Anytime Model Selection for Linear Bandits

Parnian Kassraie, Nicolas Emmenegger, Andreas Krause, Aldo Pacchiano




ETH zürich

Anytime Model Selection

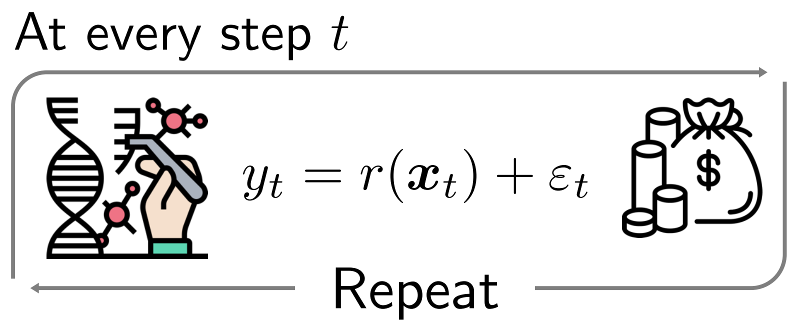


image source: flaticon

Anytime Model Selection

Solving a Linear Bandit problem :

1. Commit to a reward model (a priori)
2. Interact with the environment to maximize reward

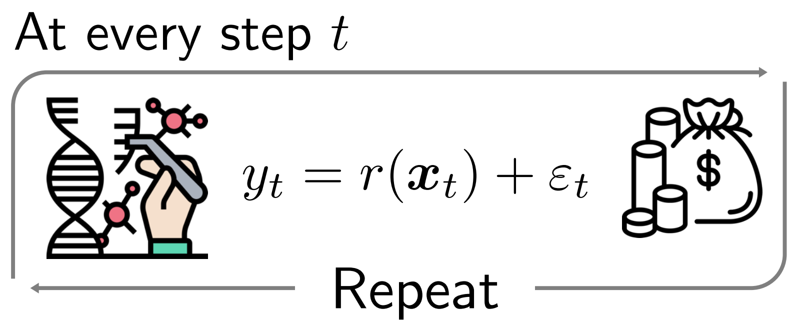


image source: flaticon

Anytime Model Selection

Solving a Linear Bandit problem :

1. Commit to a reward model (a priori)
2. Interact with the environment to maximize reward

There are many ways to model r

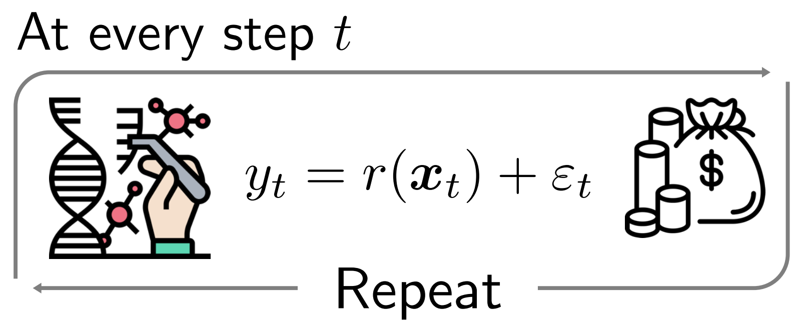


image source: flaticon

Anytime Model Selection

Solving a Linear Bandit problem :

1. Commit to a reward model (a priori)
2. Interact with the environment to maximize reward

There are many ways to model r

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

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

At every step t

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

Repeat

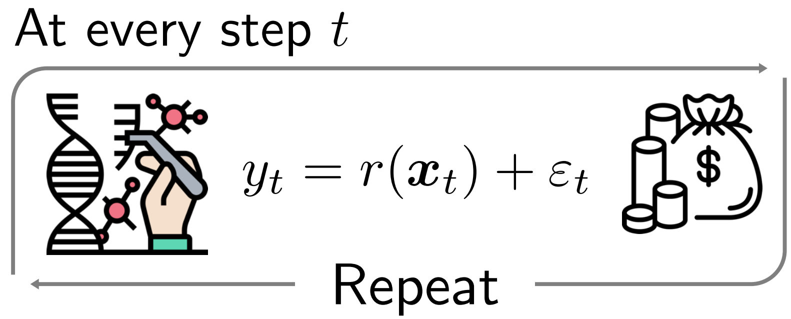
$M \gg T$ horizon/stopping time

image source: flaticon

Anytime Model Selection

Solving a Linear Bandit problem :

1. Commit to a reward model (a priori)
2. Interact with the environment to maximize reward



There are many ways to model r

$$\{\boldsymbol{\phi}_j : \mathbb{R}^{d_0} \to \mathbb{R}^d,\ j = 1, \dots, M\}$$
$$\exists j^\star \in [M] \text{ s.t. } r(\cdot) = \boldsymbol{\theta}_{j^\star}^\top \boldsymbol{\phi}_{j^\star}(\cdot)$$

$M \gg T$ horizon/stopping time

Not known a priori which model is going to yield the best algo.

... but we can guess based on emprical evidence.

image source: flaticon

Anytime Model Selection

Solving a Linear Bandit problem :

1. Commit to a reward model (a priori)
2. Interact with the environment to maximize reward

At every step t

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

Repeat

There are many ways to model r

$$\{\boldsymbol{\phi}_j : \mathbb{R}^{d_0} \to \mathbb{R}^d, \, j = 1, \ldots, M\}$$
$$\exists j^\star \in [M] \text{ s.t. } r(\cdot) = \boldsymbol{\theta}_{j^\star}^\top \boldsymbol{\phi}_{j^\star}(\cdot)$$

$M \gg T$ horizon/stopping time

Not known a priori which model is going to yield the best algo.

... but we can guess based on emprical evidence.

Anytime Model Selection problem

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

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

– Sublinear in T

– $\log M$

image source: flaticon

Online Model Selection problem

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

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

– Sublinear in T

– $\log M$

Online Model Selection problem

Find $j^\star$ … n r

$\forall T \geq 1$ … linear in T

$t=1$ … $-\log M$

Why do we need to select?
Why not just try out everything?

image source: flaticon

Online Model Selection problem

Find $j^\star$ … n r

$\forall T \geq 1$ … linear in T

$t=1$ … $- \log M$

Why do we need to select?
Why not just try out everything?

Instatiate M algorithms each using a different model

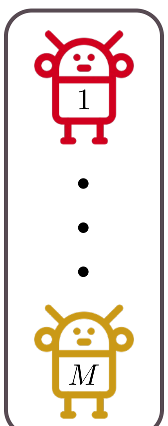


image source: flaticon

Online Model Selection problem

Find $j^\star$ … n r

$\forall T \geq 1$ … blinear in T

$t=1$ … $-\log M$

Why do we need to select?
Why not just try out everything?

Instatiate M algorithms each using a different model

Run all algorithms in parallel

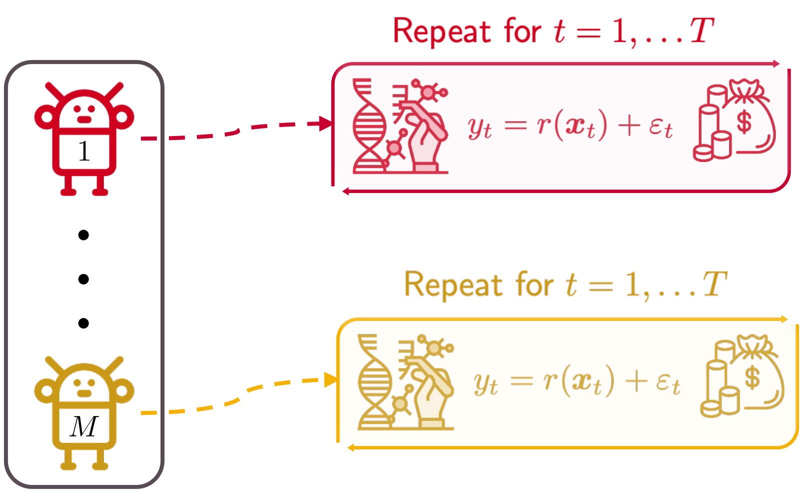


image source: flaticon

Online Model Selection problem

Find $j^\star$ … n r

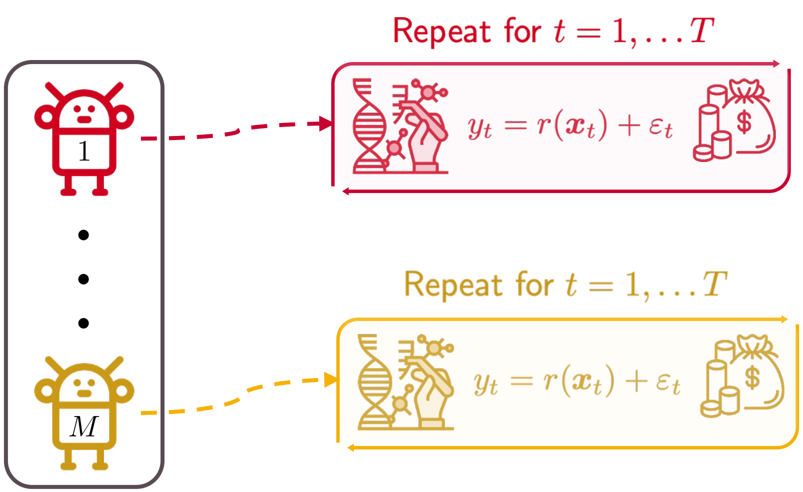
$\forall T \geq 1$ … linear in T

$t=1$ … $- \log M$

Why do we need to select?
Why not just try out everything?

Instatiate M algorithms each using a different model

Run all algorithms in parallel



Statistically expensive
$\leftrightarrow$ High regret

$\text{poly}(M)$

w.h.p.

image source: flaticon

Our Solution: Probabilistic Aggregation

- Randomly iterate over the agents and at each step play only one

Our Solution: Probabilistic Aggregation

Randomly iterate over the agents and at each step play only one

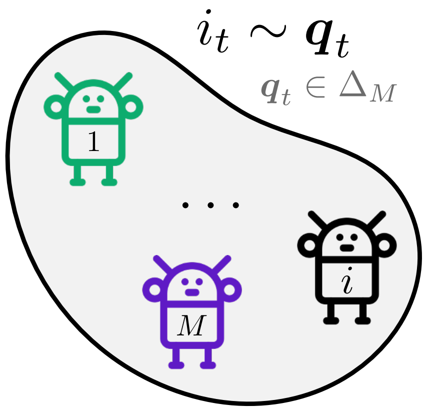


image source: flaticon

Our Solution: Probabilistic Aggregation

Randomly iterate over the agents and at each step play only one

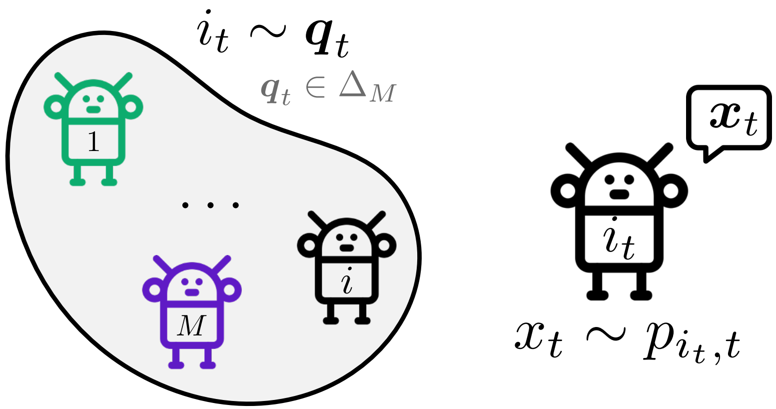


image source: flaticon

Our Solution: Probabilistic Aggregation

💡 Randomly iterate over the agents and at each step play only one

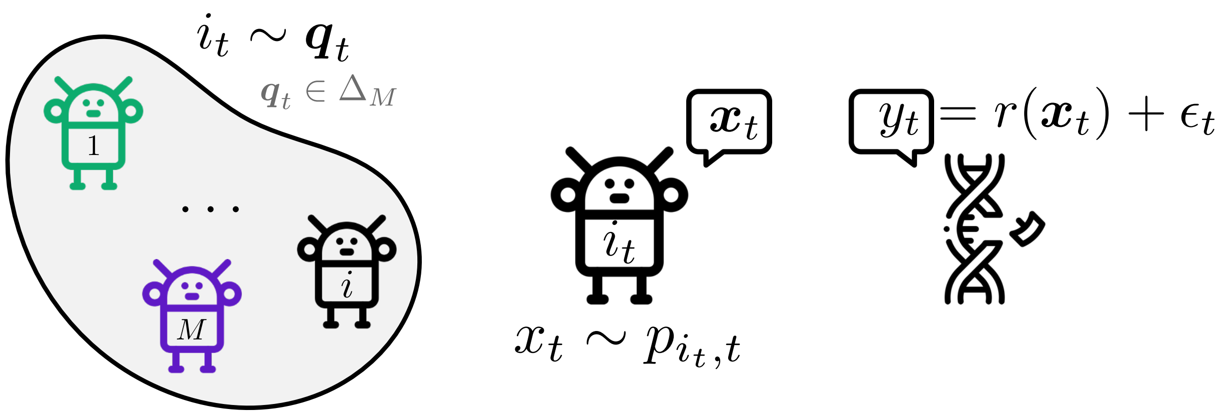
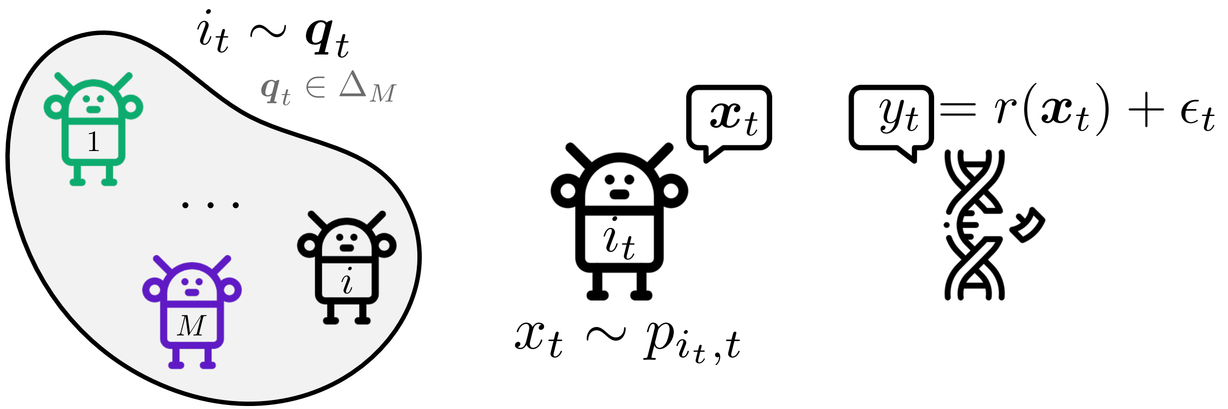


image source: flaticon

Our Solution: Probabilistic Aggregation

Randomly iterate over the agents and at each step play only one



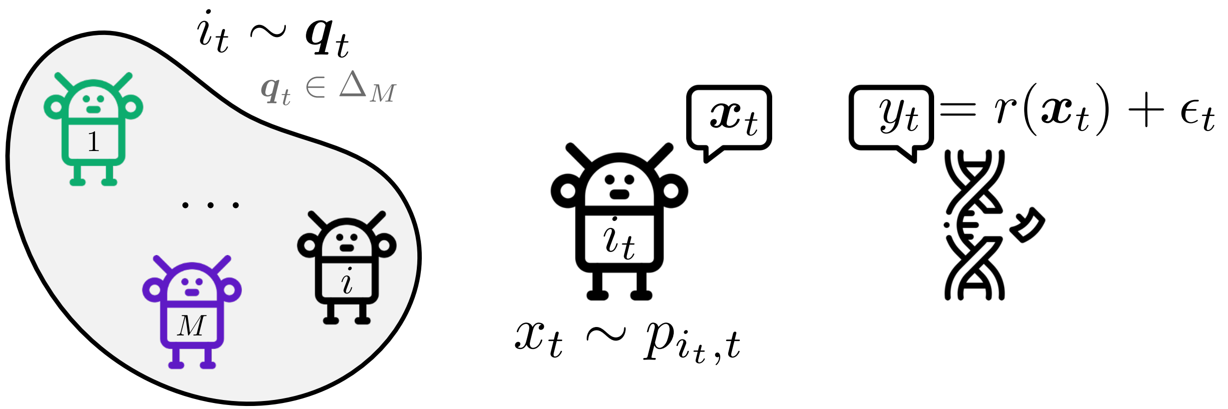
Update all agents

Update $\boldsymbol{q}_t$

image source: flaticon

Our Solution: Probabilistic Aggregation

Randomly iterate over the agents and at each step play only one



Update all agents

Update $\boldsymbol{q}_t$

Play one agent, but update all.
Reward not observed? Estimate it.

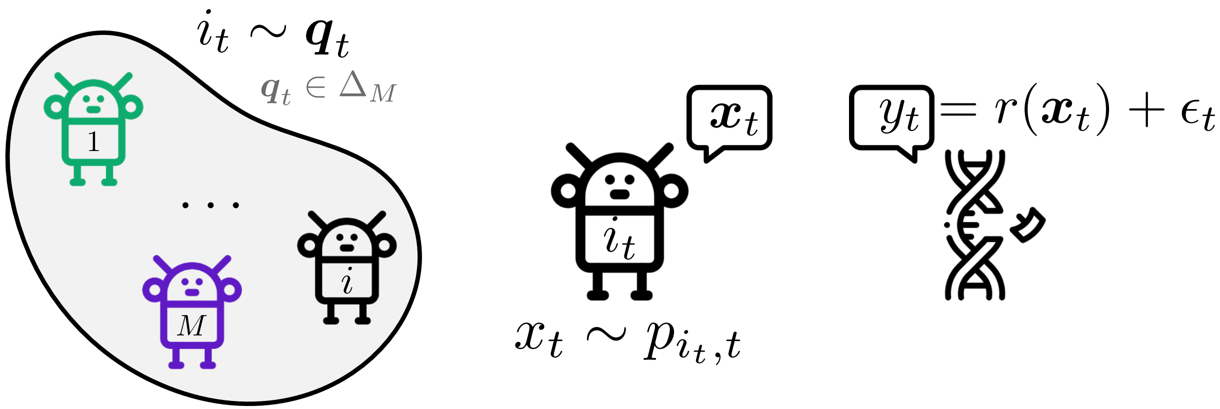
$\hat{r}_{t,j}$ for $j = 1, \ldots, M$

Choose your estimator **very** carefully!

image source: flaticon

Our Solution: Probabilistic Aggregation

Randomly iterate over the agents and at each step play only one



Update all agents

Update $\boldsymbol{q}_t$

Play one agent, but update all.
Reward not observed? Estimate it.

$\hat{r}_{t,j}$ for $j = 1, \ldots, M$

Choose your estimator **very** carefully!

Tune the probability of the agent.

$q_{t,j} \uparrow$ if $\hat{r}_{t,j} \uparrow$

Choose your update rule **very** carefully!

image source: flaticon

How to estimate and aggregate?

$\forall\, t \geq 1$

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

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

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

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

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

Theorem (Anytime Lasso Conf Seq)

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

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

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

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

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

Theorem (Anytime Lasso Conf Seq)

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

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

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

average reward of agent j

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

$$\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \ldots, \phi_M(\boldsymbol{x}))$$

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

Theorem (Anytime Lasso Conf Seq)

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

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

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

average reward of agent j

Exponential Weighting

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

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

$$\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \dots, \phi_M(\boldsymbol{x}))$$

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

Theorem (Anytime Lasso Conf Seq)

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

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

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

average reward of agent j

Exponential Weighting

estimate of the reward obtained by agent j so far

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

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

$$\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \dots, \phi_M(\boldsymbol{x}))$$

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

Theorem (Anytime Lasso Conf Seq)

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

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

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

average reward of agent j

Exponential Weighting

estimate of the reward obtained by agent j so far

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

sensitivity of updates

Putting it all together: ALExp

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 $j = 1, \ldots, M$.

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

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

Putting it all together: ALExp

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

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

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

Update selection distribution

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

prescribed in the paper

Theorem (Online Model Selection)

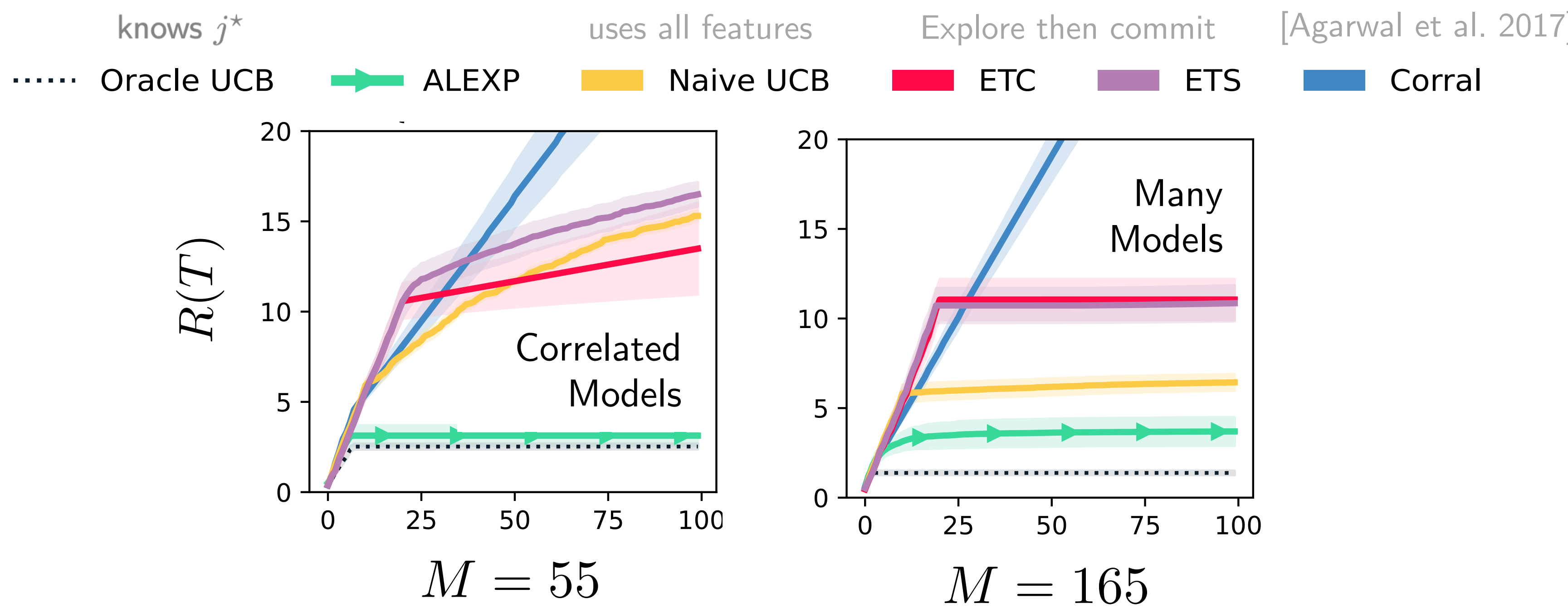
For appropriate choices of parameters,

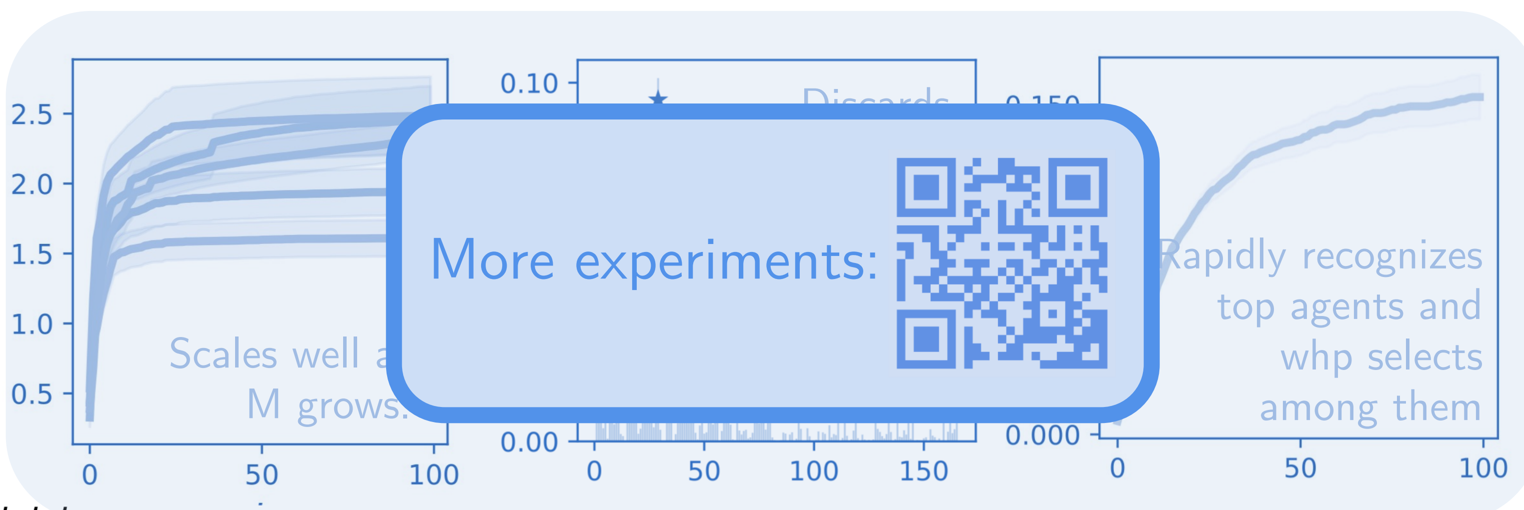
$$R(T) = \mathcal{O}\left(\sqrt{T \log^3 M} + T^{3/4}\sqrt{\log M}\right)$$

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

Synthetic Experiments

data generation & baselines described in the paper.





More experiments:

ETH zürich