Linear) Contextual Bandits

For $t = 1, \dots, T$:

1. Observe $x_t \in \mathcal{X}$
2. Take action $a_t \in [K]$
3. Incur loss $\ell_t(a_t) \in [0, 1]$

$$\text{Regret}(T) = \sum_{t=1}^{T} \ell_t(a_t) - \sum_{t=1}^{T} \ell_t(\pi^\star(x_t))$$

Linear setup:

Feature maps: $\{\phi_m\}_{m \in [M]}$, $\phi_m(x, a) \in \mathbb{R}^{d_m}$.

Realizability: $\exists \theta^\star \in \mathbb{R}^{d_{m^\star}}$, s.t. $\mathbb{E}[\ell(a) \mid x] = \langle \theta^\star, \phi_{m^\star}(x, a) \rangle$.

(Optimal policy is $\pi^\star(x_t) = \arg\max_a \langle \theta^\star, \phi_{m^\star}(x, a) \rangle$.)

With $m^\star$ known, can get $\tilde{O}(\sqrt{d_{m^\star} T \log(K)})$ regret.

[ChuLiReyzinSchapire'11]

Our Result

Main Theorem

Without knowing $m^\star$, we get:

$$\text{Regret} \le \tilde{O}(T^{2/3}(Kd_{m^\star})^{1/3}).$$

We can also achieve:

$$\text{Regret} \le \tilde{O}(\sqrt{KTd_{m^\star}} + K^{1/4}T^{3/4}).$$

*Stochastic setting, some technical assumptions required (see paper).

Model selection possible whenever problem is learnable!

Key Idea

Estimate square loss gap between two classes ($d_i < d_j$)

$$\mathcal{E}_{i,j} := \mathbb{E}_{x,a}\left(\left\langle \theta_i^\star, \phi_i(x,a) \right\rangle - \left\langle \theta_j^\star, \phi_j(x,a) \right\rangle\right)^2$$

Key Idea

Estimate square loss gap between two classes ($d_i < d_j$)

$$\mathcal{E}_{i,j} := \mathbb{E}_{x,a}\left(\left\langle\theta_i^\star, \phi_i(x,a)\right\rangle - \left\langle\theta_j^\star, \phi_j(x,a)\right\rangle\right)^2$$

Plug-in estimator has error d_j/n.

Key Idea

Estimate square loss gap between two classes ($d_i < d_j$)

$$\mathcal{E}_{i,j} := \mathbb{E}_{x,a}\left(\left\langle \theta_i^\star, \phi_i(x,a)\right\rangle - \left\langle \theta_j^\star, \phi_j(x,a)\right\rangle\right)^2$$

Plug-in estimator has error d_j/n.

Lemma: New estimator with error $\sqrt{d_j}/n + d_j/m$.

- n exploration samples, m unlabeled samples.
- Refines and generalizes **[Dicker'14,KongValiant'18]**.

Key Idea

Estimate square loss gap between two classes ($d_i < d_j$)

$$\mathcal{E}_{i,j} := \mathbb{E}_{x,a}\left(\left\langle \theta_i^\star, \phi_i(x,a)\right\rangle - \left\langle \theta_j^\star, \phi_j(x,a)\right\rangle\right)^2$$

Plug-in estimator has error d_j/n.

Lemma: New estimator with error $\sqrt{d_j}/n + d_j/m$.

- n exploration samples, m unlabeled samples.
- Refines and generalizes **[Dicker'14,KongValiant'18]**.

Algorithm:
Run CB alg with d_i, mix in exploration, test if $\mathcal{E}_{i,j} > 0$, switch to d_j if so.

Key Idea

Estimate square loss gap between two classes ($d_i < d_j$)

$$\mathcal{E}_{i,j} := \mathbb{E}_{x,a}\left(\left\langle\theta_i^\star, \phi_i(x,a)\right\rangle - \left\langle\theta_j^\star, \phi_j(x,a)\right\rangle\right)^2$$

Plug-in estimator has error d_j/n.

Lemma: New estimator with error $\sqrt{d_j}/n + d_j/m$.

- n exploration samples, m unlabeled samples.
- Refines and generalizes **[Dicker'14,KongValiant'18]**.

Algorithm:
Run CB alg with d_i, mix in exploration, test if $\mathcal{E}_{i,j} > 0$, switch to d_j if so.

Note: Cannot run LinUCB, since d_i might not be realizable.

Summary

- First model selection guarantee for contextual bandits
- Key technique: fast rates for estimating best-in-class loss.
- Open problems:
  - Can we achieve similar model selection guarantees for general policy classes?
  - Can we achieve $\boxed{\sqrt{d_{m^\star} T}}$ for all $d_{m^\star}$?

Poster #5, Wednesday @ 5:00

arXiv:1906.00531