Project Tasks

Q1 Derive Bellman optimal equations for value & action value functions.

$$V_*(s) = \max_{a \in A} q_{\pi_*}(s,a) = \max_{a \in A} E_{\pi_*}[G_t \mid S_t = s, A_t = a]$$

$$= \max_{a \in A} E_{\pi_*}[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a]$$

$$= \max_{a \in A} E_{\pi_*}[R_{t+1} + \gamma V_*(S_{t+1}) \mid S_t = s, A_t = a]$$

$$V_*(s) = \max_{a \in A} \sum_{s',r} p(s', r \mid s, a)[r + \gamma V_*(s')]$$

$$q_*(s,a) = E[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a]$$

$$= E[R_{t+1} + \gamma \max_{a'} q_*(S_{t+1}, a') \mid S_t = s, A_t = a]$$

$$q_*(s,a) = \sum_{s',r} p(s', r \mid s, a)[r + \max_{a'} q_*(s', a')]$$

Q1.1 What is optimality?

→ Refers to best possible policy agent can follow to maximize expected return

Q1.2 What is expectation & how does it relate to bellman eqⁿ?

→ Expectation :- Mean of possible values of random variable

$$E(X) = \sum p(X = x) \cdot x$$

Bellman eqⁿ is recursive relation betⁿ value function which is expected return in a state

Q1.3 Solve exercise 3.25 to 3.29.

→ See ✓ solⁿ of Chapter 3 pdf in repo provided solⁿ in

1 Intution behind Policy Iteration & Value Iteration

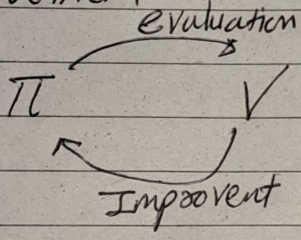
→ How are they working :-
at every point you try out policy & evaluate it., based on evaluation policy is tweaked.

Why should it word

Ans → At every step either policy is kept same or improved. Hence we converge to optimal policy.

2. What is GPI?

→ Interaction of Policy evaluation & Policy Improvement.



Loop continued till policy is stable.

Q4

Policy Iteration Time Complixity $O(S^2A)$
Space $O(S^2)$
Similar for Value Iteration

But Policy Iteration → Few Iteration req
Valie Iteration → More —"— !

For Small MDP → Policy Iteration
Large MDP → Value Iteration.