Model Selection for Sequential Inference & Optimization

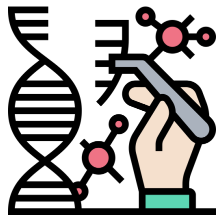
Parnian Kassraie, ETH Zurich

joint work with: Nicolas Emmenegger, Andreas Krause, Aldo Pacchiano

ETH zürich

Anytime Model Selection

At every step t



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

Repeat

image source: flaticon

Anytime Model Selection

The statistical modeling of the reward function plays a crucial role in efficiency of bandit algorithms: we choose actions based on reward estimates.

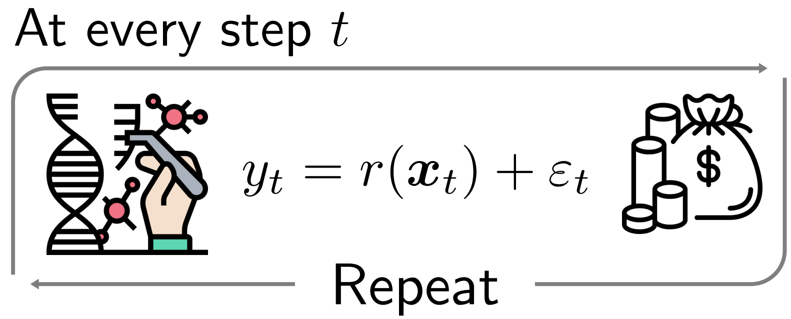
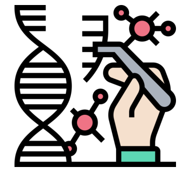


image source: flaticon

Anytime Model Selection

At every step t

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

Repeat

The statistical modeling of the reward function plays a crucial role in efficiency of bandit algorithms: we choose actions based on reward estimates.

There are many ways to model r

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

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

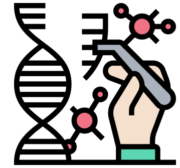
$M \gg T$ horizon/stopping time

image source: flaticon

Anytime Model Selection

The statistical modeling of the reward function plays a crucial role in efficiency of bandit algorithms: we choose actions based on reward estimates.

At every step t



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

Repeat

There are many ways to model r

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

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

$M \gg T$ 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

The statistical modeling of the reward function plays a crucial role in efficiency of bandit algorithms: we choose actions based on reward estimates.

At every step t

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

Repeat

There are many ways to model r

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

$M \gg T$ horizon/stopping time

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

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

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

Anytime Model Selection problem

Find $j^\star$ ... n r

$\forall T \geq 1$ ... blinear in T

$t=1$ ... $-\log M$

Do classical methods work here?
Is this harder than offline model selection?

Anytime Model Selection problem

Find $j^\star$ ... n r

$\forall T \geq 1$ ... blinear in T

$t=1$ ... $-\log M$

Do classical methods work here?
Is this harder than offline model selection?

Lower Data Quality

$$H_{t-1} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_{t-1}, y_{t-1})\}$$

Reward maximization → not so diverse sample

History dependence → non-i.i.d sample

Anytime Model Selection problem

Find $j^\star$ ... n r

$\forall T \geq 1$ ... blinear in T

$t=1$ ... $-\log M$

Do classical methods work here?
Is this harder than offline model selection?

Lower Data Quality

$$H_{t-1} = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_{t-1}, y_{t-1})\}$$

Reward maximization $\rightarrow$ not so diverse sample

History dependence $\rightarrow$ non-i.i.d sample

Stronger Guarantee

Valid for all stopping times

Cannot perform MS and then Inference/Opt post selection

Our Solution: Probabilistic Aggregation of Experts

Instantiate M algorithms each using a different ϕ_j to model the reward. Iterate over them.

$$\boxed{\forall\, t \geq 1}$$

image source: flaticon

Our Solution: Probabilistic Aggregation of Experts

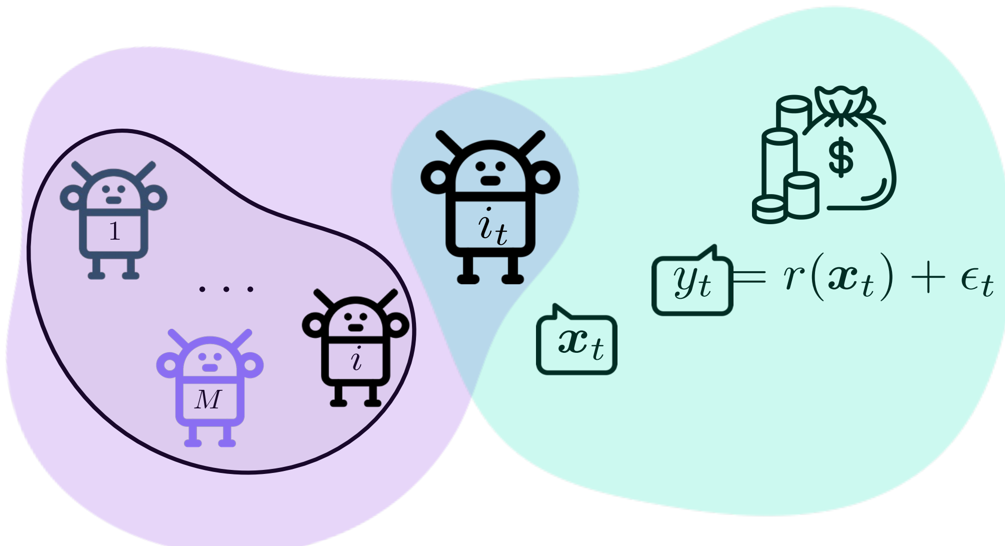
Instantiate M algorithms each using a different ϕ_j to model the reward. Iterate over them.

$\forall t \geq 1$

$i_t \sim \boldsymbol{q}_t$

$\boldsymbol{q}_t \in \Delta_M$

Model Selection



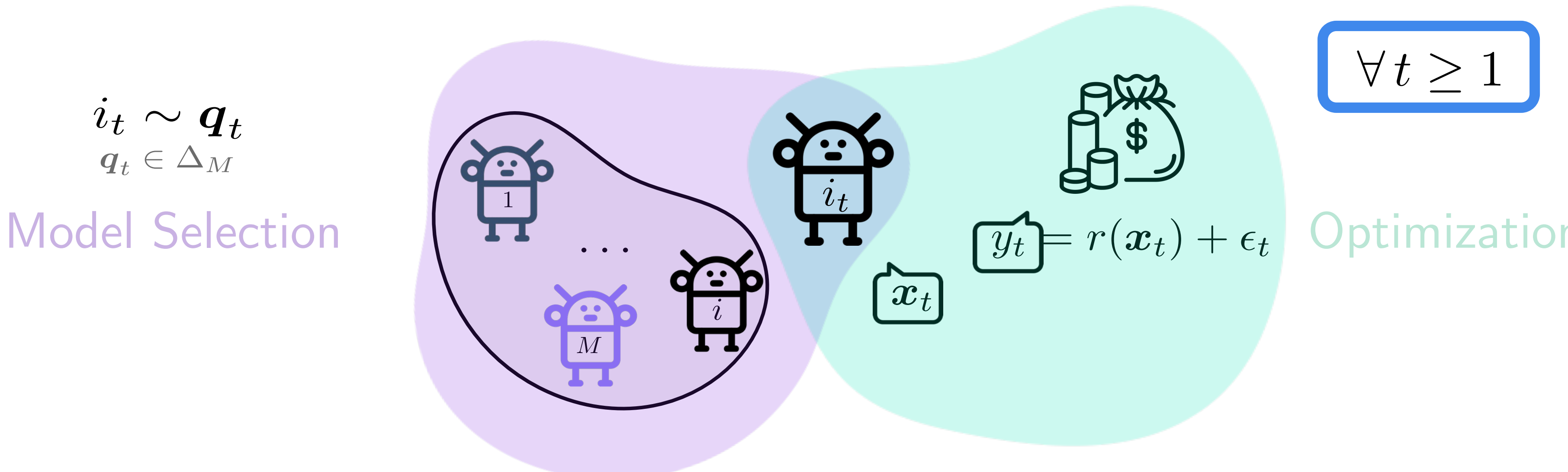
Optimization

With probability $q_{t,j}$ choose algorithm j and let them choose an action according to their action selection policy $p_{t,j} \in \mathcal{M}(\mathcal{X})$

image source: flaticon

Our Solution: Probabilistic Aggregation of Experts

Instantiate M algorithms each using a different ϕ_j to model the reward. Iterate over them.



With probability $q_{t,j}$ choose algorithm j and let them choose an action according to their action selection policy $p_{t,j} \in \mathcal{M}(\mathcal{X})$

Increase $q_{t,j}$ if the the algorithm *seems* to be lucrative

image source: flaticon

How to estimate and aggregate?

$$\forall\, t \geq 1$$

Turn lasso into a sparse online regression oracle

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

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

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

image source: flaticon

How to estimate and aggregate?

$\forall t \geq 1$

Turn lasso into a sparse online regression oracle

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

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

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

Theorem (Anytime Lasso Conf Seq)

If for all $t \geq 1$

$$\lambda_t \geq \frac{c_1}{\sqrt{t}} \sqrt{\log(M/\delta) + \sqrt{d\left(\log(M/\delta) + \boxed{(\log\log d)_+}\right)}}$$

cost of going 'time uniform'

then,

Restricted Eigenvalue property

$$\mathbb{P}\left(\forall t \geq 1 : \left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\right\|_2 \leq \frac{c_2 \lambda_t}{\kappa^2(\Phi_t, 2)}\right) \geq 1 - \delta$$

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

average reward of algo j

image source: flaticon

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)

If for all $t \geq 1$

$$\lambda_t \geq \frac{c_1}{\sqrt{t}} \sqrt{\log(M/\delta) + \sqrt{d\left(\log(M/\delta) + \boxed{(\log\log d)_+}\right)}}$$

cost of going 'time uniform'

then,

Restricted Eigenvalue property

$$\mathbb{P}\left(\forall t \geq 1: \left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\right\|_2 \leq \frac{c_2 \lambda_t}{\kappa^2(\Phi_t, 2)}\right) \geq 1 - \delta$$

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

average reward of algo j

Exponential Weighting

How lucrative algo j seems

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

image source: flaticon

ETH zürich

Result: ALEXP

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

Theorem (Regret - Informal)

For appropriate choices of parameters,

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

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

adaptive & anytime

✓

$\log M$ regret

✓

Result: ALEXP

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

Theorem (Regret - Informal)

For appropriate choices of parameters,

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

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

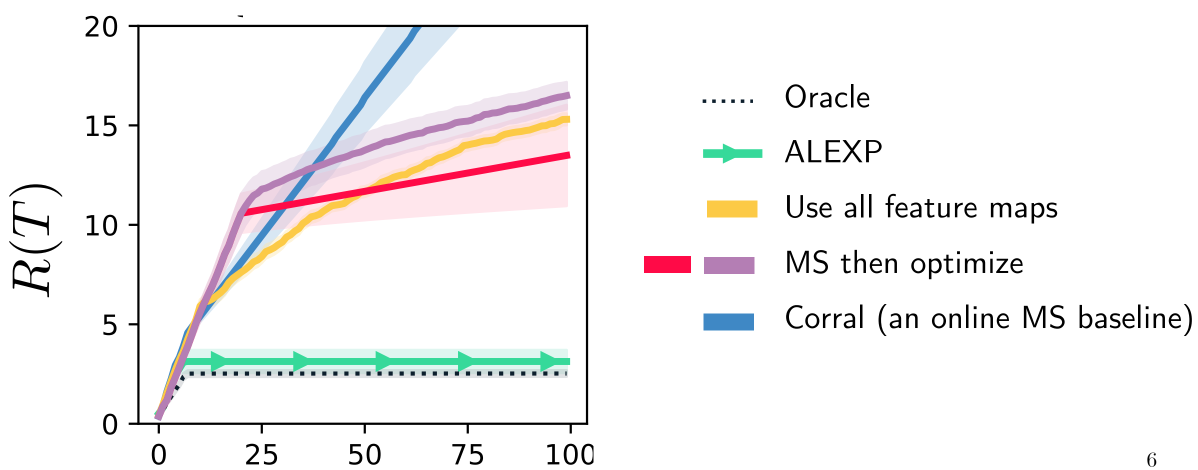
adaptive & anytime

✓

$\log M$ regret

✓

Works well even if the feature maps are correlated



Thank you!


Oracle UCB
ALEXP
Naive UCB
ETC
ETS
Corral
Scales well as
M grows.
Experiments:
Many
Models

Algorithm 1 ALEXP

Inputs: $\gamma_t,\ \eta_t,\ \lambda_t$ 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 $j = 1, \dots, M$.

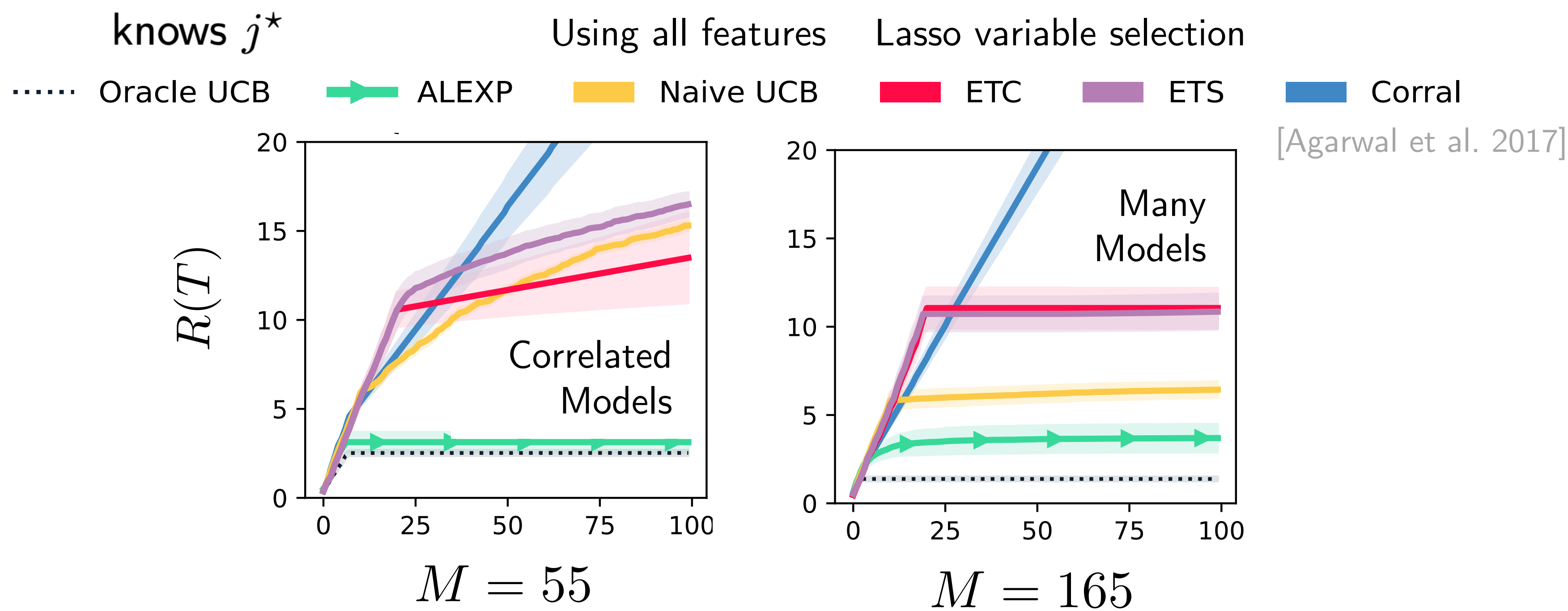
Calculate $\hat{\boldsymbol{\theta}}_t \leftarrow \mathrm{Lasso}(H_t, \lambda_t)$ and estimate

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

Update selection distribution

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

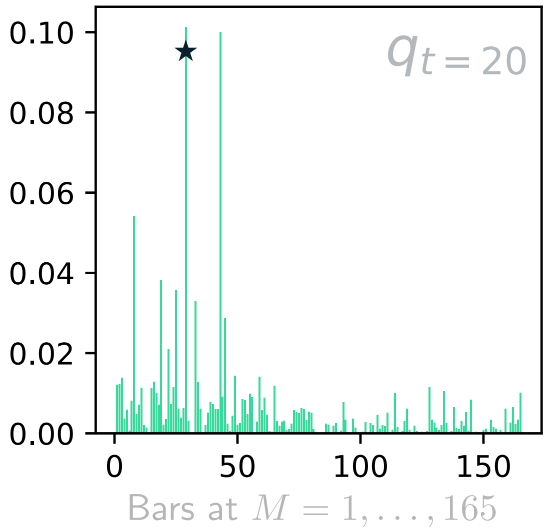
Synthetic Experiments



Model Selection Dynamics of ALExp

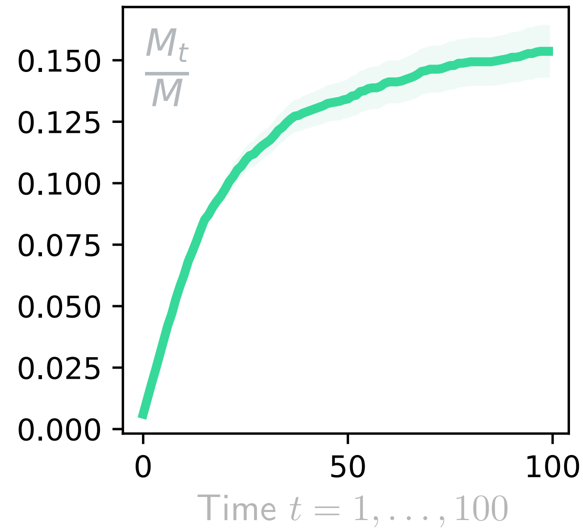
Let's see how things evolve turing training...

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

Online Model Selection problem

Find $j^\star$ n r

linear in T

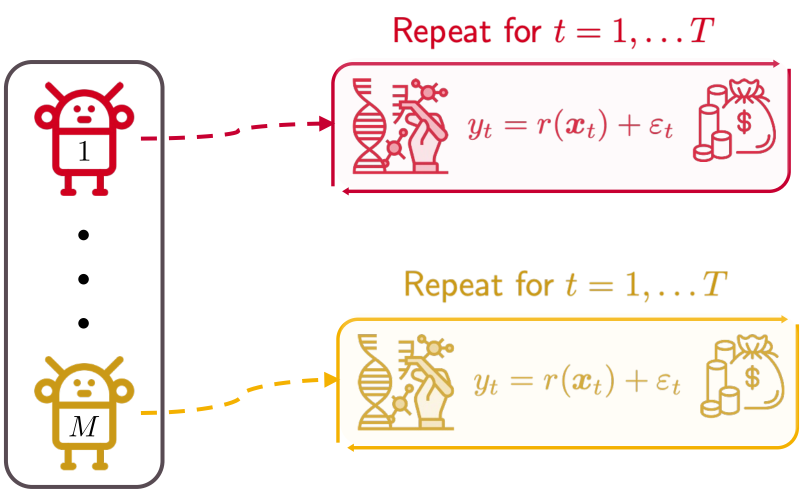
g M

$t=1$

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

$$\mathrm{poly}(M)$$

image source: flaticon

Classical 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,

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