Eluder Dimension and Potential Lemma

Presenter: Yingru Li

The Chinese University of Hong Kong, Shenzhen, China

March 13, 2021

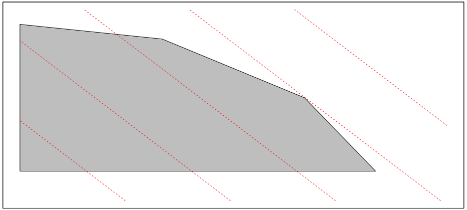
Mainly based on:
Russo, Daniel, and Benjamin Van Roy. "Learning to optimize via posterior sampling." Mathematics of Operations Research 39.4 (2014): 1221-1243.
Russo, Daniel, and Benjamin Van Roy. "Eluder Dimension and the Sample Complexity of Optimistic Exploration." NIPS. 2013.

Outline

Outline

Linear Bandit Problem

- Action space: $\mathcal{A}$
- Feature map: $\phi : \mathcal{A} \to \mathbb{R}^d$
- Mean reward of action $a \in \mathcal{A}$ is $\phi(a)^T\theta$
- $\theta \in \Theta \subset \mathbb{R}^d$ is unknown.
- Goal: Learn to solve $\max_{a\in\mathcal{A}} \phi(a)^T\theta$



Convergence to Optimality - Regret

- The agent can learn without exploring every possible action.
  The work of Dani et al. (2008), Rusmevichientong and Tsitsiklis (2010), and Abbasi-Yadkori et al. (2011) yields tight regret bounds of order

$$d\sqrt{T}$$

- Bounds exhibit no dependence on the number of actions
- What about more general model classes?

Convergence to Optimality - Regret

- The agent can learn without exploring every possible action.
  The work of Dani et al. (2008), Rusmevichientong and Tsitsiklis (2010), and Abbasi-Yadkori et al. (2011) yields tight regret bounds of order

$$d\sqrt{T}$$

- Bounds exhibit no dependence on the number of actions
- What about more general model classes?

A General Bandit Problem

- We want to solve
$$\max_{a \in \mathcal{A}} f_\theta(a)$$
- Know $f_\theta \in \mathcal{F} = \{f_p : \rho \in \Theta\}$
- Beliefs about $\theta \in \Theta$ may be encoded in terms of prior distribution.
- Agent sequentially chooses actions $A_1, A_2, \ldots$
- Choosing action A_t yields random reward with mean $f_\theta\left(A_t\right)$.

A General Bandit Problem

- Evaluate the performance up to time T by regret:

$$\text{Regret}(T) = \sum_{t=1}^{T} [\underbrace{f_\theta(A^*)}_{\text{optimal action}} - \underbrace{f_\theta(A_t)}_{\text{selected action}}]$$

Theoretical Gaurantees

Provide upper bounds on expected regret of Order up to some logarithmic factor

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

- Log-covering number:
  - Sensitivity to statistical over-fitting.
  - Closely related to concepts from statistical learning theory.
- Eluder dimension:
  - How does sampling one action reduce uncertainty about others?
  - How effectively the value of unobserved actions can be inferred from observed samples?
  - A new notion the paper introduce.

Theoretical Gaurantees

Provide upper bounds on expected regret of Order up to some logarithmic factor

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

- Log-covering number:
  - Sensitivity to statistical over-fitting.
  - Closely related to concepts from statistical learning theory.
- Eluder dimension:
  - How does sampling one action reduce uncertainty about others?
  - How effectively the value of unobserved actions can be inferred from observed samples?
  - A new notion the paper introduce.

Theoretical Gaurantees

Provide upper bounds on expected regret of Order up to some logarithmic factor

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

- Log-covering number:
  - Sensitivity to statistical over-fitting.
  - Closely related to concepts from statistical learning theory.
- Eluder dimension:
  - How does sampling one action reduce uncertainty about others?
  - How effectively the value of unobserved actions can be inferred from observed samples?
  - A new notion the paper introduce.

Theoretical Gaurantees

Provide upper bounds on expected regret of Order up to some logarithmic factor

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

- Bound holds for Thompson Sampling and a general UCB algorithm.
- Matches the best bounds available for UCB algorithms when specialized to linear or generalized linear models.

Theoretical Gaurantees

Provide upper bounds on expected regret of Order up to some logarithmic factor

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

- Bound holds for Thompson Sampling and a general UCB algorithm.
- Matches the best bounds available for UCB algorithms when specialized to linear or generalized linear models.

Outline

What about VC dimension?

- Define $S = \{x_1, \ldots, x_n\}$. Consider a binary function class $\mathcal{H}$ and the "projection" set

$$\mathcal{H}_S = \mathcal{H}_{x_1,\ldots,x_n} = \{(h(x_1), \ldots, h(x_n) : h \in \mathcal{H}\}$$

- **Growth Function**: The growth function is the maximum number of ways into which n points can be classified by the function class:

$$G_{\mathcal{H}}(n) = \sup_{x_1,\ldots,x_n} |\mathcal{H}_S|$$

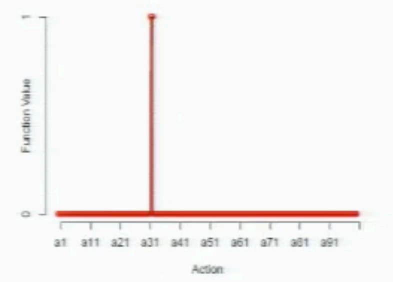
- **VC Dimension**:

$$\dim_{\mathrm{VC}}(\mathcal{H}) = \max\{n : G_{\mathcal{H}}(n) = 2^n\}$$

- VC dimension of a function class $\mathcal{H}$ is the cardinality of the largest set that it can shatter.

What about VC Dimension?

- $\mathcal{A} = \{a_1, \ldots, a_n\}$
- $\mathcal{F} = \{f_1, \ldots, f_n\}$
- $f_i(a) = \mathbb{1}[a = a_i]$

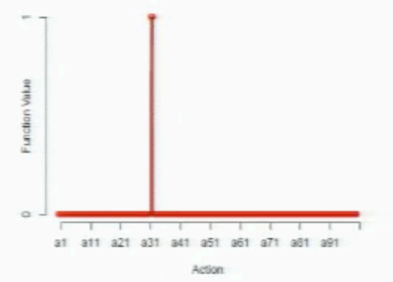


A noiseless prediction problem: Suppose A_t drawn uniformly from $\mathcal{A}$,

- $\dim_{\mathrm{VC}}(\mathcal{F}) = 1$
- Prediction error converges to $1/n$ in constant time.
  (e.g. predicting 0 or use f_1 every time.)

What about VC Dimension?

- $\mathcal{A} = \{a_1, \ldots, a_n\}$
- $\mathcal{F} = \{f_1, \ldots, f_n\}$
- $f_i(a) = \mathbb{1}[a = a_i]$



A multiarmed bandit problem: Suppose f_θ drawn uniformly from $\mathcal{F}$. then until the optimal action is identified, **Regret scales linearly with n**.

(a) Regret per round is 1

(b) At most a single function is ruled out per round

Defining Eluder Dimension - Intutitive explanation

- Elude (verb)
- evade or escape from (a danger, enemy, or pursuer), typically in a skillful or cunning way.
  "he managed to elude his pursuers by escaping into an alley"
- (of an idea or fact) fail to be grasped or remembered by (someone).
  "the logic of this eluded most people"

- A politician want to elude the reporters!
- The politician sequentially presents information to reporters.
- But each piece of information must be novel to the reporters.
- How long can he continue?

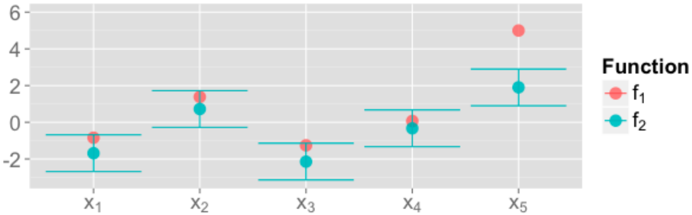
Defining Eluder Dimension - Intutitive explanation

- Elude (verb)
- evade or escape from (a danger, enemy, or pursuer), typically in a skillful or cunning way.
  "he managed to elude his pursuers by escaping into an alley"
- (of an idea or fact) fail to be grasped or remembered by (someone).
  "the logic of this eluded most people"

- A politician want to elude the reporters!
- The politician sequentially presents information to reporters.
- But each piece of information must be novel to the reporters.
- How long can he continue?

Defining Eluder Dimension - notion of (in)dependence

Eluder principle: An action a is independent of $\{a_1, \ldots, a_n\}$ if two functions that make similar predictions at $\{a_1, \ldots, a_n\}$ could differ significantly at a.



Defining Eluder Dimension - notion of (in)dependence

Definition 1 ($(\mathcal{F}, \epsilon)$-independence).

$a \in \mathcal{A}$ is ϵ-independent of $\{a_1, \ldots, a_n\} \subseteq \mathcal{A}$ with respect to $\mathcal{F}$ iff

- $\exists f, \tilde{f} \in \mathcal{F}$ satisfying

  (1) $\sqrt{\sum_{i=1}^{n} \left(f(a_i) - \tilde{f}(a_i)\right)^2} \leq \epsilon$

  satisfies $f(a) - \tilde{f}(a) > \epsilon$.

Definition 2 ($(\mathcal{F}, \epsilon)$-dependence).

$a \in \mathcal{A}$ is ϵ-dependent of $\{a_1, \ldots, a_n\} \subseteq \mathcal{A}$ with respect to $\mathcal{F}$ iff

- $\forall f, \tilde{f} \in \mathcal{F}$ satisfying

  (1) $\sqrt{\sum_{i=1}^{n} \left(f(a_i) - \tilde{f}(a_i)\right)^2} \leq \epsilon$

  satisfies $f(a) - \tilde{f}(a) \leq \epsilon$.

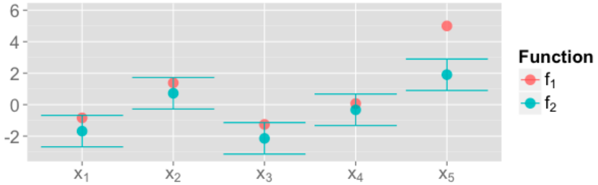


Figure: x_5 is $(\{f_1, f_2\}, 1)$-independent of $\{x_1, .., x_4\}$

Defining Eluder Dimension - notion of (in)dependence

- Let us get some understanding via the notion of linear dependence in linear algebra!
- **Claim**: $\left(\mathcal{F} := \left\{\langle\theta, \phi(\cdot)\rangle, \theta \in \mathbb{R}^d\right\}, 0\right)$-dependence $\iff$ linear dependence in $\mathbb{R}^d$.
- $a \in \mathcal{A}$ is 0-dependent of $\{a_1, \ldots, a_n\} \subseteq \mathcal{A}$ with respect to $\mathcal{F}$

$$
\begin{aligned}
&\iff \quad \forall \theta, \tilde{\theta} \in \mathbb{R}^d, \langle \theta - \tilde{\theta}, a_i \rangle = 0, \forall i \in [n] \Rightarrow \langle \theta - \tilde{\theta}, a \rangle = 0 \\
&\iff \quad \forall \theta \in \mathbb{R}^d, \langle \theta, a_i \rangle = 0, \forall i \in [n] \Rightarrow \langle \theta, a \rangle = 0 \\
&\iff \quad \forall \theta \in \mathbb{R}^d, \theta \in \mathrm{Span}(a_1, \ldots, a_n)^{\perp} \Rightarrow \langle \theta, a \rangle = 0 \\
&\iff \quad a \in \left(\mathrm{Span}(a_1, \ldots, a_n)^{\perp}\right)^{\perp} = \mathrm{Span}(a_1, \ldots, a_n)
\end{aligned}
$$

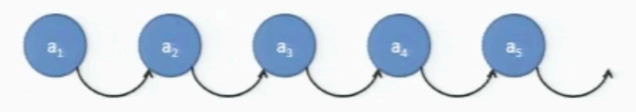
- $\iff a \in \mathcal{A}$ is linearly dependent of $\{a_1, \ldots, a_n\} \subseteq \mathcal{A}$.
- This ϵ-approximate extension is advantageous as it captures both nonlinear dependence and approximate dependence.

Defining Eluder Dimension

The eluder dimension is the length of the longest independent sequence.

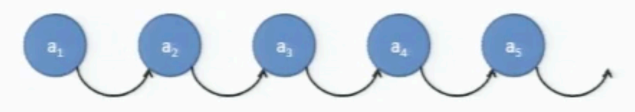
Definition 3 (Eluder dimension).

$\dim_E(\mathcal{F}, \epsilon)$ is the length of the **longest sequence of elements in $\mathcal{A}$** such that, for some $\epsilon' \geq \epsilon$, every element is $(\mathcal{F}, \epsilon')$-independent of its predecessors.



Eluder Dimension - Non-increasing in tolerance ϵ

- **Property.** $\dim_E(\mathcal{F}, \epsilon) \geq \dim_E(\mathcal{F}, \epsilon + \epsilon_0), \forall \epsilon_0 > 0.$
- **Proof** (My understanding):
  If for some $\epsilon' \geq \epsilon + \epsilon_0$, every element is $(\mathcal{F}, \epsilon')$-independent of its predecessors,
- then of course, for the above found ϵ', we have $\epsilon' \geq \epsilon + \epsilon_0 > \epsilon$, every element is $(\mathcal{F}, \epsilon')$-independent of its predecessors



- Therefore, conclude $\dim_E(\mathcal{F}, \epsilon)$ is at least the same as $\dim_E(\mathcal{F}, \epsilon + \epsilon_0)$.
- **Useful in the main proof.**

Understand Eluder Dim via comparison with VC Dim - Classical Def

- Define $S = \{x_1, \ldots, x_n\}$. Consider a binary function class $\mathcal{H}$ and the "projection" set

$$\mathcal{H}_S = \mathcal{H}_{x_1,\ldots,x_n} = \{(h(x_1), \ldots, h(x_n)) : h \in \mathcal{H}\}$$

- **Growth Function**: The growth function is the maximum number of ways into which n points can be classified by the function class:

$$G_{\mathcal{H}}(n) = \sup_{x_1,\ldots,x_n} |\mathcal{H}_S|$$

- **VC Dimension**:

$$\dim_{\mathrm{VC}}(\mathcal{H}) = \max\{n : G_{\mathcal{H}}(n) = 2^n\}$$

- VC dimension of a function class $\mathcal{H}$ is the cardinality of the largest set that it can shatter.

Understand Eluder Dim via comparison with VC Dim - New Def

Definition 4 (VC-independence).

An action a is VC-independent of $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ if for any $f, \tilde{f} \in \mathcal{F}$, there exists some $\bar{f} \in \mathcal{F}$, which agrees with f on a and with $\tilde{f}$ on $\tilde{\mathcal{A}}$; that is, $\bar{f}(a) = f(a)$ and $\bar{f}(\tilde{a}) = \tilde{f}(\tilde{a})$ for all $\tilde{a} \in \tilde{\mathcal{A}}$.

Definition 5 (VC-dependence).

An action a is VC-dependent of $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ if for any $\bar{f} \in \mathcal{F}$, there exists $f, \tilde{f} \in \mathcal{F}$, such that $\bar{f}$ cannot simultaneously agrees with f on a and with $\tilde{f}$ on $\tilde{\mathcal{A}}$; that is, $\bar{f}(a) \neq f(a)$ or $\bar{f}(\tilde{a}) \neq \tilde{f}(\tilde{a})$ for all $\tilde{a} \in \tilde{\mathcal{A}}$.

Remark 1.

By this definition, an action a is said to be VC-dependent on $\tilde{\mathcal{A}}$ if knowing the values $f \in \mathcal{F}$ takes on $\tilde{\mathcal{A}}$ could restrict the set of possible values at a.

Understand Eluder Dim via comparison with VC Dim - New Def

Definition 6 (Alternative definition of $\dim_{\mathrm{VC}}$).

The VC dimension of a class of binary-valued functions $\mathcal{H}$ with domain $\mathcal{A}$ is the largest cardinality of a set $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ such that every $a \in \tilde{\mathcal{A}}$ is VC-independent of $\tilde{\mathcal{A}} \backslash \{a\}$.

Remark 2 (Equivalence to classical definition of $\dim_{\mathrm{VC}}$ in binary output setting).

- *If $\mathcal{H}$ can shatter $\tilde{\mathcal{A}}$, it is trivial to see every $a \in \tilde{\mathcal{A}}$ is VC-independent of $\tilde{\mathcal{A}} \setminus \{a\}$.*
- *Conversely, we need to prove if every $a \in \tilde{\mathcal{A}}$ is VC-independent of $\tilde{\mathcal{A}} \setminus \{a\}$, then $\mathcal{H}$ can shatter $\tilde{\mathcal{A}}$.*

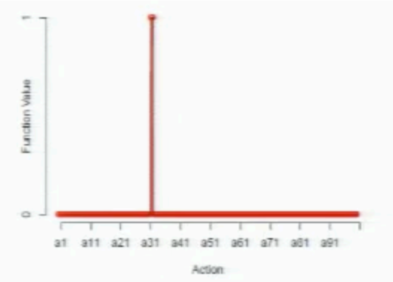
Understand Eluder Dim via comparison with VC Dim - New Def

Conversely, prove if every $a \in \tilde{\mathcal{A}}$ is VC-independent of $\tilde{\mathcal{A}} \setminus \{a\}$, then $\mathcal{H}$ can shatter $\tilde{\mathcal{A}}$.

- **My constructive proof**: Let us first assume there exists function $f_0, f_1 \in \mathcal{H}$ s.t. for all $a \in \tilde{\mathcal{A}}$, $f_1(a) = 1$ and $f_0(a) = 0$.
- W.l.o.g. let $\tilde{\mathcal{A}} = \{a_1, \cdots, a_n\}$.
- Pick $a = a_1, f = f_0, \tilde{f} = f_1$, by definition of VC-independence, there must exists $f_{01} \in \mathcal{H}$ s.t. $f_{01}(a_1) = f_0(a_1) = 0$ and $f_{01}(\tilde{\mathcal{A}} \setminus \{a_1\}) = f_1(\tilde{\mathcal{A}} \setminus \{a_1\}) = 1$.
- Similarly, pick $a = a_1, f = f_1, \tilde{f} = f_0$, there must exists $f_{10} \in \mathcal{H}$ s.t. $f_{10}(a_1) = 1$ and $f_{10}(\tilde{\mathcal{A}} \setminus \{a_1\}) = 0$. Now, a_1 is shattered.
- Recursively, pick $f \in \{f_0, f_1\}$ and $\tilde{f}_0 \in \{f_{01}, f_{10}, f_0, f_1\}$, and let $a = a_2 \in \tilde{\mathcal{A}} \setminus \{a_1\}$, there must exists $f_{001}, f_{101}, f_{110}, f_{010} \in \mathcal{H}$. Now, a_1 and a_2 is shattered.
- And finally we see $\mathcal{H}$ can shatter $\tilde{\mathcal{A}}$ by recursively find $f_{\{0,1\}^n} \in \mathcal{H}$, i.e. $\tilde{\mathcal{A}}$ is shattered.

Understand Eluder Dim via comparison with VC Dim

- $\mathcal{A} = \{a_1, \ldots, a_n\}$
- $\mathcal{F} = \{f_1, \ldots, f_n\}$
- $f_i(a) = \mathbf{1}\{a = a_i\}$



- In the above example, any two actions are VC dependent because knowing the label of one action could completely determine the value of the other action.
- However, this only happens if the sampled action has label 1.
- If it has label 0, one cannot infer anything about the value of the other action.

Understand Eluder Dim via comparison with VC Dim

- **stronger requirement**: guarantee one ~~could~~ will gain useful information through exploration.

Definition 7 (strong-dependence).

An action a is strongly dependent on a set of actions $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ if any two functions $f, \tilde{f} \in \mathcal{F}$ that agree on $\tilde{\mathcal{A}}$ agree on a; that is, the set $\{f(a) : f(\tilde{a}) = \tilde{f}(\tilde{a}), \forall \tilde{a} \in \tilde{\mathcal{A}}\}$ is a singleton. An action a is weakly independent of $\tilde{\mathcal{A}}$ if it is not strongly dependent on $\tilde{\mathcal{A}}$.

- a is strongly dependent on $\tilde{\mathcal{A}}$ if knowing the values of f on $\tilde{\mathcal{A}}$ completely determines the value of f on a.
- ϵ-Eluder dimension: Strong + ϵ-Approximate dependence
  - focusing on the possible difference $f(a) - \tilde{f}(a)$ between two functions that approximately agree on $\tilde{\mathcal{A}}$.

Understand Eluder Dim via comparison with VC Dim

- **stronger requirement**: guarantee one ~~could~~ will gain useful information through exploration.

Definition 7 (strong-dependence).

An action a is strongly dependent on a set of actions $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ if any two functions $f, \tilde{f} \in \mathcal{F}$ that agree on $\tilde{\mathcal{A}}$ agree on a; that is, the set $\{f(a) : f(\tilde{a}) = \tilde{f}(\tilde{a}), \forall \tilde{a} \in \tilde{\mathcal{A}}\}$ is a singleton. An action a is weakly independent of $\tilde{\mathcal{A}}$ if it is not strongly dependent on $\tilde{\mathcal{A}}$.

- a is strongly dependent on $\tilde{\mathcal{A}}$ if knowing the values of f on $\tilde{\mathcal{A}}$ completely determines the value of f on a.
- ϵ-Eluder dimension: Strong + ϵ-Approximate dependence
  - focusing on the possible difference $f(a) - \tilde{f}(a)$ between two functions that approximately agree on $\tilde{\mathcal{A}}$.

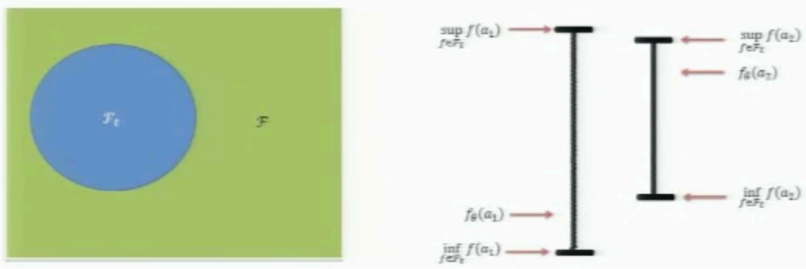
Outline

Outline

Optimism in the face of uncertainty

Act according to an "optimistic" model of the environment

- Confidence set $\mathcal{F}_t \leftarrow$ subset of $f \in \mathcal{F}$ that are statistically plausible given data.
- Play $\bar{A}_t \in \arg\max_{a \in \mathcal{A}} \left\{ \sup_{f \in \mathcal{F}_t} f(a) \right\}$.



Optimism in the face of uncertainty

Act according to an "optimistic" model of the environment

- $\mathcal{F}_t \leftarrow$ subset of $f \in \mathcal{F}$ that are statistically plausible given data.
- Play $\bar{A}_t \in \underset{a \in \mathcal{A}}{\arg\max} \left\{ \sup_{f \in \mathcal{F}_t} f(a) \right\}$.

There is a huge literature on this approach:

- Bandit problems with independent arms
  - (Lai-Robins, 1985), (Lai, 1987), (Auer, 2002), (Audibert, 2009)....
- Bandit problems with dependent arms
  - (Rusmevichientong-Tsitsiklis 2010), (Filippi et. al, 2010), (Srinivas et. al, 2012)...
- Reinforcement Learning
  - (Kearns-Singh, 2002), (Bartlett-Terwari, 2009), (Jaksch et. al 2010)...
- Monte Carlo Tree Search
  - (Kocsis-Szepesvári, 2006)...

A posterior sampling strategy

"Thompson sampling" & "probability matching":

- Sample each action according to the posterior probability it is optimal:

$$\pi_t = \mathbb{P}\left(A_t^* \in \cdot \mid H_t\right),$$

  where A_t^* is a random variable that satisfies $A_t^* \in \arg\max_{a \in \mathcal{A}} f_\theta(a)$.

- Practical implementations typically operate by, at each time t, sampling an index $\hat{\theta}_t \in \Theta$ from the distribution $\mathbb{P}\left(\theta \in \cdot \mid H_t\right)$ and then generating an action $A_t \in \arg\max_a f_{\hat{\theta}}(a)$.

The paper Learning to Optimize via Posterior Sampling

- establishes a close connection with optimistic algorithms.
- implies the analysis also bounds the Bayesian regret of TS.

Outline

Proof Sketch

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

(1) **Regret decomposition** for Optimism and Posterior Sampling;

- Upper bound regret by summation of confidence intervals at queried action sequence.

(2) Build generic **confidence sets** $\mathcal{F}_t \subset \mathcal{F}$;

- Size of $\mathcal{F}_t$ depends on the log-covering number of $\mathcal{F}$.

(3) **Key step**: Measure **the rate at which confidence intervals (bonus) shrink** $\Rightarrow$ **Regret rate**.

Proof Sketch

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

(1) **Regret decomposition** for Optimism and Posterior Sampling;

- Upper bound regret by summation of confidence intervals at queried action sequence.

(2) Build generic **confidence sets** $\mathcal{F}_t \subset \mathcal{F}$;

- Size of $\mathcal{F}_t$ depends on the log-covering number of $\mathcal{F}$.

(3) **Key step**: Measure **the rate at which confidence intervals (bonus) shrink** $\Rightarrow$ **Regret rate**.

Proof Sketch

$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}}\underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}}T}$$

(1) **Regret decomposition** for Optimism and Posterior Sampling;

- Upper bound regret by summation of confidence intervals at queried action sequence.

(2) Build generic **confidence sets** $\mathcal{F}_t \subset \mathcal{F}$;

- Size of $\mathcal{F}_t$ depends on the log-covering number of $\mathcal{F}$.

(3) **Key step**: Measure **the rate at which confidence intervals (bonus) shrink** $\Rightarrow$ **Regret rate**.

Proof Sketch

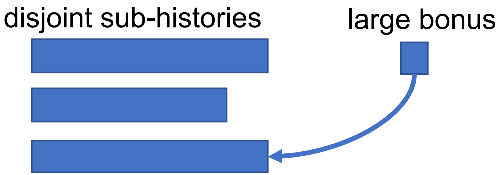
$$\sqrt{\underbrace{\dim_E\left(\mathcal{F}, T^{-1}\right)}_{\text{Eluder dimension}} \underbrace{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right)}_{\text{log-covering number}} T}$$

(1) **Regret decomposition** for Optimism and Posterior Sampling;

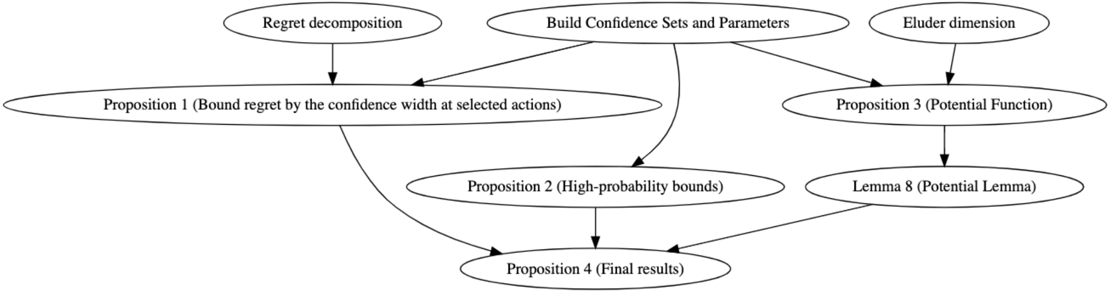
(2) Build generic **confidence sets** $\mathcal{F}_t \subset \mathcal{F}$;

(3) Measure **the rate at which confidence intervals (bonus) shrink**. (Potential lemma)

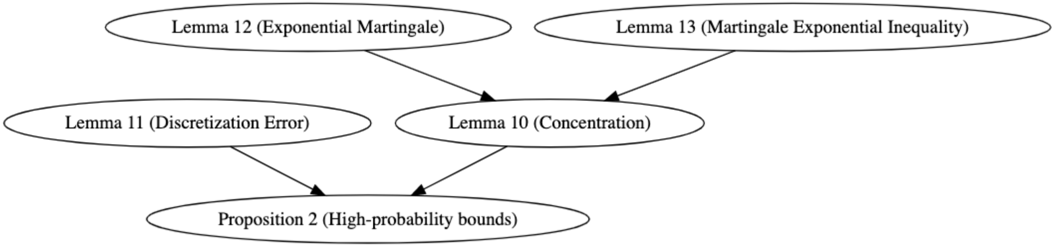
- If bonus at one action is large, then this action must be dependent on few ($\leq$) disjoint subsequences.
- Depends on the eluder dimension of $\mathcal{F}$. After some finite time, action should be dependent on at least ($\geq$) some disjoint subsequences.
- Therefore, bonus cannot be large forever.



Proof Sketch - Big Picture



Proof Sketch - Big Picture



- Moved to appendix: missing proofs

Regret decomposition

- UCB sequence $U = \{U_t \mid t \in \mathbb{N}\}$ adapted to filtration $\{\mathcal{H}_t \mid t \in \mathbb{N}\}$.
- **UCB regret decomposition**: Consider a UCB algorithm, $\bar{A}_t \in \arg\max_{a \in \mathcal{A}_t} U_t(a)$ and $A_t^* \in \arg\max_{a \in \mathcal{A}_t} f_\theta(a)$. We have the following simple regret decomposition:

$$\begin{aligned} f_\theta\left(A_t^*\right) - f_\theta\left(\bar{A}_t\right) &= f_\theta\left(A_t^*\right) - U_t\left(\bar{A}_t\right) + U_t\left(\bar{A}_t\right) - f_\theta\left(\bar{A}_t\right) \\ &\leq \left[f_\theta\left(A_t^*\right) - U_t\left(A_t^*\right)\right] + \left[U_t\left(\bar{A}_t\right) - f_\theta\left(\bar{A}_t\right)\right] \end{aligned}$$

Regret decomposition

- UCB sequence $U = \{U_t \mid t \in \mathbb{N}\}$ adapted to filtration $\{\sigma(H_t) \mid t \in \mathbb{N}\}$.
- **PS regret decomposition**: Consider a PS algorithm, conditioned on H_t, the optimal action A_t^* and the action A_t selected by posterior sampling are identically distributed, and U_t is deterministic.
  - Hence $\mathbb{E}\left[U_t\left(A_t^*\right) \mid H_t\right] = \mathbb{E}\left[U_t\left(A_t\right) \mid H_t\right]$.
  - And we have regret decomposition,

$$\begin{aligned}
\mathbb{E}\left[f_\theta\left(A_t^*\right) - f_\theta\left(A_t\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[f_\theta\left(A_t^*\right) - f_\theta\left(A_t\right) \mid H_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[U_t\left(A_t\right) - U_t\left(A_t^*\right) + f_\theta\left(A_t^*\right) - f_\theta\left(A_t\right) \mid H_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[U_t\left(A_t\right) - f_\theta\left(A_t\right) \mid H_t\right] + \mathbb{E}\left[f_\theta\left(A_t^*\right) - U_t\left(A_t^*\right) \mid H_t\right]\right] \\
&= \mathbb{E}\left[U_t\left(A_t\right) - f_\theta\left(A_t\right)\right] + \mathbb{E}\left[f_\theta\left(A_t^*\right) - U_t\left(A_t^*\right)\right]
\end{aligned}$$

Regret decoposition - Comparison

- Assume f_θ takes values in $[0, C]$. Compare decomposition for UCB and PS,

$$\text{Regret}\left(T, \pi^U, \theta\right) \overset{a.s.}{\leq} \sum_{t=1}^{T} \left[U_t\left(\bar{A}_t\right) - f_\theta\left(\bar{A}_t\right)\right] + C \sum_{t=1}^{T} \mathbb{1}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

$$\text{BayesRegret}\ \left(T, \pi^{PS}\right) \leq \mathbb{E} \sum_{t=1}^{T} \left[U_t\left(A_t\right) - f_\theta\left(A_t\right)\right] + C \sum_{t=1}^{T} \mathbb{P}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

- Important difference: the regret bound of π^U depends on the specific UCB sequence U used by the UCB algorithm in question,
- whereas the bound of π^{PS} applies simultaneously for all UCB sequences.

Regret decoposition - Comparison

- Assume f_θ takes values in $[0, C]$. Compare decomposition for UCB and PS,

$$\text{BayesRegret}\left(T, \pi^{U}\right) \leq \mathbb{E}\sum_{t=1}^{T}\left[U_t\left(\bar{A}_t\right) - f_\theta\left(\bar{A}_t\right)\right] + C\sum_{t=1}^{T}\mathbb{P}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

$$\text{BayesRegret}\left(T, \pi^{PS}\right) \leq \mathbb{E}\sum_{t=1}^{T}\left[U_t\left(A_t\right) - f_\theta\left(A_t\right)\right] + C\sum_{t=1}^{T}\mathbb{P}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

- While the Bayesian regret of a UCB algorithm depends critically on the specific choice of confidence sets,
- posterior sampling depends on the best-possible choice of confidence sets.

Regret decoposition - Comparison

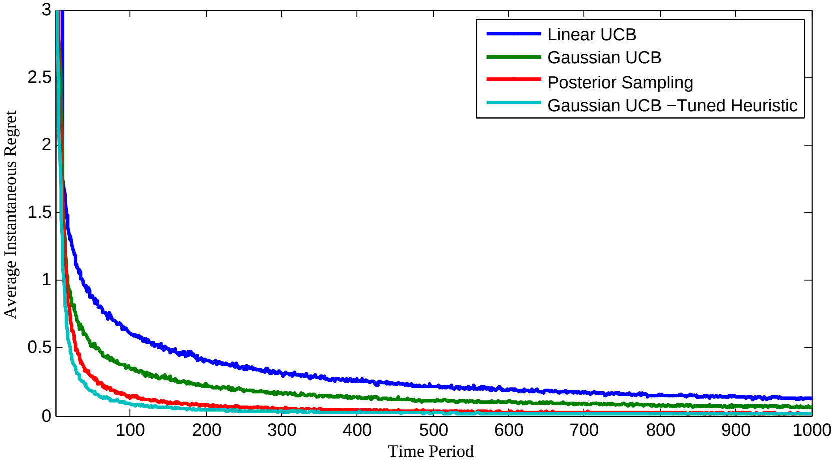
- Assume f_θ takes values in $[0, C]$. Compare decomposition for UCB and PS,

$$\text{BayesRegret}\ \left(T, \pi^U\right) \leq \mathbb{E}\sum_{t=1}^{T}\left[U_t\left(\bar{A}_t\right) - f_\theta\left(\bar{A}_t\right)\right] + C\sum_{t=1}^{T}\mathbb{P}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

$$\text{BayesRegret}\ \left(T, \pi^{PS}\right) \leq \mathbb{E}\sum_{t=1}^{T}\left[U_t\left(A_t\right) - f_\theta\left(A_t\right)\right] + C\sum_{t=1}^{T}\mathbb{P}\left(f_\theta\left(A_t^*\right) > U_t\left(A_t^*\right)\right)$$

- This is a crucial advantage when there are complicated dependencies among actions, as designing and computing with appropriate confidence sets presents significant challenges.
- This difficulty is likely the main reason that posterior sampling significantly outperforms UCB algorithms in the simulations.

Simulation results



Confidence sets and width

- Assumption 1. For all $f \in \mathcal{F}$ and $a \in \mathcal{A}$, $f(a) \in [0, C]$.
- Assumption 2. For all $t \in \mathbb{N}$, $R_t - f_\theta(A_t)$ conditioned on (H_t, θ, A_t) is σ-sub-Gaussian.
- Construct a set $\mathcal{F}_t \subset \mathcal{F}$ of functions that are statistically plausible at time t.
- Let $w_{\mathcal{F}}(a) := \sup_{\bar{f} \in \mathcal{F}} \bar{f}(a) - \inf_{f \in \mathcal{F}} f(a)$ denote the width of $\mathcal{F}$ at a.
- **Remark**: while the analysis of posterior sampling will make use of UCBs, the actual performance of posterior sampling does not depend on UCBs used in the analysis.

Confidence sets and width

- Assumption 1. For all $f \in \mathcal{F}$ and $a \in \mathcal{A}$, $f(a) \in [0, C]$.
- Assumption 2. For all $t \in \mathbb{N}$, $R_t - f_\theta(A_t)$ conditioned on (H_t, θ, A_t) is σ-sub-Gaussian.
- Construct a set $\mathcal{F}_t \subset \mathcal{F}$ of functions that are statistically plausible at time t.
- Let $w_{\mathcal{F}}(a) := \sup_{\bar{f} \in \mathcal{F}} \bar{f}(a) - \inf_{f \in \mathcal{F}} f(a)$ denote the width of $\mathcal{F}$ at a.
- **Remark**: while the analysis of posterior sampling will make use of UCBs, the actual performance of posterior sampling does not depend on UCBs used in the analysis.

Confidence sets and width

- Assumption 1. For all $f \in \mathcal{F}$ and $a \in \mathcal{A}, f(a) \in [0, C]$.
- Assumption 2. For all $t \in \mathbb{N}, R_t - f_\theta(A_t)$ conditioned on (H_t, θ, A_t) is σ-sub-Gaussian.
- Construct a set $\mathcal{F}_t \subset \mathcal{F}$ of functions that are statistically plausible at time t.
- Let $w_{\mathcal{F}}(a) := \sup_{\bar{f} \in \mathcal{F}} \bar{f}(a) - \inf_{f \in \mathcal{F}} f(a)$ denote the width of $\mathcal{F}$ at a.
- **Remark**: while the analysis of posterior sampling will make use of UCBs, the actual performance of posterior sampling does not depend on UCBs used in the analysis.

Confidence bounds and regret

Proposition 1 (Bound regret in terms of the confidence width at selected actions).
Fix any sequence $\{\mathcal{F}_t : t \in \mathbb{N}\}$, where $\mathcal{F}_t \subset \mathcal{F}$ is measurable with respect to $\sigma(H_t)$. Then for any $T \in \mathbb{N}$,

$$\textit{Regret}\left(T, \pi^U, \theta\right) \overset{a.s.}{\leq} \sum_{t=1}^{T} w_{\mathcal{F}_t}\left(\bar{A}_t\right) + C\mathbb{1}\left(f_\theta \notin \mathcal{F}_t\right)$$

$$\textit{BayesRegret}\left(T, \pi^{PS}\right) \leq \mathbb{E}\left[\sum_{t=1}^{T} w_{\mathcal{F}_t}\left(A_t\right) + C\mathbb{1}\left(f_\theta \notin \mathcal{F}_t\right)\right]$$

Confidence bounds

- Least square: $L_{2,t}(f) = \sum_1^{t-1} \left(f\left(A_t\right) - R_t\right)^2$ is the cumulative squared prediction error.
- The confidence sets constructed here are centered around least squares estimates $\hat{f}_t^{LS} \in \arg\min_{f \in \mathcal{F}} L_{2,t}(f)$.
- The sets take the form $\mathcal{F}_t := \left\{ f \in \mathcal{F} : \left\| f - \hat{f}_t^{LS} \right\|_{2,E_t} \leq \sqrt{\beta_t} \right\}$
- β_t is an appropriately chosen confidence parameter
- the empirical 2-norm $\| \cdot \|_{2,E_t}$ is defined by $\|g\|_{2,E_t}^2 = \sum_1^{t-1} g^2\left(A_k\right)$.
  - Hence $\|f - f_\theta\|_{2,E_t}^2$ measures the cumulative discrepancy between the previous predictions of f and f_θ.

Confidence bounds

Proposition 2 (High-probability bounds).

Define the confidence parameter,

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t \left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

For all $\delta > 0$ *and* $\alpha > 0$, *if*

$$\mathcal{F}_t = \left\{ f \in \mathcal{F} : \left\| f - \hat{f}_t^{LS} \right\|_{2,E_t} \leq \sqrt{\beta_t^*(\mathcal{F}, \delta, \alpha)} \right\}$$

for all $t \in \mathbb{N}$, *then*

$$\mathbb{P}\left(f_\theta \in \bigcap_{t=1}^{\infty} \mathcal{F}_t \right) \geq 1 - 2\delta$$

The shrink rate of confidence width - Key theorem

Proposition 3 (Potential Function).
If $(\beta_t \geq 0 \mid t \in \mathbb{N})$ *is a nondecreasing sequence and* $\mathcal{F}_t := \left\{ f \in \mathcal{F} : \left\| f - \hat{f}_t^{LS} \right\|_{2,E_t} \leq \sqrt{\beta_t} \right\}$, *then for all* $T \in \mathbb{N}$ *and* $\epsilon > 0$,

$$\sum_{t=1}^{T} \mathbf{1}\left(w_{\mathcal{F}_t}\left(A_t\right) > \epsilon\right) \leq \left(\frac{4\beta_T}{\epsilon^2} + 1\right) \dim_E(\mathcal{F}, \epsilon).$$

The shrink rate of confidence width - Key theorem

Lemma 8 (Potential Lemma).
If $(\beta_t \geq 0 \mid t \in \mathbb{N})$ *is a nondecreasing sequence and* $\mathcal{F}_t := \left\{ f \in \mathcal{F} : \left\| f - \hat{f}_t^{LS} \right\|_{2,E_t} \leq \sqrt{\beta_t} \right\}$*, then for all* $T \in \mathbb{N}$*,*

$$\sum_{t=1}^{T} w_{\mathcal{F}_t}(A_t) \leq 1 + \dim_E\left(\mathcal{F}, T^{-1}\right) C + 4\sqrt{\dim_E\left(\mathcal{F}, T^{-1}\right) \beta_T T}.$$

Final results

Proposition 4.

For all $T \in \mathbb{N}, \alpha > 0$ *and* $\delta \leq 1/2T$,

$$\mathbb{E}\left[\text{Regret}\left(T, \pi^{U}, \theta\right) \mid \theta\right] \leq 1 + \left[\dim_E\left(\mathcal{F}, T^{-1}\right) + 1\right] C + 4\sqrt{\dim_E\left(\mathcal{F}, T^{-1}\right) \beta_T^*(\mathcal{F}, \alpha, \delta) T}$$

$$\text{BayesRegret}\left(T, \pi^{PS}\right) \leq 1 + \left[\dim_E\left(\mathcal{F}, T^{-1}\right) + 1\right] C + 4\sqrt{\dim_E\left(\mathcal{F}, T^{-1}\right) \beta_T^*(\mathcal{F}, \alpha, \delta) T}$$

Outline

Proof of proposition 3 (Potential Function)

Step 1 If $w_t(A_t) > \epsilon$, then A_t is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(A_1, \ldots, A_{t-1})$ for $T > t$.

Step 2 In any action sequence $(a_1, \ldots, a_\tau)$, there is some element a_j that is ϵ-dependent on at least $\tau/d - 1$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$, where $d := \dim_E(\mathcal{F}, \epsilon)$.

Step 3 Now, consider taking $(a_1, \ldots, a_\tau)$ to be the subsequence $(A_{t_1}, \ldots, A_{t_\tau})$ of $(A_1, \ldots, A_T)$ consisting of elements A_t for which $w_{\mathcal{F}_t}(A_t) > \epsilon$, i.e. $w_{\mathcal{F}_{t_j}}(A_{t_j}) > \epsilon, \forall j = 1 \ldots \tau$.

- By step 1, each A_{t_j} is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(A_1, \ldots, A_{t_j - 1})$.
- It follows that each a_j is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$.
- Combining Step 2, we have $\tau/d - 1 \leq 4\beta_T/\epsilon^2$. It follows that $\tau \leq (4\beta_T/\epsilon^2 + 1)\, d$. **Done.**

Proof of proposition 3 - Step 1

Step 1 If $w_t(A_t) > \epsilon$, then A_t is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(A_1, \ldots, A_{t-1})$ for $T > t$.

- $w_{\mathcal{F}_t}(A_t) > \epsilon \Longrightarrow \exists \bar{f}, \underline{f} \in \mathcal{F}_t, \bar{f}(A_t) - \underline{f}(A_t) > \epsilon$.
- By definition, since $\bar{f}(A_t) - \underline{f}(A_t) > \epsilon$, if A_t is ϵ-dependent on a subsequence $(A_{i_1}, \ldots, A_{i_k})$ of $(A_1, \ldots, A_{t-1})$, then $\sum_{j=1}^{k} \left(\bar{f}(A_{i_j}) - \underline{f}(A_{i_j})\right)^2 > \epsilon^2$.
- It follows that, if A_t is ϵ-dependent on K disjoint subsequences of $(A_1, \ldots, A_{t-1})$, then $\|\bar{f} - \underline{f}\|_{2,E_t}^2 > K\epsilon^2$.
- By the triangle inequality, we have

$$\|\bar{f} - \underline{f}\|_{2,E_t} \leq \left\|\bar{f} - \hat{f}_t^{LS}\right\|_{2,E_t} + \left\|\underline{f} - \hat{f}_t^{LS}\right\|_{2,E_t} \leq 2\sqrt{\beta_t} \leq 2\sqrt{\beta_T}$$

- Then $K < 4\beta_T/\epsilon^2$.

Proof of proposition 3 - Step 1

Step 1 If $w_t(A_t) > \epsilon$, then A_t is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(A_1, \ldots, A_{t-1})$ for $T > t$.

- $w_{\mathcal{F}_t}(A_t) > \epsilon \Longrightarrow \exists \bar{f}, \underline{f} \in \mathcal{F}_t, \bar{f}(A_t) - \underline{f}(A_t) > \epsilon$.
- By definition, since $\bar{f}(A_t) - \underline{f}(A_t) > \epsilon$, if A_t is ϵ-dependent on a subsequence $(A_{i_1}, \ldots, A_{i_k})$ of $(A_1, \ldots, A_{t-1})$, then $\sum_{j=1}^{k} \left(\bar{f}(A_{i_j}) - \underline{f}(A_{i_j})\right)^2 > \epsilon^2$.
- It follows that, if A_t is ϵ-dependent on K disjoint subsequences of $(A_1, \ldots, A_{t-1})$, then $\|\bar{f} - \underline{f}\|_{2,E_t}^2 > K\epsilon^2$.
- By the triangle inequality, we have

$$\|\bar{f} - \underline{f}\|_{2,E_t} \leq \left\|\bar{f} - \hat{f}_t^{LS}\right\|_{2,E_t} + \left\|\underline{f} - \hat{f}_t^{LS}\right\|_{2,E_t} \leq 2\sqrt{\beta_t} \leq 2\sqrt{\beta_T}$$

- Then $K < 4\beta_T/\epsilon^2$.

Proof of proposition 3 - Step 1

Step 1 If $w_t(A_t) > \epsilon$, then A_t is ϵ-dependent on fewer than $4\beta_T/\epsilon^2$ disjoint subsequences of $(A_1, \ldots, A_{t-1})$ for $T > t$.

- $w_{\mathcal{F}_t}(A_t) > \epsilon \Longrightarrow \exists \bar{f}, \underline{f} \in \mathcal{F}_t, \bar{f}(A_t) - \underline{f}(A_t) > \epsilon$.
- By definition, since $\bar{f}(A_t) - \underline{f}(A_t) > \epsilon$, if A_t is ϵ-dependent on a subsequence $(A_{i_1}, \ldots, A_{i_k})$ of $(A_1, \ldots, A_{t-1})$, then $\sum_{j=1}^{k} \left(\bar{f}\left(A_{i_j}\right) - \underline{f}\left(A_{i_j}\right)\right)^2 > \epsilon^2$.
- It follows that, if A_t is ϵ-dependent on K disjoint subsequences of $(A_1, \ldots, A_{t-1})$, then $\|\bar{f} - \underline{f}\|_{2,E_t}^2 > K\epsilon^2$.
- By the triangle inequality, we have

$$\|\bar{f} - \underline{f}\|_{2,E_t} \leq \left\|\bar{f} - \hat{f}_t^{LS}\right\|_{2,E_t} + \left\|\underline{f} - \hat{f}_t^{LS}\right\|_{2,E_t} \leq 2\sqrt{\beta_t} \leq 2\sqrt{\beta_T}$$

- Then $K < 4\beta_T/\epsilon^2$.

Proof of proposition 3 - Step 2 Intuition

Step 2 In any action sequence $(a_1, \ldots, a_\tau)$, there is some element a_j that is ϵ-dependent on at least $\tau/d - 1$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$, where $d := \dim_E(\mathcal{F}, \epsilon)$.

- Let us again get some intuition from linear algebra! ($\mathcal{F}$ linear function class and $\epsilon = 0$)
- ϵ-dependency now become linear dependency.
- W.l.o.g let $\tau = Kd + 1$. Sampling basis of $\mathbb{R}^d$ one by one:

$$a_1 = e_1, a_2 = e_2, \ldots, a_d = e_d, \ldots, a_{id+j} = e_j, \ldots, a_\tau = e_1$$

- Form every round of sampled basis $B_i = \{a_{(i-1)d+1}, \ldots, a_{(i)d}\}$ as a subsequence, $i = 1, \ldots, K$
- then a_τ is linearly dependent on all previous constructed disjoint subsequences, which is $K > \tau/d - 1$

Proof of proposition 3 - Step 2 Formal constructive proof

Step 2 In any action sequence $(a_1, \ldots, a_\tau)$, there is some element a_j that is ϵ-dependent on at least $\tau/d - 1$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$, where $d := \dim_E(\mathcal{F}, \epsilon)$.

- For an integer K satisfying $Kd + 1 \leq \tau \leq Kd + d$, we will construct K disjoint subsequences $B_1, \ldots, B_K$.
- First let $B_i = (a_i)$ for $i = 1, \ldots, K$. If a_{K+1} is ϵ-dependent on each subsequence $B_1, \ldots, B_K$, our claim is established.
- Otherwise, select a subsequence B_i s.t. a_{K+1} is ϵ-independent and append a_{K+1} to B_i.
- Repeat this process for elements with indices $j > K + 1$ until a_j is ϵ-dependent on each subsequence or $j = \tau$.
- In the latter scenario $(j = \tau)$, $\sum_i |B_i| \geq Kd$,
- and since each element of a subsequence B_i is ϵ-independent of its predecessors, $|B_i| = d$.
- In this case, a_τ must be ϵ-dependent on each subsequence.

Proof of proposition 3 - Step 2 Formal constructive proof

Step 2 In any action sequence $(a_1, \ldots, a_\tau)$, there is some element a_j that is ϵ-dependent on at least $\tau/d - 1$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$, where $d := \dim_E(\mathcal{F}, \epsilon)$.

- For an integer K satisfying $Kd + 1 \leq \tau \leq Kd + d$, we will construct K disjoint subsequences $B_1, \ldots, B_K$.
- First let $B_i = (a_i)$ for $i = 1, \ldots, K$. If a_{K+1} is ϵ-dependent on each subsequence $B_1, \ldots, B_K$, our claim is established.
- Otherwise, select a subsequence B_i s.t. a_{K+1} is ϵ-independent and append a_{K+1} to B_i.
- Repeat this process for elements with indices $j > K + 1$ until a_j is ϵ-dependent on each subsequence or $j = \tau$.
- In the latter scenario $(j = \tau)$, $\sum_i |B_i| \geq Kd$,
- and since each element of a subsequence B_i is ϵ-independent of its predecessors, $|B_i| = d$.
- In this case, a_τ must be ϵ-dependent on each subsequence.

Proof of proposition 3 - Step 2 Formal constructive proof

Step 2 In any action sequence $(a_1, \ldots, a_\tau)$, there is some element a_j that is ϵ-dependent on at least $\tau/d - 1$ disjoint subsequences of $(a_1, \ldots, a_{j-1})$, where $d := \dim_E(\mathcal{F}, \epsilon)$.

- For an integer K satisfying $Kd + 1 \leq \tau \leq Kd + d$, we will construct K disjoint subsequences $B_1, \ldots, B_K$.
- First let $B_i = (a_i)$ for $i = 1, \ldots, K$. If a_{K+1} is ϵ-dependent on each subsequence $B_1, \ldots, B_K$, our claim is established.
- Otherwise, select a subsequence B_i s.t. a_{K+1} is ϵ-independent and append a_{K+1} to B_i.
- Repeat this process for elements with indices $j > K + 1$ until a_j is ϵ-dependent on each subsequence or $j = \tau$.
- In the latter scenario $(j = \tau)$, $\sum_i |B_i| \geq Kd$,
- and since each element of a subsequence B_i is ϵ-independent of its predecessors, $|B_i| = d$.
- In this case, a_τ must be ϵ-dependent on each subsequence.

Proof of Lemma 8 (Potential Lemma)

- Write $d = \dim_E\left(\mathcal{F}, T^{-1}\right)$ and $w_t = w_t\left(A_t\right)$.
- Reorder the sequence $\left(w_1, \ldots, w_T\right) \rightarrow \left(w_{i_1}, \ldots, w_{i_T}\right)$, where $w_{i_1} \geq w_{i_2} \geq \cdots \geq w_{i_T}$.
- $\sum_{t=1}^T w_{\mathcal{F}_t}\left(A_t\right) = \sum_{t=1}^T w_{i_t} =$

$$\sum_{t=1}^T w_{i_t} \mathbf{1}\left\{w_{i_t} \leq T^{-1}\right\} + \sum_{t=1}^T w_{i_t} \mathbf{1}\left\{w_{i_t} > T^{-1}\right\} \leq 1 + \sum_{t=1}^T w_{i_t} \mathbf{1}\left\{w_{i_t} \geq T^{-1}\right\}$$

- We know $w_{i_t} \leq C$. In addition,

$$w_{i_t} > \epsilon \Longleftrightarrow \sum_{k=1}^T \mathbf{1}\left(w_{\mathcal{F}_k}\left(A_k\right) > \epsilon\right) \geq t.$$

- By Proposition 3 (Potential Function), this can only occur if $t < \left(\left(4\beta_T\right)/\epsilon^2 + 1\right) \dim_E(\mathcal{F}, \epsilon)$.

Proof of Lemma 8 (Potential Lemma)

- For $\epsilon \geq T^{-1}, \dim_E(\mathcal{F}, \epsilon) \leq \dim_E\left(\mathcal{F}, T^{-1}\right) = d$, since $\dim_E\left(\mathcal{F}, \epsilon\right)$ is non-increasing in tolerance ϵ.
- Therefore, when $w_{i_t} > \epsilon \geq T^{-1}, t < \left(\left(4\beta_T\right)/\epsilon^2 + 1\right) d$, which implies $\epsilon < \sqrt{\left(4\beta_T d\right)/(t-d)}$.
- This shows that if $w_{i_t} > T^{-1}$, for $\epsilon \geq T^{-1}$, taking $\epsilon \uparrow w_{i_t}$, then

$$w_{i_t} \leq \min\left\{C, \sqrt{\left(4\beta_T d\right)/(t-d)}\right\}.$$

- Therefore,

$$\sum_{t=1}^{T} w_{i_t} \mathbf{1}\left\{w_{i_t} > T^{-1}\right\} \leq dC + \sum_{t=d+1}^{T} \sqrt{\frac{4d\beta_T}{t-d}} \leq dC + 2\sqrt{d\beta_T} \int_{t=0}^{T} \frac{1}{\sqrt{t}} dt = dC + 4\sqrt{d\beta_T T}.$$

Outline

Outline

Remark - confidence parameter β^*

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

- (FINITE FUNCTION CLASSES). When $\mathcal{F}$ is finite, $\beta_t^*(\mathcal{F}, \delta, 0) = 8\sigma^2 \log(|\mathcal{F}|/\delta)$.
- (LINEAR MODELS). Consider a d-dimensional linear model $f_\rho(a) := \langle \phi(a), \rho \rangle$.
  - Fix $\gamma = \sup_{a \in \mathcal{A}} \|\phi(a)\|$ and $s = \sup_{\rho \in \Theta} \|\rho\|$.
  - Hence, for all $\rho_1, \rho_2 \in \mathcal{F}$, we have $\|f_{\rho_1} - f_{\rho_2}\|_\infty \leq \gamma \|\rho_1 - \rho_2\|$.
  - An α-covering of $\mathcal{F}$ can therefore be attained through an (α/γ)-covering of $\Theta \subset \mathbb{R}^d$.
  - Such a covering requires $O\left((1/\alpha)^d\right)$ elements, and it follows that, $\log N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right) = O(d \log(1/\alpha))$.
  - If α is chosen to be $1/t^2$, the second term in β_t^* tends to zero, and therefore, $\beta_t^*\left(\mathcal{F}, \delta, 1/t^2\right) = O(d \log(t/\delta))$.

Remark - confidence parameter β^*

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

- (FINITE FUNCTION CLASSES). When $\mathcal{F}$ is finite, $\beta_t^*(\mathcal{F}, \delta, 0) = 8\sigma^2 \log(|\mathcal{F}|/\delta)$.
- (LINEAR MODELS). Consider a d-dimensional linear model $f_\rho(a) := \langle \phi(a), \rho \rangle$.
  - Fix $\gamma = \sup_{a \in \mathcal{A}} \|\phi(a)\|$ and $s = \sup_{\rho \in \Theta} \|\rho\|$.
  - Hence, for all $\rho_1, \rho_2 \in \mathcal{F}$, we have $\|f_{\rho_1} - f_{\rho_2}\|_\infty \leq \gamma \|\rho_1 - \rho_2\|$.
  - An α-covering of $\mathcal{F}$ can therefore be attained through an (α/γ)-covering of $\Theta \subset \mathbb{R}^d$.
  - Such a covering requires $O\left((1/\alpha)^d\right)$ elements, and it follows that, $\log N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right) = O(d \log(1/\alpha))$.
  - If α is chosen to be $1/t^2$, the second term in β_t^* tends to zero, and therefore, $\beta_t^*\left(\mathcal{F}, \delta, 1/t^2\right) = O(d \log(t/\delta))$.

Remark - confidence parameter β^*

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

- (FINITE FUNCTION CLASSES). When $\mathcal{F}$ is finite, $\beta_t^*(\mathcal{F}, \delta, 0) = 8\sigma^2 \log(|\mathcal{F}|/\delta)$.
- (LINEAR MODELS). Consider a d-dimensional linear model $f_\rho(a) := \langle \phi(a), \rho \rangle$.
  - Fix $\gamma = \sup_{a \in \mathcal{A}} \|\phi(a)\|$ and $s = \sup_{\rho \in \Theta} \|\rho\|$.
  - Hence, for all $\rho_1, \rho_2 \in \mathcal{F}$, we have $\|f_{\rho_1} - f_{\rho_2}\|_\infty \leq \gamma \|\rho_1 - \rho_2\|$.
  - An α-covering of $\mathcal{F}$ can therefore be attained through an (α/γ)-covering of $\Theta \subset \mathbb{R}^d$.
  - Such a covering requires $O\left((1/\alpha)^d\right)$ elements, and it follows that, $\log N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right) = O(d \log(1/\alpha))$.
  - If α is chosen to be $1/t^2$, the second term in β_t^* tends to zero, and therefore, $\beta_t^*\left(\mathcal{F}, \delta, 1/t^2\right) = O(d \log(t/\delta))$.

Remark - confidence parameter β^*

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

- (FINITE FUNCTION CLASSES). When $\mathcal{F}$ is finite, $\beta_t^*(\mathcal{F}, \delta, 0) = 8\sigma^2 \log(|\mathcal{F}|/\delta)$.
- (LINEAR MODELS). Consider a d-dimensional linear model $f_\rho(a) := \langle \phi(a), \rho \rangle$.
  - Fix $\gamma = \sup_{a \in \mathcal{A}} \|\phi(a)\|$ and $s = \sup_{\rho \in \Theta} \|\rho\|$.
  - Hence, for all $\rho_1, \rho_2 \in \mathcal{F}$, we have $\|f_{\rho_1} - f_{\rho_2}\|_\infty \leq \gamma \|\rho_1 - \rho_2\|$.
  - An α-covering of $\mathcal{F}$ can therefore be attained through an (α/γ)-covering of $\Theta \subset \mathbb{R}^d$.
  - Such a covering requires $O\left((1/\alpha)^d\right)$ elements, and it follows that, $\log N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right) = O(d \log(1/\alpha))$.
  - If α is chosen to be $1/t^2$, the second term in β_t^* tends to zero, and therefore, $\beta_t^*\left(\mathcal{F}, \delta, 1/t^2\right) = O(d \log(t/\delta))$.

Remark - confidence parameter β^*

$$\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right).$$

- (GENERALIZED LINEAR MODELS). Consider the case of a d -dimensional generalized linear model $f_\theta(a) := g(\langle\phi(a), \theta\rangle)$, where g is an increasing Lipschitz continuous function.
  - Fix $g, \gamma = \sup_{a \in \mathcal{A}} \|\phi(a)\|$ and $s = \sup_{\rho \in \Theta} \|\rho\|$.
  - Then, the previous argument shows $\log N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right) = O(d \log(1/\alpha))$.
  - Again, choosing $\alpha = 1/t^2$ yields a confidence parameter $\beta_t^*\left(\mathcal{F}, \delta, 1/t^2\right) = O(d \log(t/\delta))$.

Remark - relate β^* to Kolmogorov dimension

Definition 9 (Kolmogorov dimension).

The Kolmogorov dimension of a function class $\mathcal{F}$ is given by

$$\dim_K(\mathcal{F}) = \limsup_{\alpha \downarrow 0} \frac{\log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)\right)}{\log\left(1/\alpha\right)}.$$

Example : $\dim_K\left(\mathbb{R}^d\right) = d$

- $\beta_t^*(\mathcal{F}, \delta, \alpha) := 8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right)$

$$\beta_t^*\left(\mathcal{F}, 1/t^2, 1/t^2\right) = 8\sigma^2\left[\frac{\log\left(N\left(\mathcal{F}, 1/t^2, \|\cdot\|_\infty\right)\right)}{\log\left(t^2\right)} + 1\right]\log\left(t^2\right) + 2\frac{t}{t^2}\left(8C + \sqrt{8\sigma^2 \ln\left(4t^2\delta\right)}\right)$$
$$= 16\left(1 + o(1) + \dim_K(\mathcal{F})\right)\log t$$

- $\limsup_{t\to\infty} \log\left(N\left(\mathcal{F}, 1/t^2, \|\cdot\|_\infty\right)\right)/\log\left(t^2\right) = \dim_K(\mathcal{F})$.

Outline

Equivelant definition of eluder dimension

- The ϵ-eluder dimension of a class of functions $\mathcal{F}$ is the length of the longest sequence $a_1, \ldots, a_\tau$ such that for some $\epsilon' \geq \epsilon$

$$w_k := \sup \left\{ (f_{\rho_1} - f_{\rho_2})(a_k) : \sqrt{\sum_{i=1}^{k-1} (f_{\rho_1} - f_{\rho_2})^2 (a_i)} \leq \epsilon', \rho_1, \rho_2 \in \Theta \right\} > \epsilon'$$

for each $k \leq \tau$

Eluder dim for Finite action spaces

- Any action is ϵ'-dependent on itself since

$$\sup\left\{(f_{\rho_1} - f_{\rho_2})(a) : \sqrt{(f_{\rho_1} - f_{\rho_2})^2(a)} \leq \epsilon' \rho_1, \rho_2 \in \Theta\right\} \leq \epsilon'$$

  Therefore, for all $\epsilon > 0$, the ϵ-eluder dimension of $\mathcal{A}$ is bounded by $|\mathcal{A}|$

Eluder dim for Linear model

Proposition 5.

Suppose $\Theta \subset \mathbb{R}^d$ *and* $f_\theta(a) = \theta^T \phi(a)$. *Assume there exist constants* γ *and* S *such that for all* $a \in \mathcal{A}$ *and* $\rho \in \Theta, \|\rho\|_2 \leq S$, *and* $\|\phi(a)\|_2 \leq \gamma$. *Then*

$$\dim_E(\mathcal{F}, \epsilon) \leq 3d(e/(e-1)) \ln \left\{3 + 3((2S)/\epsilon)^2\right\} + 1$$

- To simplify the notation, define w_k as in previous page, $\phi_k = \phi\left(a_k\right), \rho = \rho_1 - \rho_2$, and $\Phi_k = \sum_{i=1}^{k-1} \phi_i \phi_i^T$.
- In this case, $\sum_{i=1}^{k-1}\left(f_{\rho_1} - f_{\rho_2}\right)^2\left(a_i\right) = \rho^T \Phi_k \rho$, and by the triangle inequality $\|\rho\|_2 \leq 2S$.
- The proof follows by bounding the number of times $w_k > \epsilon'$ can occur.

Eluder dim for Linear model - Proof Sketch

Step 1. If $w_k \geq \epsilon'$, then $\phi_k^T V_k^{-1} \phi_k \geq \frac{1}{2}$ where $V_k := \Phi_k + \lambda I$ and $\lambda = (\epsilon'/(2S))^2$.

Step 2. If $w_i \geq \epsilon'$ for each $i < k$, then $\det V_k \geq \lambda^d \left(1 + \frac{1}{2}\right)^{k-1}$ and det $V_k \leq \left(\left(\gamma^2(k-1)\right)/d + \lambda\right)^d$.

Step 3. Complete proof by solving k with the upper and lower bound of $\det V_k$.

Eluder dim for Linear model - Proof

Step 1. If $w_k \geq \epsilon'$, then $\phi_k^T V_k^{-1} \phi_k \geq \frac{1}{2}$ where $V_k := \Phi_k + \lambda I$ and $\lambda = (\epsilon'/(2S))^2$.

- We find

$$
\begin{aligned}
w_k &\leq \max\left\{\rho^T \phi_k : \rho^T \Phi_k \rho \leq (\epsilon')^2, \rho^T I \rho \leq (2S)^2\right\} \\
&\leq \max\left\{\rho^T \phi_k : \rho^T V_k \rho_k \leq 2(\epsilon')^2\right\} = \sqrt{2}(\epsilon')^2 \|\phi_k\|_{V_k^{-1}}.
\end{aligned}
$$

- The second inequality follows because any ρ that is feasible for the first maximization problem must satisfy $\rho^T V_k \rho \leq (\epsilon')^2 + \lambda(2S)^2 = 2(\epsilon')^2$.
- The third inequality follows by Cauchy-Schwarz inequality.
- By this result, $w_k \geq \epsilon'$ implies $\|\phi_k\|_{V_k^{-1}}^2 \geq 1/2$

Eluder dim for Linear model - Proof

Step 2. If $w_i \geq \epsilon'$ for each $i < k$, then $\det V_k \geq \lambda^d \left(\frac{3}{2}\right)^{k-1}$ and det $V_k \leq \left(\left(\gamma^2(k-1)\right)/d + \lambda\right)^d$.

- Since $V_k = V_{k-1} + \phi_k \phi_k^T$, using the matrix determinant lemma,

$$\det V_k = \det V_{k-1}\left(1 + \phi_t^T V_k^{-1} \phi_t\right) \geq \det V_{k-1}\left(\frac{3}{2}\right) \geq \cdots \geq \det[\lambda I]\left(\frac{3}{2}\right)^{k-1} = \lambda^d \left(\frac{3}{2}\right)^{k-1}$$

- Recall that det V_k is the product of the eigenvalues of V_k, whereas trace $[V_k]$ is the sum.
- By AM-GM inequality, $\det V_k$ is maximized when all eigenvalues are equal. This implies

$$\det V_k \leq \left(\left(\text{trace}\,[V_k]\right)/d\right)^d \leq \left(\left(\gamma^2(k-1)\right)/d + \lambda\right)^d.$$

Eluder dim for Linear model - Proof

Step 3. Manipulating the result of Step 2 shows k must satisfy the inequality: $\left(\frac{3}{2}\right)^{(k-1)/d} \leq \alpha_0[(k-1)/d]+1$, where $\alpha_0 = \gamma^2/\lambda = (2S\gamma/\epsilon')^2$. Let $B(x,\alpha) = \max\left\{B : (1+x)^B \leq \alpha B + 1\right\}$.

- The number of times $w_k > \epsilon'$ can occur is bounded by $dB\left(1/2, \alpha_0\right) + 1$
- Note that any $B \geq 1$ must satisfy the inequality $\ln\{1+x\}B \leq \ln\{1+\alpha\} + \ln B$. Since $\ln\{1+x\} \geq x/(1+x)$, using the transformation of variables $y = B[x/(1+x)]$ gives

$$
\begin{aligned}
y \leq \ln\{1+\alpha\} + \ln\frac{1+x}{x} + \ln y \leq \ln\{1+\alpha\} + \ln\frac{1+x}{x} + \frac{y}{e} \\
\implies \quad y \leq \frac{e}{e-1}\left(\ln\{1+\alpha\} + \ln\frac{1+x}{x}\right)
\end{aligned}
$$

- This implies $B(x,\alpha) \leq ((1+x)/x)(e/(e-1))(\ln\{1+\alpha\} + \ln((1+x)/x))$.

Elliptical potential lemma

- Let $A_1, A_2, \cdots$ be a sequence of vectors in $\mathbb{R}^d$ that satisfy $\|A_t\|_2 \leq 1$ for all $t \geq 1$. For a fixed constant λ with $\lambda \geq 1$, define the sequence of covariance matrices $\{\Sigma_t\}_{t \geq 0}$ as follows:

$$\boldsymbol{\Sigma}_1^{-1} := \lambda \mathbb{I}_d \quad , \quad \boldsymbol{\Sigma}_t^{-1} := \lambda \mathbb{I}_d + \sum_{\tau=1}^{t-1} A_\tau A_\tau^\top$$

- The elliptical potential lemma then asserts that

$$\sum_{t=1}^{T} A_t^\top \boldsymbol{\Sigma}_t A_t \leq 2 \log \frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}_{T+1}} \leq 2d \log \left(1 + \frac{T}{\lambda d}\right)$$

Information theoretic perspective of the elliptical potential lemma

- Suppose $R_t = \theta^\top A_t + \mathcal{N}(0,1)$ and $\mathcal{D} = (A_1, R_1, \ldots, A_{t-1}, R_{t-1})$
- Information gain of the new observation A_t, R_t,

$$\begin{aligned}
\mathrm{I}\left(\theta; A_t, R_t \mid \mathcal{D}\right) &= \mathrm{H}(\theta \mid \mathcal{D}) - \mathrm{H}\left(\theta \mid \mathcal{D}, A_t, R_t\right) \\
&= (1/2)\mathbb{E}\left[\log \frac{\det\left(\boldsymbol{\Sigma}_t\right)}{\det(\boldsymbol{\Sigma}_{t+1})} \mid \mathcal{D}\right], \quad \text{where } \boldsymbol{\Sigma}_{t+1}^{-1} = \boldsymbol{\Sigma}_t^{-1} + A_t A_t^\top \\
&= (1/2)\mathbb{E}\left[\log\det\left(I + \boldsymbol{\Sigma}_t^{1/2} A_t A_t^\top \boldsymbol{\Sigma}_t^{1/2}\right) \mid \mathcal{D}\right] \\
&= (1/2)\mathbb{E}\left[\log\left(1 + A_t^\top \boldsymbol{\Sigma}_t A_t\right) \mid \mathcal{D}\right]
\end{aligned}$$

- Mutual information between the model parameter and history observations:

$$\mathrm{I}\left(\theta; A_1, R_1, \cdots, A_T, R_T\right) = (1/2)\mathbb{E}\left[\log \frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}_{T+1}}\right]$$

Eluder dim for Generalized linear models

Proposition 6.

Suppose $\Theta \subset \mathbb{R}^d$ *and* $f_\theta(a) = g\left(\theta^T \phi(a)\right)$ *where* $g(\cdot)$ *is a differentiable and strictly increasing function. Assume that there exist constants* $\underline{h}, \bar{h}, \gamma$, *and* S *such that for all* $a \in \mathcal{A}$ *and* $\rho \in \Theta, 0 < \underline{h} \leq g'\left(\rho^T \phi(a)\right) \leq \bar{h}, \|\rho\|_2 \leq S$, *and* $\|\phi(a)\|_2 \leq \gamma$. *Then*

$$\dim_E(\mathcal{F}, \epsilon) \leq 3dr^2(e/(e-1)) \ln\left\{3r^2 + 3r^2((2S\bar{h})/\epsilon)^2\right\} + 1$$

- Similar to the linear case.

Step 1. If $w_k \geq \epsilon'$, then $\phi_k^T V_k^{-1} \phi_k \geq 1/\left(2r^2\right)$ where $V_k := \Phi_k + \lambda I$ and $\lambda = (\epsilon'/(2S\underline{h}))^2$.

Step 2. If $w_i \geq \epsilon'$ for each $i < k$, then det $V_k \geq \lambda^d \left(\frac{3}{2}\right)^{k-1}$ and det $V_k \leq \left(\left(\gamma^2(k-1)\right)/d + \lambda\right)^d$.

Step 3. Complete proof by comparing the lower and upper bound of $\det V_k$ to solve k.

Eluder dim for Generalized linear models

Step 1. If $w_k \geq \epsilon'$, then $\phi_k^T V_k^{-1} \phi_k \geq 1/\left(2r^2\right)$ where $V_k := \Phi_k + \lambda I$ and $\lambda = (\epsilon'/(2S\underline{h}))^2$.

- By definition $w_k \leq \max\left\{g\left(\rho^T\phi_k\right) : \sum_{i=1}^{k-1} g\left(\rho^T\phi\left(a_i\right)\right)^2 \leq \left(\epsilon'\right)^2, \rho^T I\rho \leq (2S)^2\right\}$.
- By the uniform bound on $g'(\cdot)$ this is less than $\max\left\{\bar{h}\rho^T\phi_k : \underline{h}^2\rho^T\Phi_k\rho \leq \left(\epsilon'\right)^2, \rho^T I\rho \leq (2S)^2\right\} \leq \max\left\{\bar{h}\rho^T\phi_k : \underline{h}^2\rho^T V_k\rho \leq 2\left(\epsilon'\right)^2\right\} = \sqrt{2\left(\epsilon'\right)^2/r^2}\,\|\phi_k\|_{V_k^{-1}}$.

Eluder dim for Generalized linear models

Step 2. If $w_i \geq \epsilon'$ for each $i < k$, then $\det V_k \geq \lambda^d \left(\frac{3}{2}\right)^{k-1}$ and $\det V_k \leq \left(\left(\gamma^2(k-1)\right)/d + \lambda\right)^d$.

Step 3. The above inequalities imply k must satisfy $\left(1 + 1/\left(2r^2\right)\right)^{(k-1)/d} \leq \alpha_0[(k-1)/d]$, where $\alpha_0 = \gamma^2/\lambda$.

- Therefore, as in the linear case, the number of times $w_k > \epsilon'$ can occur is bounded by $dB\left(1/\left(2r^2\right), \alpha_0\right) + 1$.
- Plugging these constants into the earlier bound $B(x, \alpha) \leq ((1+x)/x)(e/(e-1))(\ln\{1+\alpha\} + \ln((1+x)/x))$ and using $1 + x \leq 3/2$, yields the result.

Conclusion

- MABs (RL) / Online Learning require fundamentally different notions of model complexity.
- Huge value in having a unified conceptual understanding.

Outline

Other notion of complexity for online (sequential) learning

- **Sequential Rademacher Complexity**
- A. Rakhlin and K. Sridharan. Online non-parametric regression. In Conference on Learning Theory, pages 1232-1264, 2014.
- A. Rakhlin and K. Sridharan. On martingale extensions of vapnik-chervonenkis theory with applications to online learning. In Measures of Complexity, pages 197-215. Springer, 2015.
- A. Rakhlin, K. Sridharan, and A. Tewari. Sequential complexities and uniform martingale laws of large numbers. Probability Theory and Related Fields, 161(1-2):111-153, 2015.

Eluder dimension and its relation to RL

- Eluder Dimension applied to model-based RL [Osband and Van Roy 14', Szepesvari and Mengdi Wang et al. 20']
- Eluder Dimension applied to value-based RL [WSY20]
- Bellman Rank [JKALS17]
- Bellman Eluder Dimension [JLM21]

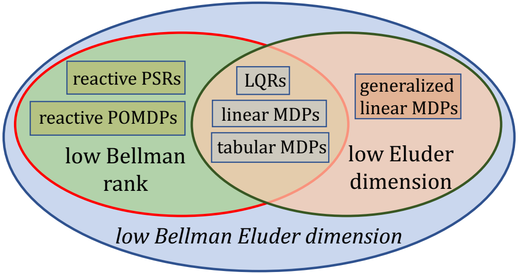


Figure: A schematic summarizing relations among families of RL problems

Outline

Proof of Proposition 2

Lemma 10 (Concentration).

For any $\delta > 0$ *and* $f : \mathcal{A} \mapsto \mathbb{R}$, *with probability at least* $1 - \delta$,

$$L_{2,t}(f) \geq L_{2,t}(f_\theta) + \frac{1}{2}\|f - f_\theta\|_{2,E_t}^2 - 4\sigma^2 \log(1/\delta)$$

simultaneously for all $t \in \mathbb{N}$.

Lemma 11 (Discretization error).

If f^α *satisfies* $\|f - f^\alpha\|_\infty \leq \alpha$, *then with probability at least* $1 - \delta$,

$$\left|\frac{1}{2}\|f^\alpha - f_\theta\|_{2,E_t}^2 - \frac{1}{2}\|f - f_\theta\|_{2,E_t}^2 + L_{2,t}(f) - L_{2,t}(f^\alpha)\right| \leq \alpha t \left[8C + \sqrt{8\sigma^2 \ln(4t^2/\delta)}\right] \quad \forall t \in \mathbb{N}$$

Proof of Proposition 2

- Let $\mathcal{F}^\alpha \subset \mathcal{F}$ be an α-cover of $\mathcal{F}$ in the sup norm in the sense that, for any $f \in \mathcal{F}$, there is an $f^\alpha \in \mathcal{F}^\alpha$ such that $\|f^\alpha - f\|_\infty \leq \epsilon$.
- By a union bound, with probability at least $1 - \delta$,

$$L_{2,t}\left(f^\alpha\right) - L_{2,t}\left(f_\theta\right) \geq \frac{1}{2}\left\|f^\alpha - f_\theta\right\|_{2,E_t} - 4\sigma^2 \log\left(\left|\mathcal{F}^\alpha\right|/\delta\right) \quad \forall t \in \mathbb{N}, \quad f \in \mathcal{F}^\alpha$$

- Therefore, with probability at least $1 - \delta$ for all $t \in \mathbb{N}$ and $f \in \mathcal{F}$

$$\begin{aligned} L_{2,t}(f) - L_{2,t}\left(f_\theta\right) \geq & \frac{1}{2}\left\|f - f_\theta\right\|_{2,E_t}^2 - 4\sigma^2 \log\left(\left|\mathcal{F}^\alpha\right|/\delta\right) \\ & + \underbrace{\min_{f^\alpha \in \mathcal{F}^\alpha}\left\{\frac{1}{2}\left\|f^\alpha - f_\theta\right\|_{2,E_t}^2 - \frac{1}{2}\left\|f - f_\theta\right\|_{2,E_t}^2 + L_{2,t}(f) - L_{2,t}\left(f^\alpha\right)\right\}}_{\text{Discretization error}}. \end{aligned}$$

Proof of Proposition 2

- Lemma 11 (Discretization error) asserts that with probability at least $1-\delta$, the discretization error is bounded for all t by $\alpha\eta_t$, where $\eta_t := t\left[8C + \sqrt{8\sigma^2 \ln\left(4t^2/\delta\right)}\right]$.
- Since the least squares estimate $\hat{f}_t^{LS}$ has lower squared error than f_θ by definition, we find with probability at least $1-2\delta$

$$\frac{1}{2}\left\|\hat{f}_t^{LS} - f_\theta\right\|_{2,E_t}^2 \leq 4\sigma^2 \log\left(\left|\mathcal{F}^\alpha\right|/\delta\right) + \alpha\eta_t$$

- Equivalently,

$$\left\|\hat{f}_t^{LS} - f_\theta\right\|_{2,E_t} \leq \sqrt{8\sigma^2 \log\left(N\left(\mathcal{F}, \alpha, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha\eta_t} \stackrel{\text{def}}{=} \sqrt{\beta_t^*(\mathcal{F}, \delta, \alpha)}$$

Proof of Lemma 10 for proposition 2 - Exponential martingale

- Consider random variables $(Z_n \mid n \in \mathbb{N})$ adapted to the filtration $(\mathcal{H}_n : n = 0, 1, \ldots)$.
- Assume $\mathbb{E}\left[\exp\left\{\lambda Z_i\right\}\right]$ is finite for all λ.
- Define the conditional mean $\mu_i = \mathbb{E}\left[Z_i \mid \mathcal{H}_{i-1}\right]$.
- We define the conditional cumulant generating function of the centered random variable $[Z_i - \mu_i]$ by $\psi_i(\lambda) = \log \mathbb{E}\left[\exp\left(\lambda\left[Z_i - \mu_i\right]\right) \mid \mathcal{H}_{i-1}\right]$. Let

$$M_n(\lambda) = \exp\left\{\sum_{i=1}^{n} \lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda)\right\}$$

Lemma 12 (Exponential martingale).

$(M_n(\lambda) \mid n \in \mathbb{N})$ *is a martingale, and* $\mathbb{E}M_n(\lambda) = 1$

Lemma 13 (Martingale exponential inequality).

For all $x \geq 0$ *and* $\lambda \geq 0, \mathbb{P}\left(\sum_1^n \lambda Z_i \leq x + \sum_1^n\left[\lambda\mu_i + \psi_i(\lambda)\right], \forall n \in \mathbb{N}\right) \geq 1 - e^{-x}$.

Proof of Lemma 10 for proposition 2

- We set $\mathcal{H}_{t-1}$ to be the σ-algebra generated by (H_t, A_t, θ).
- By assumptions, $\epsilon_t := R_t - f_\theta(A_t)$ satisfies $\mathbb{E}[\epsilon_t \mid \mathcal{H}_{t-1}] = 0$, and $\mathbb{E}[\exp\{\lambda\epsilon_t\} \mid \mathcal{H}_{t-1}] \leq \exp\{(\lambda^2\sigma^2)/2\}$ a.s. for all λ.
- Define $Z_t = (f_\theta(A_t) - R_t)^2 - (f(A_i) - R_t)^2$
- By definition, $\sum_1^T Z_t = L_{2,T+1}(f_\theta) - L_{2,T+1}(f)$.
- Some calculation shows that $Z_t = -(f(A_t) - f_\theta(A_t))^2 + 2(f(A_t) - f_\theta(A_t))\epsilon_t$. Therefore the conditional mean and conditional cumulant generating function satisfy, $\mu_t = \mathbb{E}[Z_t \mid \mathcal{H}_{t-1}] = -(f(A_t) - f_\theta(A_t))^2$

$$\begin{aligned}\psi_t(\lambda) &= \log\mathbb{E}[\exp(\lambda[Z_t - \mu_t]) \mid \mathcal{H}_{t-1}] \\ &= \log\mathbb{E}[\exp(2\lambda(f(A_t) - f_\theta(A_t))\epsilon_t) \mid \mathcal{H}_{t-1}] \leq \frac{(2\lambda[f(A_t) - f_\theta(A_t)])^2\sigma^2}{2}\end{aligned}$$

Proof of Lemma 10 for proposition 2

- Applying Lemma 11 shows that, for all $x \geq 0, \lambda \geq 0$

$$\mathbb{P}\left(\sum_{k=1}^{t} \lambda Z_k \leq x - \lambda \sum_{k=1}^{t} \left(f\left(A_k\right) - f_\theta\left(A_k\right)\right)^2 + \frac{\lambda^2}{2}\left(2f\left(A_k\right) - 2f_\theta\left(A_k\right)\right)^2 \sigma^2 \forall t \in \mathbb{N}\right) \geq 1 - e^{-x}$$

- Or rearranging terms

$$\mathbb{P}\left(\sum_{k=1}^{t} Z_k \leq \frac{x}{\lambda} + \sum_{k=1}^{t} \left(f\left(A_k\right) - f_\theta\left(A_k\right)\right)^2 \left(2\lambda\sigma^2 - 1\right) \forall t \in \mathbb{N}\right) \geq 1 - e^{-x}$$

- Choosing $\lambda = 1/\left(4\sigma^2\right), x = \log(1/\delta)$, and using the definition of $\sum_1^t Z_k$ implies

$$\mathbb{P}\left(L_{2,t}(f) \geq L_{2,t}\left(f_\theta\right) + \frac{1}{2}\left\|f - f_\theta\right\|_{2,E_t}^2 - 4\sigma^2 \log(1/\delta), \forall t \in \mathbb{N}\right) \geq 1 - \delta$$

Proof of Lemma 11 for proposition 2

- Since any two functions in $f, f^\alpha \in \mathcal{F}$ satisfy $\|f - f^\alpha\|_\infty \le C$, it is enough to consider $\alpha \le C$. We find

$$\left|\left(f^\alpha\right)^2(a) - (f)^2(a)\right| \le \max_{-\alpha \le y \le \alpha} \left|(f(a)+y)^2 - f(a)^2\right| = 2f(a)\alpha + \alpha^2 \le 2C\alpha + \alpha^2$$

- which implies

$$\begin{aligned}\left|\left(f^\alpha(a) - f_\theta(a)\right)^2 - \left(f(a) - f_\theta(a)\right)^2\right| &= \left|\left[\left(f^\alpha\right)(a)^2 - f(a)^2\right] + 2f_\theta(a)\left(f(a) - f^\alpha(a)\right)\right| \\ &\le 4C\alpha + \alpha^2 \\ \left|\left(R_t - f(a)\right)^2 - \left(R_t - f^\alpha(a)\right)^2\right| &= \left|2R_t\left(f^\alpha(a) - f(a)\right) + f(a)^2 - f^\alpha(a)^2\right| \\ &\le 2\alpha\left|R_t\right| + 2C\alpha + \alpha^2\end{aligned}$$

Proof of Lemma 11 for proposition 2

- Summing over t, we find that the left-hand side of Lemma 11 is bounded by

$$\sum_{k=1}^{t-1} \left(\frac{1}{2} \left[4C\alpha + \alpha^2 \right] + \left[2\alpha \left| R_k \right| + 2C\alpha + \alpha^2 \right] \right) \leq \alpha \sum_{k=1}^{t-1} \left(6C + 2 \left| R_k \right| \right)$$

- Because ϵ_k is sub-Gaussian, $\mathbb{P}\left(|\epsilon_k| > \sqrt{2\sigma^2 \ln(2/\delta)} \right) \leq \delta$. By a union bound,

$$\mathbb{P}\left(\exists k \in [t-1] \text{ s.t. } |\epsilon_k| > \sqrt{2\sigma^2 \ln\left(4t^2/\delta\right)} \right) \leq \frac{\delta}{2} \sum_{k=1}^{t-1} \frac{1}{t^2} \leq \delta$$

- Since $|R_k| \leq C + |\epsilon_k|$, this shows that with probability at least $1 - \delta$ the discretization error is bounded for all t by $\alpha\eta_t$, where $\eta_t := t \left[8C + 2\sqrt{2\sigma^2 \ln\left(4t^2/\delta\right)} \right]$

Proof of Lemma 12 for Lemma 10

- By definition,

$$\mathbb{E}\left[M_1(\lambda) \mid \mathcal{H}_0\right] = \mathbb{E}\left[\exp\left\{\lambda\left[Z_1 - \mu_1\right] - \psi_1(\lambda)\right\} \mid \mathcal{H}_0\right] = \mathbb{E}\left[\exp\left\{\lambda\left[Z_1 - \mu_1\right]\right\} \mid \mathcal{H}_0\right] / \exp\left\{\psi_1(\lambda)\right\}$$

- Then, for any $n \geq 2$,

$$\begin{aligned}
\mathbb{E}\left[M_n(\lambda) \mid \mathcal{H}_{n-1}\right] &= \mathbb{E}\left[\exp\left\{\sum_{i=1}^{n-1} \lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda)\right\} \exp\left\{\lambda\left[Z_n - \mu_n\right] - \psi_n(\lambda)\right\} \mid \mathcal{H}_{n-1}\right] \\
&= \exp\left\{\sum_{i=1}^{n-1} \lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda)\right\} \mathbb{E}\left[\exp\left\{\lambda\left[Z_n - \mu_n\right] - \psi_n(\lambda)\right\} \mid \mathcal{H}_{n-1}\right] \\
&= \exp\left\{\sum_{i=1}^{n-1} \lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda)\right\} = M_{n-1}(\lambda)
\end{aligned}$$

Proof of lemma 13 for Lemma 10

- For any $\lambda, M_n(\lambda)$ is a martingale with $\mathbb{E}M_n(\lambda) = 1$. Therefore, for any stopping time $\tau, \mathbb{E}M_{\tau\wedge n}(\lambda) = 1$. For arbitrary $x \geq 0$, define $\tau_x = \inf\{n \geq 0 \mid M_n(\lambda) \geq x\}$ and note that τ_x is a stopping time corresponding to the first time M_n crosses the boundary at x.
- Then $\mathbb{E}M_{\tau_r\wedge n}(\lambda) = 1$ and by Markov's inequality,

$$x\mathbb{P}\left(M_{\tau_x\wedge n}(\lambda) \geq x\right) \leq \mathbb{E}M_{\tau_x\wedge n}(\lambda) = 1$$

- Note that the event $\{M_{\tau_x\wedge n}(\lambda) \geq x\} = \bigcup_{k=1}^{n}\{M_k(\lambda) \geq x\}$.
- So we have shown that for all $x \geq 0$ and $n \geq 1$

$$\mathbb{P}\left(\bigcup_{k=1}^{n}\{M_k(\lambda) \geq x\}\right) \leq \frac{1}{x}$$

Proof of lemma 13 for Lemma 10

- For all $x \geq 0$ and $n \geq 1$

$$\mathbb{P}\left(\bigcup_{k=1}^{n}\{M_k(\lambda) \geq x\}\right) \leq \frac{1}{x}$$

- Taking the limit as $n \to \infty$, and applying the monotone convergence theorem shows $\mathbb{P}\left(\bigcup_{k=1}^{\infty}\{M_k(\lambda) \geq x\}\right) \leq 1/x$, or

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty}\{M_k(\lambda) \geq e^x\}\right) \leq e^{-x}.$$

- Recall $M_n(\lambda) = \exp\left\{\sum_{i=1}^{n}\lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda)\right\}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}\left\{\sum_{i=1}^{n}\lambda\left[Z_i - \mu_i\right] - \psi_i(\lambda) \geq x\right\}\right) \leq e^{-x}. \quad \square$$