Quiz - 2.

1. What are the parameters to a state value function?

↳ $$V^{\pi}(s) = E_{\pi}\{R_t | s_t = s\} = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \middle| s_t = s\right\}$$

E_{π} → expected value given that the agent follows policy π

t → time step.

π → policy mapping from each state. $s \in S$

$a \in A$ → action.

2. What are the parameters to an action - value function?

↳ $$Q^{\pi}(s, a) = E_{\pi}\{R_t | s_t = s, a_t = a\}$$

$$= E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \middle| s_t = s, a_t = a\right\}$$

a → action s → state.

π → policy.

3. What is meant by a policy (π)? Is ~~it~~ it deterministic or stochastic?

↳ A policy is a rule used by an agent to decide what actions to take.

Stochastic policy is determined by π

$$a_t \sim \pi_{\theta}(\cdot | s_t)$$

Agent's actions are not fixed but are chosen randomly

4 False

Under an optimal policy π^*, every state must have the highest possible state value compared to any other policy.

5. Stateless → so state transitions not considered

Expected reward : Sum of (Prob of each outcome × Reward of each outcome)

A : $0.9 \times 10 + 0.1 \times 20. = 11$

B : $0.1 \times 10 + 0.9 \times 20 + 0.1 \times (-30) = 17.$

C : $0.1 \times 20 + 0.9 \times (-30) + 0.1 \times 40 = -25$

D : $0.1 \times (-30) + 0.9 \times 40 = 33$

Highest expected reward → D.

By aiming at region D, the agent is expected to maximize its total points over 1000 games.

6. Yes:

References:

- spinningup. openai. com
- incomplete ideas. net