AVD894 - Reinforcement learning
Assignment - 1

17-09-2025

SAURABH KUMAR
SC22B146

Proving convergence of the value iteration algorithm for policy evaluation.

Sol^n: $v^\pi(s) = \sum_a \pi(a|s) \sum_r p(r|s,a)\, r + \gamma \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) \cdot v^\pi(s')$

In the matrix form,

$$\bar{v}^\pi = \bar{r}^\pi + \gamma\, p^\pi \bar{v}^\pi$$

To compute $v^\pi(s)$, we have the value iteration algorith that starts with an estimate $v_0(s)\ \forall s \in S$ which is initialized to 0.

Then, for every k we define

$$\bar{v}_{k+1}^\pi = \bar{r}^\pi + \gamma\, p^\pi \bar{v}_k^\pi$$

To prove that :

$$\lim_{k\to\infty} v_k^\pi = v^\pi, \ \forall\ s \in S$$

Define the error, $\delta_k = v_k - v^\pi$.

Substitute $v_{k+1} = \delta_{k+1} + v_\pi$ and $v_k = \delta_k + v_\pi$

into $v_{k+1} = r^\pi + \gamma P^\pi v_k$.

$$\Rightarrow \quad \delta_{k+1} + v^\pi = r^\pi + \gamma P^\pi (\delta_k + v^\pi)$$

$$\Rightarrow \quad \delta_{k+1} = -v^\pi + r^\pi + \gamma P^\pi \delta_k + \gamma P^\pi v^\pi$$

$$= \gamma P^\pi \delta_k - \cancel{v^\pi} + \underbrace{(r^\pi + \gamma P^\pi v^\pi)}_{\cancel{v^\pi}}$$

$$= \gamma P^\pi \delta_k$$

$$= \gamma^2 P^{\pi 2} \delta_{k-1} = \dots = \gamma^{k+1} P^{\pi\, k+1} \delta_0 .$$

As $0 \le P_\pi^k \le 1\ \forall\ k$ and $\gamma < 1$, so $\gamma^k \to 0$,

hence $\delta_{k+1} = \gamma^{k+1} P_\pi^{k+1} \delta_0 \to 0$ as $k \to \infty$.

As $\delta_k \to 0$, $v_k \to v^\pi$.

Contraction Mapping:

Take $f(v_1) = r^{\pi} + \gamma P^{\pi} v_1^{\pi}$

$f(v_2) = r^{\pi} + \gamma P^{\pi} v_2^{\pi}$

$$|f(v_1) - f(v_2)| = |\gamma P^{\pi} v_1^{\pi} - \gamma P^{\pi} v_2^{\pi}|$$

$$\leq \gamma \, |v_1^{\pi} - v_2^{\pi}| \quad , \text{ as } P^{\pi} \text{ sums to } 1.$$

Thus, for $0 \leq \gamma < 1$, the mapping $\bar{v}(\cdot)$ to $\bar{v}_{k+1}(\cdot)$ is a contraction map.

Because the mapping is a contraction, iterating from any starting v_0 gives

$$|v_k - v^{\pi}|_{\infty} \leq \gamma \, |v_{k+1} - v^{\pi}|_{\infty}$$

$$\leq \gamma^{k} \, |v_0 - v^{\pi}|_{\infty}$$

$\therefore \; |v_k - v^{\pi}|_{\infty} \rightarrow 0$ as $k \rightarrow \infty$,

hence $v_k \rightarrow v^{\pi}$ as $k \rightarrow \infty$.