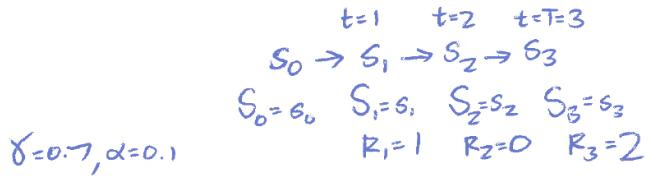


EE4400 Data Engineering and Deep Learning
Tutorial 5 – Reinforcement Learning

Q1.



$$\begin{aligned}
 (a) \quad G_t &= R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots \\
 t=0: \quad G_0 &= R_1 + \gamma R_2 + \gamma^2 R_3 \\
 &= 1 + (0.7)(0) + (0.7)^2(2) = 1.98 \\
 t=1: \quad G_1 &= R_2 + \gamma R_3 \\
 &= (0.7)(2) = 1.4 \\
 t=2: \quad G_2 &= R_3 = 2
 \end{aligned}$$

$$\begin{aligned}
 (b) \quad V(S_t) &\leftarrow V(S_t) + \alpha [R_{t+1} + \gamma V(S_{t+1}) - V(S_t)] \\
 t=0: \quad V(s_0) &\leftarrow V(s_0) + \alpha [R_1 + \gamma V(s_1) - V(s_0)] \\
 &= 0.1 + (0.1)[1 + (0.7)(0.1) - (0.1)] = 0.197 \\
 t=1: \quad V(s_1) &\leftarrow V(s_1) + \alpha [R_2 + \gamma V(s_2) - V(s_1)] \\
 &= 0.1 + (0.1)[0 + (0.7)(0.1) - (0.1)] = 0.097 \\
 t=2: \quad V(s_2) &\leftarrow V(s_2) + \alpha [R_3 + \gamma V(s_3) - V(s_2)] \\
 &= 0.1 + (0.1)[2 + (0.7)(0) - (0.1)] = 0.290
 \end{aligned}$$

s_3 is terminal state

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_{a \in A(S_{t+1})} Q(S_{t+1}, a) - Q(S_t, A_t)]$$

$$\begin{aligned}
 t=0: \quad Q(s_0, R) &\leftarrow Q(s_0, R) + \alpha [R_1 + \gamma \max_{a \in A(s_1)} Q(s_1, a) - Q(s_0, R)] \\
 &= 0.1 + (0.1)[1 + (0.7) \max_{\substack{a \in A(s_1) \\ 0.1}} \{Q(s_1, L), Q(s_1, R)\} - 0.1] = 0.197 \\
 t=1: \quad Q(s_1, R) &\leftarrow Q(s_1, R) + \alpha [R_2 + \gamma \max_{a \in A(s_2)} Q(s_2, a) - Q(s_1, R)] \\
 &= 0.1 + (0.1)[0 + (0.7) \max_{\substack{a \in A(s_2) \\ 0.1}} \{Q(s_2, L), Q(s_2, R)\} - 0.1] = 0.097 \\
 t=2: \quad Q(s_2, R) &\leftarrow Q(s_2, R) + \alpha [R_3 + \gamma \max_{a \in A(s_3)} Q(s_3, a) - Q(s_2, R)] \\
 &= 0.1 + (0.1)[2 + (0.7)(0) - 0.1] = 0.290
 \end{aligned}$$

s_3 is terminal state