Online Model Selection

Parnian Kassraie, ETH Zurich

ETH zürich
Learning &
Adaptive Systems

Sequential Decision-Making & Bandits: Problem

At every step t

Choose actions $\boldsymbol{x}_t$

unknown reward

$$y_t = r(\boldsymbol{x}_t) + \varepsilon_t$$

obsv.noise

Receive feedback y_t

Repeat

Goal: Choose actions that give a high reward

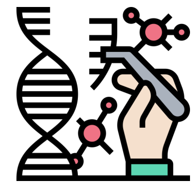
$$R(T) = \sum_{t=1}^{T} r(\boldsymbol{x}^\star) - r(\boldsymbol{x}_t)$$

Motivation: maximize r using the fewest queries

image source: flaticon

Sequential Decision-Making & Bandits: Solutions

To take actions at every step:



$$y_t = r(\boldsymbol{x}_t) + \varepsilon_t$$

- Estimate the reward function

based on:
Statistical model for the reward e.g. r is a linear function
history $H_{t-1} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_{t-1}, y_{t-1})\}$

- Use reward estimate to choose the next action

Many principles: optimism, expected improvement, entropy search

(better) estimate r
explore

maximize r
exploit

Heavily rely on the choice of model ⟶ Model selection is key!

Sequential Decision-Making & Bandits: Solutions

$y_t = r(\boldsymbol{x}_t) + \varepsilon_t$

To take actions at every step:

- Estimate the reward function

based on: Statistical model for the reward e.g. r is a linear function

history $H_{t-1} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_{t-1}, y_{t-1})\}$ samples are non-i.i.d

- Use reward estimate to choose the next action

(better) estimate r
explore

maximize r
exploit

samples are not so diverse

Model selection in this setting is not fun and games...

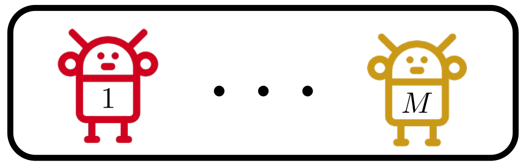
Open problem: when is (efficient) online model selection possible?

[Agarwal et al. 2017]

image source: flaticon

Blackbox Approaches

Model Selection as online search through a bag of algorithms.



Static Experts

Get M suggestions $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}$

Choose one $\boldsymbol{x}_t$

Be as good as:
best-in-hindsight expert

: Exponential Weights

$\log M$

[Haussler et al., 1998, Auer et al., 2002b, McMahan and Streeter 2009, Foster et al., 2017]

Adaptive Experts

Get M suggestions $\boldsymbol{x}_t^{(1)}, \ldots, \boldsymbol{x}_t^{(M)}$

Choose one $\boldsymbol{x}_t^{(j)}$

Update expert j

Be as good as: the global maxima

: A meta bandit algorithm

$\text{poly}(M)$

[Maillard and Munos, 2011, Agarwal et al., 2017, Pacchiano et al., 2020, Luo et al., 2022]

Problem Setting in this Talk

$$\boldsymbol{x}_t \in \mathcal{X} \subset \mathbb{R}^{d_0}$$
$$y_t = r(\boldsymbol{x}_t) + \varepsilon_t$$

i.i.d. zero-mean sub-gaussian noise

The reward is linearly parametrized by an unknown feature map

Model Class $\{\boldsymbol{\phi}_j : \mathbb{R}^{d_0} \to \mathbb{R}^d, \, j = 1, \dots, M\}$ $\quad M \gg T$

$\exists j^\star \in [M]$ s.t. $r(\cdot) = \boldsymbol{\theta}_{j^\star}^\top \boldsymbol{\phi}_{j^\star}(\cdot)$

+ typical bdd assump. $\|r\|_\infty \leq B$

Online Model Selection problem:

Find $j^\star$ while maximizing for the unknown r

$$R(T) = \sum_{t=1}^{T} r(\boldsymbol{x}^\star) - r(\boldsymbol{x}_t)$$

– Sublinear in T

– $\log M$

Warm-up Solution: Explore then Commit

For T_0 steps, take i.i.d. samples following a uniform, or "diverse" distribution

Use Group Lasso for implicit model selection

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M) \in \mathbb{R}^{dM}$$

$$\hat{\boldsymbol{\theta}} = \arg\min \frac{1}{T_0} \|\boldsymbol{y} - \Phi\boldsymbol{\theta}\|_2^2 + \lambda \sum_{j=1}^{M} \|\boldsymbol{\theta}_j\|_2$$

$$\boldsymbol{\phi}(\boldsymbol{x}) = (\boldsymbol{\phi}_1(\boldsymbol{x}), \dots, \boldsymbol{\phi}_M(\boldsymbol{x}))$$

For the remaining steps, always do

$$\boldsymbol{x}_t = \arg\max \hat{\boldsymbol{\theta}}^\top \boldsymbol{\phi}(\boldsymbol{x})$$

Under good choice of T_0 and λ satisfies,

$$R(T) = \mathcal{O}(\sqrt[3]{T^2 \log M}) \quad \text{w.h.p.}$$

(matches lower bound in certain action domains)

Warm-up Solution: Explore then Commit

For T_0 steps, take i.i.d. samples following a uniform, or "diverse" distribution

Incur high regret of $2BT_0$

Use the Lasso for implicit model selection

$$\hat{\boldsymbol{\theta}} = \arg\min \frac{1}{T_0} \|\boldsymbol{y} - \Phi\boldsymbol{\theta}\|_2^2 + \lambda \sum_{j=1}^{M} \|\boldsymbol{\theta}_j\|_2$$

Relies on Lasso variable selection

For the remaining steps, always do

$$\boldsymbol{x}_t = \arg\max \hat{\boldsymbol{\theta}}^\top \boldsymbol{\phi}(\boldsymbol{x})$$

Is not any-time: only works if horizon T is known in advance (doubling trick aside)

Under good choice of T_0 and λ satisfies,

$$R(T) = \mathcal{O}(\sqrt[3]{T^2 \log M}) \quad \text{w.h.p.}$$

(matches lower bound in certain action domains)

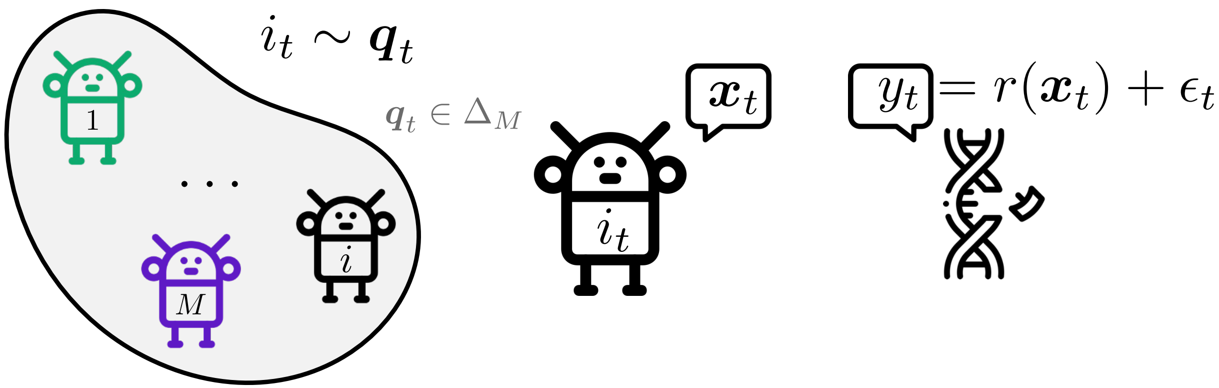
Online Model Selection

Instead of commiting to a single model,

Randomly iterate over the models and at each step choose one

Instatiate M "agents"

- Agent j only uses ϕ_j to model the reward
- Has action selection strategy $p_{t,j} \in \mathcal{M}(\mathcal{X})$ which is updated at every step e.g. UCB



Update $\boldsymbol{q}_t$

Update all agents

Requires having observed the reward for the choice of each agent

Reward not observed? Hallucinate it.

image source: flaticon

How to hallucinate rewards

Turn group lasso into a sparse online regression oracle

$$\forall t \geq 1 \qquad \hat{\boldsymbol{\theta}}_t = \arg\min \frac{1}{t} ||\boldsymbol{y}_t - \Phi_t \boldsymbol{\theta}||_2^2 + \lambda_t \sum_{j=1}^{M} ||\boldsymbol{\theta}_j||_2$$

Theorem (Anytime Lasso Conf Seq)

For appropriate choice of $(\lambda_t)_{t\geq 1}$,

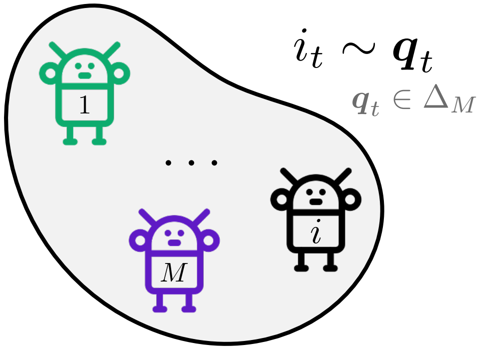
$$\mathbb{P}\left(\forall t \geq 1 : \left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\right\|_2 \lessapprox \sqrt{\frac{\log(M/\delta)}{t}}\right) \geq 1 - \delta$$

Hallucinate the reward of agent j as

$$\hat{r}_{t,j} = \mathbb{E}_{\boldsymbol{x}\sim p_{t,j}} \hat{\boldsymbol{\theta}}_t^\top \boldsymbol{\phi}(\boldsymbol{x})$$

$p_{t,j} \in \mathcal{M}(\mathcal{X})$ action selection strategy

How to iterate over agents



increase $q_{t,j}$ if $\hat{r}_{t,j}$ was high

Exponential Weighting

Estimate of the reward obtained by agent j so far

$$q_{t,j} = \frac{\exp(\eta_t \sum_{s=1}^{t-1} \hat{r}_{s,j})}{\sum_{i=1}^{M} \exp(\eta_t \sum_{s=1}^{t-1} \hat{r}_{s,i})}$$

sensitivity of updates

$$\hat{r}_{t,j} = \mathbb{E}_{\boldsymbol{x} \sim p_{t,j}} \hat{\boldsymbol{\theta}}_t^\top \boldsymbol{\phi}(\boldsymbol{x})$$

Putting it all together: ALExp

Find $j^\star$ while maximizing for the unknown r

Anytime **Exp**onential weighting algorithm with **L**asso reward estimates

Algorithm 1 ALEXP

Inputs: γ_t, η_t, λ_t for $t \geq 1$

for $t \geq 1$ **do**

Draw $\boldsymbol{x}_t \sim (1-\gamma_t)\sum_{j=1}^M q_{t,j} p_{t,j} + \gamma_t \text{Unif}(\mathcal{X})$

Observe $y_t = r(\boldsymbol{x}_t) + \epsilon_t$.

Append history $H_t = H_{t-1} \cup \{(\boldsymbol{x}_t, y_t)\}$.

Update agents $p_{t,j}$ for $j = 1, \dots, M$.

Calculate $\hat{\boldsymbol{\theta}}_t \leftarrow \text{Lasso}(H_t, \lambda_t)$ and estimate

$$\hat{r}_{t,j} \leftarrow \mathbb{E}_{\boldsymbol{x} \sim p_{t+1,j}}[\hat{\boldsymbol{\theta}}_t^\top \boldsymbol{\phi}(\boldsymbol{x})]$$

Update selection distribution

$$q_{t+1,j} \leftarrow \frac{\exp(\eta_t \sum_{s=1}^t \hat{r}_{s,j})}{\sum_{i=1}^M \exp(\eta_t \sum_{s=1}^t \hat{r}_{s,i})}$$

Theorem (Online Model Selection)

For appropriate choices of parameters, ALEXP with a UCB oracle agent satisfies

$$R(T) = \mathcal{O}\left(\sqrt{T \log^3 M} + T^{3/4}\sqrt{\log M}\right)$$

w.h.p. simultaneously for all $T \geq 1$.

[Open problem of Agarwal et al. 2017 in the Linear case]

Probably not tight? Lower bounds not clear.

A classic interpretation of ALExp [for bandit enthusiasts]

… is almost an Exp4 algorithm.

each expert is adaptive

as oppose to static experts with pre-set sequence of actions/advices

regression oracle is Lasso

as oppose to Importance Weighted Estimator or OLS

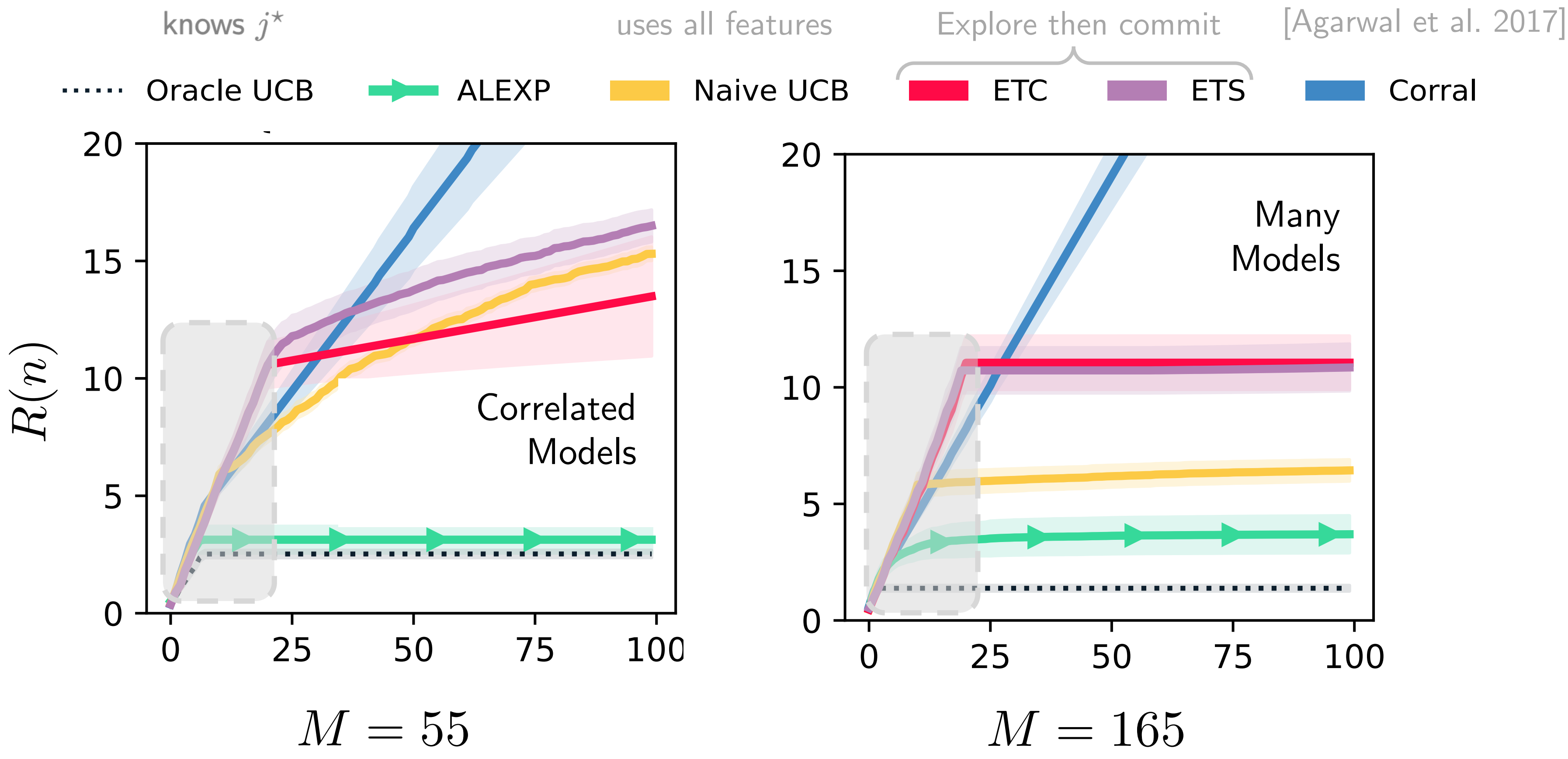
In this context, regret w.r.t. to agent j roughly is..

$$R(T, j) \leq \frac{\log M}{\eta} + \eta \sum_{t=1}^{T} \mathrm{Var}^2(\hat{\boldsymbol{\theta}}_t) + \sum_{t=1}^{T} \mathrm{Bias}(\hat{\boldsymbol{\theta}}_t)$$

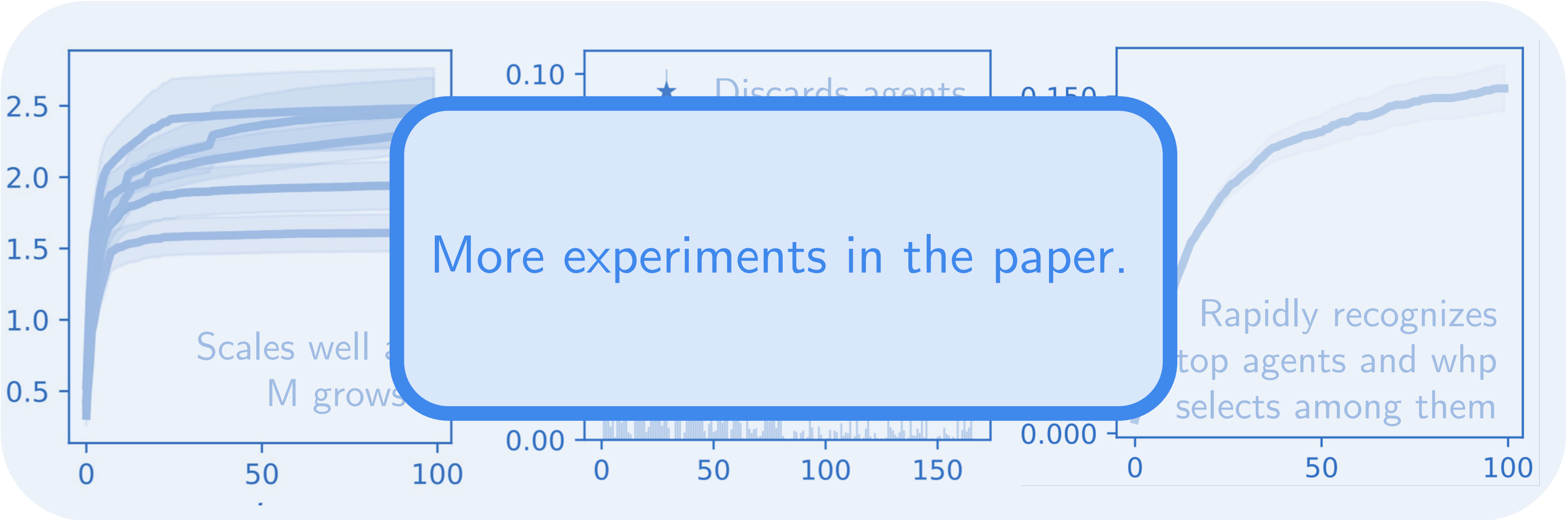
$\hat{\boldsymbol{\theta}}_t \in \mathbb{R}^{dM}$	$\mathrm{Var}^2(\hat{\boldsymbol{\theta}}_t)$	$\mathrm{Bias}(\hat{\boldsymbol{\theta}}_t)$
OLS/IW	M	0
Lasso	$\log M$	$\log M$

Model Selection for Optimistic algorithms

data generation & baselines
described in the paper.



If I am running out of time:



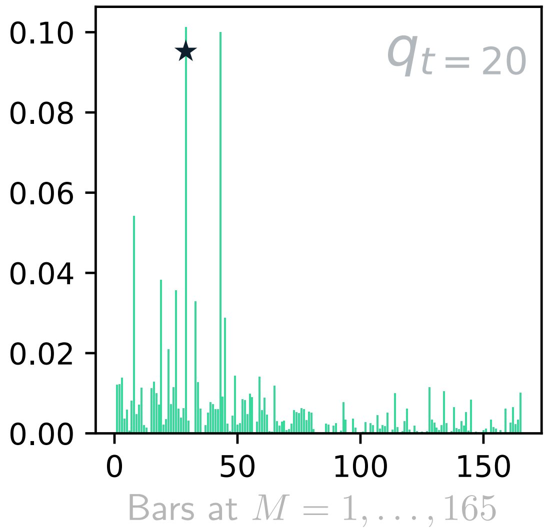
More experiments in the paper.

If not...

Model Selection Dynamics of ALExp

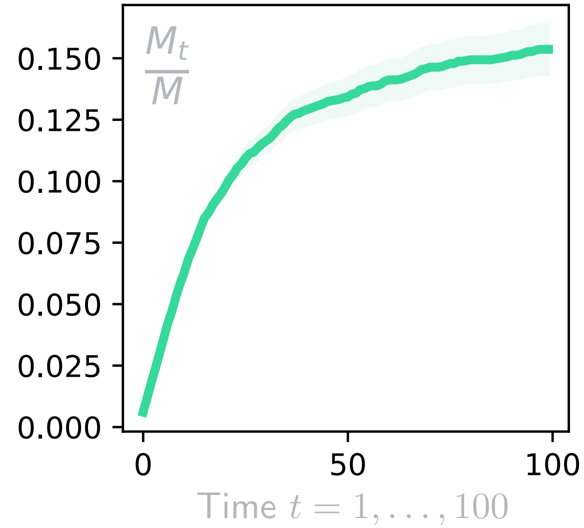
Let's see how things evolve turing training...

Distribution over the models
at time t=20



Discards agents without
having queried them

Number of visited agents
Total number of agents



Rapidly recognizes top agents
and whp selects among them

What's left open?

1. Is exploration necessary for model selection?

$$\gamma_t \sim t^{-1/4}$$

Algorithm 1 ALEXP

Inputs: γ_t, η_t, λ_t for $t \geq 1$
for $t \geq 1$ **do**
  Draw $\boldsymbol{x}_t \sim (1-\gamma_t)\sum_{j=1}^{M} q_{t,j} p_{t,j} + \gamma_t \text{Unif}(\mathcal{X})$
  Observe $y_t = r(\boldsymbol{x}_t) + \epsilon_t$.
  Append history $H_t = H_{t-1} \cup \{(\boldsymbol{x}_t, y_t)\}$.
  Update agents $p_{t,j}$ for $j = 1, \ldots, M$.
  Calculate $\hat{\boldsymbol{\theta}}_t \leftarrow \text{Lasso}(H_t, \lambda_t)$ and estimate rewards
  Update selection distribution

connected to lowerbounds on min-eigenvals of covariance matrix

some new results: *pure* exploration is not necessary.

What's left open?

1. Is exploration necessary for model selection?

2. For what other model classes (efficient) model selection is possible?

Linear ✔

$$\{\boldsymbol{\phi}_j : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^d,\, j = 1, \dots, M\}$$

$$\exists j^\star \in [M] \text{ s.t. } r(\cdot) = \boldsymbol{\theta}_{j^\star}^\top \boldsymbol{\phi}_{j^\star}(\cdot)$$

Blackbox Class of size M?

$\mathrm{Poly}(M)$ lower bound?

Infinite class with bounded eluder dimension?

$\log \tilde{d}$ upper bound?

Thank you!

Based on: "Anytime Model Selection in Linear Bandits." NeurIPS, 2023

Joint work with Nicolas Emmenegger, Andreas Krause, and Aldo Pacchiano