National University of Singapore
College of Design & Engineering – ECE

EE4400 Data Engineering and Deep Learning
Tutorial 5 – Reinforcement Learning

Q1.

$t=1 \quad t=2 \quad t=T=3$

$S_0 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3$

$S_0=s_0 \quad S_1=s_1 \quad S_2=s_2 \quad S_3=s_3$

$\gamma=0.7, \alpha=0.1 \qquad R_1=1 \quad R_2=0 \quad R_3=2$

(a) $G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots$

$t=0: \; G_0 = R_1 + \gamma R_2 + \gamma^2 R_3$

$= 1 + (0.7)(0) + (0.7)^2(2) = 1.98$

$t=1: \; G_1 = R_2 + \gamma R_3$

$= (0.7)(2) = 1.4$

$t=2: \; G_2 = R_3 = 2$

(b) $V(S_t) \leftarrow V(S_t) + \alpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$

$t=0: \; V(s_0) \leftarrow V(s_0) + \alpha[R_1 + \gamma V(s_1) - V(s_0)]$

$= 0.1 + (0.1)[1 + (0.7)(0.1) - (0.1)] = 0.197$

$t=1: \; V(s_1) \leftarrow V(s_1) + \alpha[R_2 + \gamma V(s_2) - V(s_1)]$

$= 0.1 + (0.1)[0 + (0.7)(0.1) - (0.1)] = 0.097$

$t=2: \; V(s_2) \leftarrow V(s_2) + \alpha[R_3 + \gamma V(s_3) - V(s_2)]$

$= 0.1 + (0.1)[2 + (0.7)(0) - (0.1)] = 0.290$

(s_3 is terminal state)

Q2. $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_{a \in A(S_{t+1})} Q(S_{t+1}, a) - Q(S_t, A_t)]$

$t=0: \; Q(s_0, R) \leftarrow Q(s_0, R) + \alpha[R_1 + \gamma \max_{a \in A(s_1)} Q(s_1, a) - Q(s_0, R)]$

$= 0.1 + (0.1)[1 + (0.7)\max\{\underbrace{Q(s_1, L)}_{0.1}, \underbrace{Q(s_1, R)}_{0.1}\} - 0.1] = 0.197$

$t=1: \; Q(s_1, R) \leftarrow Q(s_1, R) + \alpha[R_2 + \gamma \max_{a \in A(s_2)} Q(s_2, a) - Q(s_1, R)]$

$= 0.1 + (0.1)[0 + (0.7)\max\{\underbrace{Q(s_2, L)}_{0.1}, \underbrace{Q(s_2, R)}_{0.1}\} - 0.1] = 0.097$

$t=2: \; Q(s_2, R) \leftarrow Q(s_2, R) + \alpha[R_3 + \gamma \max_{a \in A(s_3)} Q(s_3, a) - Q(s_2, R)]$

$= 0.1 + (0.1)[2 + (0.7)(0) - 0.1] = 0.290$

(s_3 is terminal state)