CS 542 Stats RL: Homework 3

October 16, 2024

Submission deadline: Nov. 1 (Friday) before class.

1. Bisimulation and Bellman-completeness (5 pts)
Let $M = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$ be an MDP and $\phi : \mathcal{S} \to \mathcal{S}_\phi$ be a state abstraction. Let $\mathcal{F}^\phi$ be the set of all possible functions over $\mathcal{S} \times \mathcal{A}$ with value range $[0, V_{\max}]$ ($V_{\max} = R_{\max}/(1-\gamma) > 0$) that are piece-wise constant under ϕ. That is, for any $f \in \mathcal{F}^\phi$, $\forall s^{(1)}, s^{(2)}$ such that $\phi(s^{(1)}) = \phi(s^{(2)})$, we always have $f(s^{(1)}, a) = f(s^{(2)}, a), \forall a \in \mathcal{A}$.

Prove that the following two conditions are equivalent:

1. ϕ is a bisimulation for M.
2. $\mathcal{F}^\phi$ is closed under $\mathcal{T}$, the Bellman optimality operator of M. That is, $\mathcal{T} f \in \mathcal{F}^\phi, \forall f \in \mathcal{F}^\phi$.

Hint: For (2) $\Rightarrow$ (1), try to prove that $\neg$ (1) $\Rightarrow$ $\neg$ (2). That is, if ϕ is not a bisimulation, you should be able to construct $f \in \mathcal{F}^\phi$ such that $\mathcal{T} f \notin \mathcal{F}^\phi$.

2. (5 pts) In the FQE analysis, we assumed that $\|d_t^\pi/\mu\|_\infty$ is bounded for all t, where $\mu \in \Delta(\mathcal{S} \times \mathcal{A})$ is the data distribution. What if we instead assume that $\|d^\pi/\mu\|_\infty$ is bounded, that is, we only cover the discounted occupancy d^π as a whole?

(1) (2 pts) To put things more formally, define $C_t^\pi := \|d_t^\pi/\mu\|_\infty$, and $C^\pi := \|d^\pi/\mu\|_\infty$. Upper bound C_t^π as a function of C^π, and also upper bound C^π as a function of $\{C_t^\pi\}_{t \geq 0}$.

(2) (3 pts) Perform the FQE analysis using C^π (in class what we used is essentially $\max_t C_t^\pi$). To make your life easier, let's assume that FQE produces $f_0, f_1, \ldots, f_K$ that satisfies

$$\|f_k - \mathcal{T}^\pi f_{k-1}\|_{2,\mu} \leq \epsilon, \quad \forall k.$$

Your task is to **give a bound** on $|\mathbb{E}_{s \sim d_0}[f_K(s,\pi)] - J(\pi)|$ as a function of ϵ, γ, K, $V_{\max} = R_{\max}/(1-\gamma)$, and C^π, in a form similar to the bound given in the class. Hint: the easiest way is to start with Eq.(5) in HW2.

3. Refined coverage coefficient (5 pts) In the FQE/FQI analysis, whenever we use the concentrability condition, it is to perform a change of measure in the form of (for FQI $\mathcal{T}^\pi$ should be replaced by $\mathcal{T}$, but the story is similar)

$$\|f - \mathcal{T}^\pi f'\|_{2,d_t^\pi} \leq \sqrt{C_t^\pi}\|f - \mathcal{T}^\pi f'\|_{2,\mu}$$

for some choices of f and f' (e.g., $f = f_k$ and $f' = f_{k-1}$). So naturally, we can replace the definition of C_t^π with the following one, which can be potentially tighter by leveraging the structure of the $\mathcal{F}$ class:[1]

$$C_t^\pi(\mathcal{F}) := \max_{f,f' \in \mathcal{F}} \frac{\|f - \mathcal{T}^\pi f'\|_{2,d_t^\pi}^2}{\|f - \mathcal{T}^\pi f'\|_{2,\mu}^2}. \tag{1}$$

Now consider $C_t^\pi(\mathcal{F})$ in the "linear-completeness" setting, that is,

1. $\mathcal{F}$ is the linear class induced from feature $\phi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^d$, i.e., $\mathcal{F} = \{(s,a) \to \phi(s,a)^\top \theta : \theta \in \mathbb{R}^d\}$. [2]
2. $\mathcal{F}$ that satisfies Bellman-completeness w.r.t. π, i.e., $\mathcal{T}^\pi f \in \mathcal{F} \, \forall f \in \mathcal{F}$.

Let $\sigma_{\min}$ be the smallest eigenvalue of $\Sigma_\mu := \mathbb{E}_{(s,a)\sim\mu}[\phi(s,a)\phi(s,a)^\top] \in \mathbb{R}^{d\times d}$ and assume that

- $\sigma_{\min} > 0$.
- $\|\phi(s,a)\| \leq 1$ (here the norm is the standard L_2 norm for vectors).

Your tasks:

(1) (4 pts) Derive an upper bound on $C_t^\pi(\mathcal{F})$ as a function of $1/\sigma_{\min}$.
Hint: (1) The properties of π and d_t^π do not matter at all: the bound holds even if we replace d_t^π with an arbitrary distribution over $\mathcal{S} \times \mathcal{A}$. [3] (2) For matrix A, its smallest eigenvalue can be written as $\min_{\|x\|=1} x^\top A x$.

(2) (1 pts) The tabular setting is a special case when $d = |\mathcal{S} \times \mathcal{A}|$ and $\phi(s,a) = \mathbf{e}_{(s,a)}$, i.e., a vector with the coordinate indexed by (s,a) being 1 and all other coordinates being 0. Give an explicit expression of $\sigma_{\min}$ as a function of μ.
Hint: what kind of special structure does Σ_μ possess in this case?

Remark For any definition of C^π, the corresponding definition for FQI is typically $C = \max_\pi C^\pi$. When we use the raw density ratio to define C^π, in general C will have to scale with $|\mathcal{A}|$, which cannot handle large action spaces. The result here shows that tightening C^π by leveraging the structure of $\mathcal{F}$ can potentially avoid dependence on $|\mathcal{A}|$, as data distribution μ now only needs to cover the directions occupied by d_t^π in the feature space $\mathbb{R}^d$.

[1] C_π measures how well μ covers policy π, and if we want the analog of concentrability (i.e., covering all policies), we can take $\max_\pi C_\pi$.
[2] Here ϕ is some general state-action feature map, and should not be confused with the ϕ in Q1 which is a state abstraction.
[3] Thus sometimes this bound can be quite loose.

4. (Optional; 3 pts) In the same setting as Q3 (linear complete $\mathcal{F}$), **first show that** $C_t^\pi(\mathcal{F})$ in Eq. 1 has a more refined upper bound, given in matrix form:

$$C_t^\pi(\mathcal{F}) \le \sigma_{\max}(\Sigma_\mu^{-1/2}\Sigma_{d_t^\pi}\Sigma_\mu^{-1/2}), \tag{2}$$

where $\Sigma_{(\cdot)} = \mathbb{E}_{(\cdot)}[\phi\phi^\top]$ is the feature covariance matrix under the distribution specified in the subscript, and $\sigma_{\max}(\cdot)$ is the largest eigenvalue of a matrix. For simplicity we assume that Σ_μ is invertible.

Now, it turns out that a more refined analysis in FQE can further replace $C_t^\pi(\mathcal{F})$ with a tighter quantity, $\bar{C}_t^\pi(\mathcal{F})$ (you can take this statement as given, but it should be clear if you prove Q2 using Eq.(5) from HW2):

$$\bar{C}_t^\pi(\mathcal{F}) := \max_{f,f' \in \mathcal{F}} \frac{(\mathbb{E}_{d_t^\pi}[f - \mathcal{T}^\pi f'])^2}{\|f - \mathcal{T}^\pi f'\|_{2,\mu}^2}.$$

Your second task is to give an upper bound of $\bar{C}_t^\pi(\mathcal{F})$ in matrix form that is analogous to the RHS of Eq.(2), and the bound can depend on $\Sigma_{(\cdot)}$ and $\mathbb{E}_{(\cdot)}[\phi]$ for $(\cdot) = \mu, d_t^\pi$, i.e., the first and second order moments of μ and d_t^π.

After deriving your bound, make a brief comment about how it compares to Eq.(2) qualitatively. Are there situations where one is bounded but the other can be arbitrarily large?