Carissa Ying Geok Teng (A0205190R/E0425113)

1)

Uπ\*t=1(A) = Uπ\*t=0(A) + α(R(A) + γUπ\*t=0(A) - Uπ\*t=0(A))

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

= -0.125

Uπ\*t=2(A) = Uπ\*t=1(A) + α(R(A) + γUπ\*t=1(B) - Uπ\*t=1(A))

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

= 0.1375

2)

π(s1) = argmax Q(s1, a) = a1

π(s2) = argmax Q(s2, a) =a1

3)

SARSA:

Q(s1, b) = Q(s1, b) + α(r + γQ(s2, b) – Q(s1, b))

= 2 + 0.2\*(1 + 0.8\*2 – 2)

= 2.12

Q-learning:

Q(s1, b) = Q(s1, b) + α(r + γ max Q(s2, action) – Q(s1, b))

= 2 + 0.2\*(1 + 0.8\*4 – 2)

= 2.44