Carissa Ying Geok Teng (A0205190R/E0425113)

1)

$U^{\pi^*}_{t=1}(A) = U^{\pi^*}_{t=0}(A) + \alpha(R(A) + \gamma U^{\pi^*}_{t=0}(A) - U^{\pi^*}_{t=0}(A))$

$= -0.1 + 0.5*(-0.1 + 0.5*(-0.1) – (-0.1))$

$= -0.125$

$U^{\pi^*}_{t=2}(A) = U^{\pi^*}_{t=1}(A) + \alpha(R(A) + \gamma U^{\pi^*}_{t=1}(B) - U^{\pi^*}_{t=1}(A))$

$= -0.125 + 0.5*(-0.1 + 0.5*1 – (-0.125))$

$= 0.1375$

2)

$\pi(s_1) = \text{argmax } Q(s_1, a) = a_1$

$\pi(s_2) = \text{argmax } Q(s_2, a) = a_1$

3)

SARSA:

$Q(s_1, b) = Q(s_1, b) + \alpha(r + \gamma Q(s_2, b) – Q(s_1, b))$

$= 2 + 0.2*(1 + 0.8*2 – 2)$

$= 2.12$

Q-learning:

$Q(s_1, b) = Q(s_1, b) + \alpha(r + \gamma \max Q(s_2, \text{action}) – Q(s_1, b))$

$= 2 + 0.2*(1 + 0.8*4 – 2)$

$= 2.44$