Name: Tazmeen Afroz Roll No: 22P-9252

Section: BAI-6A Reinforcement Learning

Home Task 1

State-Value Function

$$V_\pi(A) = 0.5\,[0.9\,(V_\pi(A))] + 0.25\,[5 + 0.9\,(V_\pi(B))] + 0.25\,[0.9\,(V_\pi(C))]$$

$$V_\pi(B) = 0.5\,[5 + 0.9\,(V_\pi(B))] + 0.25\,[0.9\,(V_\pi(A))] + 0.25\,[0.9\,(V_\pi(D))]$$

$$V_\pi(C) = 0.5\,[0.9\,(V_\pi(C))] + 0.25\,[0.9\,(V_\pi(A))] + 0.25\,[0.9\,V_\pi(D)]$$

$$V_\pi(D) = 0.25\,[5 + 0.9\,(V_\pi(B)] + 0.25\,[0.9\,(V_\pi(C))] + 0.5\,[0.9\,(V_\pi(D))]$$

$V_\pi(A) = \frac{25}{2} = 12.5$ $V_\pi(B) = \frac{325}{22} = 14.77$ Best State

$V_\pi(C) = \frac{225}{22} = 10.22$ $V_\pi(D) = \frac{25}{2} = 12.5$

---

state-action values

$$\sum_{s'} \sum_{r} p(s', r \mid s, a)\,[r + \gamma \sum_{a} \pi(a \mid s')\, q_\pi(s', a')]$$

Actions → L→left, R→Right U→up D→Down

For state A

$$Q_\pi(A,L) = 1 \times [0.9\,(\tfrac{1}{4}[Q_\pi(A,L) + Q_\pi(A,R) + Q_\pi(A,U) + Q_\pi(A,D)]$$

$$Q_\pi(A,R) = [5 + 0.9\,(\tfrac{1}{4}[Q_\pi(B,L) + Q_\pi(B,R) + Q_\pi(B,U) + Q_\pi(B,D)]$$

$$Q_\pi(A,U) = [0.9\,(\tfrac{1}{4}[Q_\pi(A,L) + Q_\pi(A,R) + Q_\pi(A,U) + Q_\pi(A,D))$$

$$Q_\pi(A,D) = [0.9\,(\tfrac{1}{4}[Q_\pi(C,L) + Q_\pi(C,R) + Q_\pi(C,U) + Q_\pi(C,D)])$$

For state B

$$Q_\pi(B,L) = [0.9\,(\tfrac{1}{4}[Q_\pi(A,L) + Q_\pi(A,R) + Q_\pi(A,U) + Q_\pi(A,D)]]$$

$$Q_\pi(B,R) = [5 + 0.9\,(\tfrac{1}{4}[Q_\pi(B,L) + Q_\pi(B,R) + Q_\pi(B,U) + Q_\pi(B,D)]$$

$$Q_\pi(B,U) = [5 + 0.9\,(\tfrac{1}{4}[Q_\pi(B,L) + Q_\pi(B,R) + Q_\pi(B,U) + Q_\pi(B,D)]]$$

$$Q_\pi(B,D) = [0.9\,(\tfrac{1}{4}[Q_\pi(D,L) + Q_\pi(D,R) + Q_\pi(D,U) + Q_\pi(D,D)])$$

For state C

$Q_\pi(C,L) = [0.9(\frac{1}{4}[Q_\pi(C,L) + Q_\pi(C,R) + Q_\pi(C,U) + Q_\pi(C,D)]]$

$Q_\pi(C,R) = [0.9(\frac{1}{4}[Q_\pi(D,L) + Q_\pi(D,R) + Q_\pi(D,U) + Q_\pi(D,D)]])$

$Q_\pi(C,U) = [0.9(\frac{1}{4}[Q_\pi(A,L) + Q_\pi(A,R) + Q_\pi(A,U) + Q_\pi(A,D)]])$

$Q_\pi(C,D) = [0.9(\frac{1}{4}[Q_\pi(C,L) + Q_\pi(C,R) + Q_\pi(C,U) + Q_\pi(C,D)]])$

For state D

$Q_\pi(D,L) = [0.9(\frac{1}{4}[Q_\pi(C,L) + Q_\pi(C,R) + Q_\pi(C,U) + Q_\pi(C,D)]]$

$Q_\pi(D,R) = [0.9(\frac{1}{4}[Q_\pi(D,L) + Q_\pi(D,R) + Q_\pi(D,U) + Q_\pi(D,D)]]$

$Q_\pi(D,U) = [5 + 0.9(\frac{1}{4}[Q_\pi(B,L) + Q_\pi(B,R) + Q_\pi(B,U) + Q_\pi(B,D)]]$

$Q_\pi(D,D) = [0.9(\frac{1}{4}[Q_\pi(D,L) + Q_\pi(D,R) + Q_\pi(D,U) + Q_\pi(D,D)]]$

For A

$Q_\pi(A,L) = 0.9([V(A)]) = 0.9(12.5) = 11.25$

Best → $Q_\pi(A,R) = 5 + 0.9([V(B)]) = 5 + 0.9(14.77) = 18.293$

$Q_\pi(A,U) = [0.9(V(A))] = 0.9(12.5) = 11.25$

$Q_\pi(A,D) = [0.9(V(C))] = 0.9(10.22) = 9.198$

For B

$Q_\pi(B,L) = 0.9(V(A)) = 0.9(12.5) = 11.25$

Best → $Q_\pi(B,R) = [5 + 0.9(V(B))] = 5 + 0.9(14.77) = 18.293$

$Q_\pi(B,U) = [5 + 0.9(V(B))] = 5 + 0.9(14.77) = 18.293$

$Q_\pi(B,D) = [0.9(V(D))] = 0.9(12.5) = 11.25$

For C

$Q_\pi(C,L) = 0.9(V(C)) = 0.9(10.22) = 9.198$

$Q_\pi(C,R) = 0.9(V(D)) = 0.9(12.5) = 11.25$

$Q_\pi(C,U) = 0.9(V(A)) = 0.9(12.5) = 11.25$ Best

$Q_\pi(C,D) = 0.9(V(C)) = 0.9(10.22) = 9.198$

For D

$Q_\pi(D,L) = 0.9(V(C)) = 0.9(10.22) = 9.198$

$Q_\pi(D,R) = 0.9(V(D)) = 0.9(12.5) = 11.25$

Best $Q_\pi(D,U) = 5 + 0.9(V(B)) = 5 + 0.9(14.77) = 18.293$

$Q_\pi(D,D) = 0.9(V(D)) = 0.9(12.5) = 11.25$