CS 542 Stats RL Homework 1

Name: Kai-Jie Lin
October 1, 2024

1. Evaluation error to decision loss (5 pts)

(1) There are K items, $1, 2, \dots, K$. The i-th item has value $v_i \in \mathbb{R}$. Let $v^\star := \max_{i\in[K]} v_i$, where $[K] := \{1, 2, \dots, K\}$, and $i^\star = \arg\max_{i\in[K]} v_i$.

An agent chooses the j-th item, where $j = \arg\max_{i\in[K]} u_i$, and $\{u_i\}_{i=1}^K$ are K real numbers. Let

$$\epsilon := \max_{i\in[K]} |u_i - v_i|.$$

Upper-bound $v^\star - v_j$ as a function of ϵ. Prove your result.

Proof:

$$v^* - v_j = v^* - u_j + u_j - v_j \le v^* - u_j + \epsilon \le v_{i^*} - u_{i^*} + \epsilon \le 2\epsilon \qquad \square$$

(2) If we further have $\forall i$, $u_i \le v_i$, can you improve the bound? What about $\forall i$, $u_i \ge v_i$?

Proof: If $\forall i, u_i \le v_i$, then $v^* - v_j \le v^* - u_j \le v^* - u_{i^*} \le \epsilon$. If $\forall i, u_i \ge v_i$, then $v^* - v_j \le u_j - v_j \le \epsilon$. $\square$

(3) State and prove an upper bound of $|u_j - v^\star|$.

Claim that:

$$|u_j - v^*| \le 3\epsilon.$$

Proof.

$$\begin{aligned} &|u_j - v^*| = |u_j - v_j + v_j - v^*| \\ &\le |u_j - v_j| + |v_j - v^*| && \text{(triangle inequality)} \\ &\le 2\epsilon + \epsilon = 3\epsilon. && \text{(by (1))} \quad \square \end{aligned}$$

2. Loss of using a smaller γ (5 pts)
Suppose we are given an MDP $M = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$. In the lecture we have seen that heavy discounting leads to faster convergence of planning, so we may run some planning algorithm using $\gamma' < \gamma$. The question is, how lossy is the obtained policy when we evaluate it in the original MDP?

More concretely, let's ignore the details of the planning algorithm and say we can compute the optimal policy for $M' = (\mathcal{S}, \mathcal{A}, P, R, \gamma')$, that is, a new MDP that is the same as M in all parameters except that its discount factor is γ' instead of γ. Let $\pi^\star_{\gamma'}$ denote the optimal policy of M'. Prove a bound on $\|V^\star_M - V^{\pi^\star_{\gamma'}}_M\|_\infty$. Here the subscript M is not necessary and only used to emphasize that both value functions are defined in the original M (i.e., w.r.t. γ, not γ'). Your bound should scale with $\gamma - \gamma'$, that is, when γ' is close to γ, the loss will be small. (If your bound is correct but loose by significant factor(s), you may lose up to 1 point.)

Proof:

$$\left\| V^*_M - V^{\pi^*_{\gamma'}}_M \right\|_\infty$$

$$\leq \left\| V^{\pi^*}_M - V^{\pi^*_{\gamma'}}_{M'} \right\|_\infty \qquad (V^{\pi^*_{\gamma'}}_{M'}(s) = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma' r_t | \pi^*_{\gamma'}, s\right] \leq \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma r_t | \pi^*_{\gamma'}, s\right] = V^{\pi^*_\gamma}_{M'}(s))$$

$$\leq \left\| V^{\pi^*}_M - V^{\pi^*}_{M'} \right\|_\infty \qquad (\pi^*_{\gamma'} \text{ is optimal for } M')$$

$$= \left\| \mathbb{E}\left[\sum_{t=2}^{\infty} (\gamma - \gamma')^(t-1) r_t | \pi^*, \cdot\right] \right\|_\infty \leq \frac{R_{\max}}{1 - \gamma + \gamma'}(\gamma - \gamma'). \qquad \square$$

3. Perturbation bound for d^π

Let π_1, π_2 be two policies, and ϵ be their distance, defined as follows:

$$\epsilon := \max_{s\in\mathcal{S}} \|\pi_1(\cdot|s) - \pi_2(\cdot|s)\|_1.$$

That is, we consider the L_1 difference between their action distributions (which is proportional to the TV distance) in each state, and then take the worst-case error over all possible states. You are asked to upper-bound $d^{\pi_1} - d^{\pi_2}$ using ϵ, where $d^\pi = (1-\gamma)(d_0^\top(I-\gamma P^\pi)^{-1})^\top$ is the discounted state occupancy of π when the initial state s_1 is drawn from some initial distribution $d_0 \in \Delta(\mathcal{S})$.

(1) (5 pts) Use ϵ to upper-bound $\|d^{\pi_1} - d^{\pi_2}\|_\infty$. Your bound should go to 0 when $\epsilon = 0$. [1]

Proof:

$$\begin{aligned}
\|d^{\pi_1} - d^{\pi_2}\|_\infty &= (1-\gamma)\left\|\mathbb{E}[\sum_{t=1}^{\infty}\gamma^{t-1}\mathbb{I}[s_t = s]|s_1 \sim d_0, \pi_1] - \mathbb{E}[\sum_{t=1}^{\infty}\gamma^{t-1}\mathbb{I}[s_t = s]|s_1 \sim d_0, \pi_2]\right\|_\infty \\
&= (1-\gamma)\left\|V^{\pi_1} - V^{\pi_2}\right\|_\infty \\
&= \left\|\mathbb{E}_{s'\sim d^{s,\pi_2}}\left[Q^{\pi_1}(s', \pi_1(s')) - Q^{\pi_1}(s', \pi_2(s'))\right]\right\|_\infty & \text{(performance difference lemma)} \\
&\leq \left\|\mathbb{E}_{s'\sim d^{s,\pi_2}}\left[\mathbb{P}(s|s', \cdot)(\pi_1(\cdot|s') - \pi_2(\cdot|s'))\right]\right\|_\infty \\
&\leq |\mathcal{S}| \max_{s\in\mathcal{S}} \|\pi_1(\cdot|s) - \pi_2(\cdot|s)\|_1 = |\mathcal{S}|\epsilon. & \square
\end{aligned}$$

(2) Optional (5 pts) What if we want a bound on $\|d^{\pi_1} - d^{\pi_2}\|_1$? While we can obtain a bound by $\|d^{\pi_1} - d^{\pi_2}\|_1 \leq |\mathcal{S}|\|d^{\pi_1} - d^{\pi_2}\|_\infty$ and plugging the bound from (1) into the RHS, this will incur a dependence on $|\mathcal{S}|$. Can you prove a bound that does not have such a dependence?

Proof:

$$\begin{aligned}
&\|d^{\pi_1} - d^{\pi_2}\|_1 = \|\gamma(P^{\pi_1})^\top d^{\pi_1} - \gamma(P^{\pi_2})^\top d^{\pi_2}\|_1 & (d^\pi = (1-\gamma)d_0 + \gamma(P^\pi)^\top d^\pi) \\
&= \gamma\|(P^{\pi_1})^\top d^{\pi_1} - (P^{\pi_1})^\top d^{\pi_2} + (P^{\pi_1})^\top d^{\pi_2} + (P^{\pi_2})^\top d^{\pi_2}\|_1 \\
&\leq \gamma\|d^{\pi_1} - d^{\pi_2} + (P^{\pi_1} - P^{\pi_2})d^{\pi_2}\|_1 & \text{(data preprocessing inequality)} \\
&\leq \gamma\|d^{\pi_1} - d^{\pi_2}\|_1 + \gamma\|(P^{\pi_1} - P^{\pi_2})^\top d^{\pi_2}\|_1 & \text{(triangle inequality)} \\
&\implies \|d^{\pi_1} - d^{\pi_2}\|_1 \leq \frac{\gamma}{1-\gamma}\|(P^{\pi_1} - P^{\pi_2})^\top d^{\pi_2}\|_1 \\
&= \frac{\gamma}{1-\gamma}\|(\sum_a (\pi_1(a|s) - \pi_2(a|s))P(s'|s,a))^\top d^{\pi_2}\|_1 \\
&\leq \frac{\gamma}{1-\gamma}\epsilon
\end{aligned}$$

[1] Note: this result is strictly inferior to the one in (2). If you choose to work on (2) and are confident about your solution, you can skip this question, as a correct solution to (2) will be automatically counted also as a correct solution to (1).