AI3000: REINFORCEMENT LEARNING

VELMA DHATRI REDDY
AI21BTECH11030

ASSIGNMENT-2

(a) Bellman optimality operator is given by

1) $B(V(S)) = \max_a \left[R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V(s') \right]$

To prove contraction under max-norm we have to show that

$$\| B(V_1) - B(V_2) \|_\infty \leq \beta \| V_1 - V_2 \|_\infty$$

$$\| B(V_1) - B(V_2) \| = \| \max_a (R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_1(s')) - \max_{a_1} (R_s^{a_1} + \gamma \sum_{s' \in S} P_{ss'}^{a_1} (V_2(s')))$$

$\leq \gamma \sum_{s' \in S} P_{ss'}^a | V_1(s') - V_2(s') |$ (we got an equality by replacing a_1 with a_2)

Taking the maximum over s on both sides, we get

$$\| B(V_1) - B(V_2) \|_\infty \leq \gamma \| V_1 - V_2 \|_\infty$$

∴ Hence, proved.

1-(b) We can say from Banach fixed point theorem that the iterative policy evaluation algorithm converges geometrically

It states that <V, ||·||> be a complete normed vector space and let $L: V \to V$ be a γ-contraction mapping. Then iterative application of L converges to a unique fixed point in V independent of the starting point.

As Bellman optimality operator is a γ-contraction the algorithm converges.

Q7 We can prove mathematically by

~~from 1(a) we get~~ $\cancel{\|V_{K+1} - V^*\|_\infty \leq}$

$$\cancel{V_{K+1} - V^*(s) = \max_a \Big[\sum_{s'} P(s'|s,a) [R(s,a,s') + \gamma V_K(s') - \max}$$

$$V_{K+1} = \max_a \Big[\sum_{s' \in S} P_{ss'}^a (R_{ss'}^a + \gamma V_K(s')) \Big]$$

$$V^*(s) = \max_a \Big[\sum_{s' \in S} P_{ss'}^a (R_{ss'}^a + \gamma V_*(s')) \Big]$$

On subtracting the above equation we get

$$V_{K+1} - V^*(s) =$$

$$\max_a \Big[\sum_{s'} P_{ss'}^a (R_{ss'}^a + \gamma V_k(s')) \Big] - \max_{a_1} \Big[\sum_{s'} P_{ss'}^{a_1} (R_{ss'}^{a_1} + \gamma V_*(s')) \Big]$$

$$\leq \max_a \Big[\sum_{s'} P_{ss'}^a \cdot \gamma (V_k(s') - V^*(s'))) \leq \gamma \sum_{s'} P_{ss'} \max_a (V_k(s') - V^*(s'))$$

On applying infinity norm on both sides:

$$\|V_{K+1} - V^*(s)\|_\infty \leq \gamma \| V_K - V^* \| \quad \text{—— (1)}$$

On applying the above inequality till V_1 we get

$$\| V_{K+1} - V^\pi \|_\infty \leq \gamma^k \|V_1 - V^\pi \|_\infty$$

Hence, proved that it converges geometrically.

1.(c) Let $m > n \geq 1$

$$\|V_m - V_n\| = \|V_m - V_{m-1} + V_{m-1} + \cdots + V_{n-1} - V_n\|$$

$$\leq \|V_m - V_{m-1}\| + \|V_{m-1} - V_{m-2}\| + \cdots + \|V_{n-1} - V_n\|$$

As the iteration ~~of~~ value algorithm converges

$$\|V_{n+1} - V_n\| \leq \gamma \|V_n - V_{n-1}\| \leq \cdots \leq \gamma^n \|V_1 - V_0\|$$

$$\|V^* - V_{k+1}\| = \lim_{l \to \infty} \|V_l - V_{k+1}\| \leq \lim_{l \to \infty} \|V_l - V_{l-1}\| + \|V_{l-1} - V_{l-2}\| + \cdots + \|V_k - V_{k+1}\|$$

$$= \lim_{l \to \infty} \gamma^{k+1} \|V_k - V_{k+1}\| \sum_{i=0}^{l-k-1} \gamma^i$$

$$= \frac{\gamma^{k+1} \varepsilon}{1-\gamma}$$

Hence, $\|V^* - V_{k+1}\| \leq \dfrac{\gamma^{k+1} \varepsilon}{1-\gamma}$