CS 542 Stats RL Homework 2

Name: Kai-Jie Lin
October 31, 2024

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, $\mathcal{T} f \in \mathcal{F}^\phi$, $\forall f \in \mathcal{F}^\phi$.

Proof:
We first show that $(2) \implies (1):$
Suppose ϕ is a bisimulation for M. $\forall f \in Fcal^\phi$, $\forall s^{(1)}, s^{(2)}$ such that $\phi(s^{(1)}) = \phi(s^{(2)})$, we have $R(s^{(1)}, a) = R(s^{(2)}, a)$ and $P(s'|s^{(1)}, a) = P(s'|s^{(2)}, a)$.
$(\mathcal{T} f)(s^{(1)}, a) = R(s^{(1)}, a) + \gamma \left\langle P(\cdot|s^{(1)}, a), V_f \right\rangle = R(s^{(2)}, a) + \gamma \left\langle P(\cdot|s^{(2)}, a), V_f \right\rangle = (\mathcal{T} f)(s^{(2)}, a)$, where $V_f(s) = \sum_{a \in \mathcal{A}} \pi(a|s) f(s, a)$. $\implies \mathcal{T} f \in \mathcal{F}^\phi, \forall f \in \mathcal{F}^\phi$.
Then we show that $(1) \implies (2):$ by proving $\neg(2) \implies \neg(1)$
Suppose ϕ is not a bisimulation for M. $\exists s^{(1)}, s^{(2)}$ such that $\phi(s^{(1)}) = \phi(s^{(2)})$ and $R(s^{(1)}, a) \neq R(s^{(2)}, a)$ or $P(s'|s^{(1)}, a) \neq P(s'|s^{(2)}, a)$.
$(\mathcal{T} f)(s^{(1)}, a) = R(s^{(1)}, a) + \gamma \left\langle P(\cdot|s^{(1)}, a), V_f \right\rangle \neq R(s^{(2)}, a) + \gamma \left\langle P(\cdot|s^{(2)}, a), V_f \right\rangle = (\mathcal{T} f)(s^{(2)}, a)$, where $V_f(s) = \sum_{a \in \mathcal{A}} \pi(a|s) f(s, a)$. $\implies \mathcal{T} f \notin \mathcal{F}^\phi, \forall f \in \mathcal{F}^\phi$.
We have shown that $(1) \iff (2)$.

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 $\{C_t^\pi\}_{t>0}$.

Lemma 1: For any $t > 0$, we have $d^\pi \geq \gamma^{t-1}(1-\gamma)d_t^\pi$, $\forall(s,a)$ pairs.
Proof: $d^\pi = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t d_t^\pi = (1-\gamma)(\sum_{i=0}^{t-1}\gamma^{t-1}d_i^\pi + \gamma^{t-1}d_t^\pi + \sum_{i=t+1}^{\infty}\gamma^{i-1}d_t^\pi) \geq (1-\gamma)\gamma^{t-1}d_t^\pi$.

Claim 1: $C_t^\pi \leq \frac{1}{\gamma^{t-1}(1-\gamma)}C^\pi$.

Proof: $C_t^\pi = \left\|\frac{d_t^\pi}{\mu}\right\|_\infty \leq \frac{1}{\gamma^{t-1}(1-\gamma)}\left\|\frac{d^\pi}{\mu}\right\|_\infty = \frac{1}{\gamma^{t-1}(1-\gamma)}C^\pi$ by Lemma 1.

Claim 2: $C^\pi \leq (1-\gamma)\sum_{t=1}^{\infty}\gamma^{t-1}C_t^\pi$.

Proof: $C^\pi = \left\|\frac{d^\pi}{\mu}\right\|_\infty = \|(1-\gamma)\sum_{t=0}^{\infty}\gamma^t d_t^\pi/\mu\|_\infty \leq (1-\gamma)\sum_{t=1}^{\infty}\gamma^{t-1}\|d_t^\pi/\mu\|_\infty = (1-\gamma)\sum_{t=1}^{\infty}\gamma^{t-1}C_t^\pi$.

(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.

Claim 3:

$$|\hat{J}(\pi) - J(\pi)| \leq \frac{c^\pi K}{(1-\gamma)}\varepsilon$$

Proof:

$$\begin{aligned}
&|\hat{J}(\pi) - J(\pi)| = |\mathbb{E}_{s\sim d_0}[f_k(s,\pi)] - J(\pi)| \\
&= \left|\left(\sum_{t=1}^{K}\gamma^{t-1}\mathbb{E}_{d_t^\pi}[f_{k-t+1} - \mathcal{T}^\pi f_{k-t}]\right) - \mathbb{E}\left[\sum_{t=k+1}^{\infty}\gamma^{t-1}r_t \mid \pi, d_0\right]\right| \\
&\leq \left|\sum_{t=1}^{k}\gamma^{t-1}\mathbb{E}_{d_t^\pi}[f_{k-t+1} - \mathcal{T}^\pi f_{k-t}]\right| = \left|\sum_{t=1}^{k}\gamma^{t-1}\|f_{k-t+1} - \mathcal{T}^\pi f_{k-t}\|_{1,d_t^\pi}\right| \\
&\leq \left|\sum_{t=1}^{K}\gamma^{t-1}\|f_{k-t+1} - \mathcal{T}^\pi f_{k-t}\|_{2,d_t^\pi}\right| \leq \left|\sum_{t=1}^{K}\gamma^{t-1}\sqrt{C_t^\pi}\,\|f_{k-t+1} - \mathcal{T}^\pi f_{k-t}\|_{2,\mu}\right| \\
&\leq \left|\sum_{t=1}^{K}\gamma^{t-1}\frac{C^\pi}{\gamma^{t-1}(1-\gamma)}\varepsilon\right| = \frac{c^\pi K}{(1-\gamma)}\varepsilon
\end{aligned}$$

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)

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

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:

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

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. $\mathcal{F}$ 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

- $\sigma_{\min} > 0$.
- $\|\phi(s,a)\| \le 1$ (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 $1/\sigma_{\min}$.

Claim: $C_t^\pi(F) \le \frac{1}{\sigma_{\min}}$

Proof:

$$\begin{aligned} C_t^\pi(F) &= \max_{f,f'\in\mathcal{F}} \frac{\|f - \mathcal{T}^\pi f'\|_{2,d_t^\pi}^2}{\|f - \mathcal{T}^\pi f'\|_{2,\mu}^2} \\ &= \max_\theta \frac{\theta^\top \Sigma_{d_t^\pi}\theta}{\theta^\top\Sigma_\mu\theta} && \text{(Since linear completeness, } f - T^\pi f' \text{ can be written as } \phi(s,a)^\top\theta \text{ for some } \theta) \\ &=\le \frac{1}{\sigma_{\min}} && \text{(since } |\phi(s,a)| \le 1,\ \theta^\top\Sigma_{d_t^\pi}\theta = \mathbb{E}[(\phi(s,a)^\top\theta)^2] \le |\theta|^2 \le 1) \end{aligned}$$

(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.

$$\Sigma_\mu = \mathbb{E}_{(s,a)-\mu}\left[\phi(s,a)\phi(sa)^\top\right] = \mu I$$
$$\sigma_{\min} = \min_{(x|=1} x^\top\Sigma_\mu x = \min_{(s,a)} \mu$$

The smallest eigenvalue is simply the smallest probability mass assigned to any state-action pair by μ.