Assignment 4)

1) "Return" $G_t = \sum_{i=t+1}^{\infty} \gamma^{i-t-1} \cdot R_i$

"Value Function" $V^{\pi}(s) = \mathbb{E}_{\pi, P_R}[G_t | S_t = s] \quad \forall s \in N, \forall t$

"Action-Value Function" $Q^{\pi}(s,a) = \mathbb{E}_{\pi, P_R}[G_t | (S_t = s, A_t = a)] \quad \forall s \in N, \forall t$

$$\Rightarrow V^{\pi}(s) = \sum_{a \in A} \pi(s,a) \cdot Q^{\pi}(s,a) \quad \forall s \in N$$

$$\Rightarrow Q^{\pi}(s,a) = R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') \cdot V^{\pi}(s') \quad \forall s \in N, a \in A$$

Value Evaluation Algorithm:

$$V_{i+1}(s) = B^*(V_i)(s) \quad \forall s \in N$$

$$= \max_{a \in A} \{ R(s,a) + \gamma \sum_{s' \in N} P(s,a,s') \cdot V_i(s') \}$$

$k=1$

s_1: $V_1 = \max_{a \in A} \{ R(s,a) + \sum_{s' \in N} P(s,a,s') \cdot V_0(s') \}$

a_1: $8 + (0.2 \cdot 10 + 0.6 \cdot 1 + 0.2 \cdot 0) = 10.6$

a_2: $10 + (0.1 \cdot 10 + 0.2 \cdot 1 + 0.7 \cdot 0) = 10.3$

$V_1(s_1) = 10.6$

s_2: a_1: $1.0 + (0.3 \cdot 10 + 0.3 \cdot 1 + 0.4 \cdot 0) = 4.3$

a_2: $-1.0 + (0.5 \cdot 10 + 0.3 \cdot 1 + 0.2 \cdot 0) = 4.3$

$V_1(s_2) = 4.3$

$k=2$:

s_1:

a_1: $8 + (0.2 \cdot 10.6 + 0.6 \cdot 4.3 + 0) = 12.70$

a_2: $10 + (0.1 \cdot 10.6 + 0.2 \cdot 4.3 + 0) = 11.92$

$V_2(s_1) = 12.70$

s_2: a_1: $1.0 + (0.3 \cdot 10.6 + 0.3 \cdot 4.3 + 0) = 5.47$

a_2: $-1.0 + (0.5 \cdot 10.6 + 0.3 \cdot 4.3 + 0) = 5.59$

$V_2(s_2) = 5.59$

$$V_0 = \begin{bmatrix} 10 \\ 1 \\ 0 \end{bmatrix} \qquad V_1 = \begin{bmatrix} 10.6 \\ 4.3 \\ 0 \end{bmatrix} \qquad V_2 = \begin{bmatrix} 12.70 \\ 5.59 \\ 0 \end{bmatrix}$$

$q_1(s_1, a_1) = 10.6$ $\qquad$ $q_1(s_1, a_2) = 10.3$

$q_1(s_2, a_1) = 4.3$ $\qquad$ $q_1(s_2, a_2) = 4.3$

$q_2(s_1, a_1) = 12.70$ $\qquad$ $q_2(s_1, a_2) = 11.92$

$q_2(s_2, a_1) = 5.47$ $\qquad$ $q_2(s_2, a_2) = 5.59$

arbitrary since a_1 and a_2 are equivalent

$\tilde{\pi}_1(s_1) = a_1$ $\qquad$ $\tilde{\pi}_1(s_2) = a_2$

$\tilde{\pi}_2(s_1) = a_1$ $\qquad$ $\tilde{\pi}_2(s_2) = a_2$

$\tilde{\pi}_K(s_1)$ will always be a_1, since $q.(s_1, a_1)$ is strictly larger than $q.(s_1, a_2)$ for nonnegative value function values of s_1 and s_2

$\tilde{\pi}_K(s_2)$ will always be a_2, since in $q.(s_2, a_2)$, a much larger probability is placed on transitioning to s_1, where the value function is very large. This more than makes up for $R(s_2, a_2) < R(s_2, a_1)$