#1 a)

First 2 iterations of the Value iteration alg.:

step 0:

$V_0(s_1) = 10.0$

$V_0(s_2) = 1.0$

$V_0(s_3) = 0.0$

step 1:

$$q_1(s_1, a_1) = R(s_1, a_1) + \gamma\big(P(s_1, a_1, s_1)V_0(s_1) + P(s_1, a_1, s_2)V_0(s_2)\big)$$

$$= 8.0 + 1.0(0.2 \cdot 10.0 + 0.6 \cdot 1.0)$$

$$= 10.6.$$

$$q_1(s_1, a_2) = R(s_1, a_2) + \gamma\big(P(s_1, a_2, s_1)V_0(s_1) + P(s_1, a_2, s_2)V_0(s_2)\big)$$

$$= 10.0 + 1.0(0.1 \cdot 10.0 + 0.2 \cdot 1.0)$$

$$= 11.2.$$

$$q_1(s_2, a_1) = R(s_2, a_1) + \gamma\big(P(s_2, a_1, s_1)V_0(s_1) + P(s_2, a_1, s_2)V_0(s_2)\big)$$

$$= 1.0 + 1.0(0.3 \cdot 10 + 0.3 \cdot 1.0)$$

$$= 4.3$$

$$q_1(s_2, a_2) = R(s_2, a_2) + \gamma\big(P(s_2, a_2, s_1)V_0(s_1) + P(s_2, a_2, s_2)V_0(s_2)\big)$$

$$= -1.0 + 1.0\,(0.5 \cdot 10 + 0.3 \cdot 1.0)$$

$$= 4.3.$$

$$V_1(s_1) = \max_{a \in A}\{q_1(s_1, a)\}$$

$$= 11.2.$$

$$V_1(s_2) = \max_{a \in A}\{q_1(s_2, a)\}$$

$$= 4.3.$$

$\pi_1(s_1) = a_2$

$\pi_1(s_2) = a_1$ (a_1 and a_2 produce an equivalent greedy improvement).

Step 2:

$$q_2(s_1, a_1) = R(s_1, a_1) + \gamma\big(P(s_1, a_1, s_1)V_1(s_1) + P(s_1, a_1, s_2)V_1(s_2)\big)$$

$$= 8.0 + 1.0\,(0.2 \cdot 11.2 + 0.6 \cdot 4.3)$$

$$= 12.82$$

$$q_2(s_1, a_2) = R(s_1, a_2) + \gamma\big(P(s_1, a_2, s_1)\, v_1(s_1) + P(s_1, a_2, s_2)\, v_1(s_2)\big)$$

$$= 10.0 + 1.0(0.1 \cdot 11.2 + 0.2 \cdot 4.3)$$

$$= 11.98$$

$$q_2(s_2, a_1) = R(s_2, a_1) + \gamma\big(P(s_2, a_1, s_1)\, v_1(s_1) + P(s_2, a_1, s_2)\, v_1(s_2)\big)$$

$$= 1.0 + 1.0(0.3 \cdot 11.2 + 0.3 \cdot 4.5)$$

$$= 5.65$$

$$q_2(s_2, a_2) = R(s_2, a_2) + \gamma\big(P(s_2, a_2, s_1)\, v_1(s_1) + P(s_2, a_2, s_2)\, v_1(s_2)\big)$$

$$= -1 + 1(0.5 \cdot 11.2 + 0.3 \cdot 4.3)$$

$$= 5.89$$

$$v_2(s_1) = \max_{a \in A} \{ q_2(s_1, a) \}$$

$$= 12.82$$

$$v_2(s_2) = \max_{a \in A} \{ q_2(s_2, a) \}$$

$$= 5.89$$

$$\pi_2(s_1) = a_1$$

$$\pi_2(s_2) = a_2$$

b) for $k = 3, 4 \ldots$ we have:

$$q_k(s_1, a_1) - q_k(s_1, a_2)$$

$$= \underbrace{R(s_1, a_1) - R(s_1, a_2)}_{=-2} + \underbrace{\gamma}_{=1} \Big(\sum_{s' \in N} \underbrace{\big(P(s_1, a_1, s') - P(s_1, a_2, s')\big)}_{>0} V_{k-1}(s') \Big)$$

$$= -2 + \big(0.1\, V_{k-1}(s_1) + 0.4\, V_{k-1}(s_2) \big)$$

at $k = 3$, $V_{k-1}(s_1) = 12{,}82$, $V_{k-1}(s_2) = 5{,}89$

$\Rightarrow \quad 0{,}1 \cdot 12{,}82 + 0{,}4 \cdot 5{,}89 = 3{,}638 > 2$

$\Rightarrow \quad q_{k=3}(s_1, a_1) - q_{k=3}(s_1, a_2) > 0$.

$\Rightarrow \quad \pi_{k=3}(s_1) = a_1$

Since $V_{k+1}(s) \geq V_k(s)$ for all k, all s by Banach's Thm, then for all $k = 3, 4, \ldots$,

$$q_k(s_1, a_1) - q_k(s_1, a_2) > 0$$

$$\Rightarrow \pi_k(s_1) = a_1$$

$\boxtimes$

Similar argument for $\pi_k(s_2)$.

#3

- state at time t: $s \in S = \{(\text{"employed"}, j_1), \ldots$
$(\text{"employed"}, j_n), (\text{"unemployed"}, j_1) \ldots (\text{"unemployed"}, j_n)\}$
Which represent the state of being "employed" at job i, $i=1,\ldots,n$, or "unemployed" and offered job i, $i=1,\ldots,n$, at the beginning of day t.
- action at time t $a_t \in A = \{\text{"nothing"}, \text{"accept"}, \text{"decline"}\}$

The set of possible actions depends on the state.
We express that as $A(s)$, where $s \in S$.

$A(\text{"unemployed"}, j_i) = \{\text{"accept"}, \text{"decline"}\}$, $i=1,\ldots,n$.
$A(\text{"employed"}, j_i) = \{\text{"nothing"}\}$, $i=1,\ldots,n$.

- $P(s, a, s'): S \times A \times S \to [0,1]$

$$
P(s,a,s') = \begin{cases}
(p_h), & \text{if } s = (\text{"unemployed"}, j_i),\ a = \text{"decline"} \\
 & \quad s' = (\text{"unemployed"}, j_h),\ i = 1, \dots, n;\ h = 1, \dots, n \\
(\alpha \cdot p_h), & \text{if } s = (\text{"unemployed"}, j_i),\ a = \text{"accept"} \\
 & \quad s' = (\text{"unemployed"}, j_h),\ i = 1, \dots, n;\ h = 1, \dots, n \\
(1-\alpha), & \text{if } s = (\text{"unemployed"}, j_i),\ a = \text{"accept"} \\
 & \quad s' = (\text{"employed"}, j_i),\ i = 1, \dots, n \\
(\alpha \cdot p_h), & \text{if } s = (\text{"employed"}, j_i),\ a = \text{"nothing"} \\
 & \quad s' = (\text{"unemployed"}, j_h),\ i = 1, \dots, n;\ h = 1, \dots, n \\
(1-\alpha), & \text{if } s = (\text{"employed"}, j_i),\ a = \text{"nothing"} \\
 & \quad s' = (\text{"employed"}, j_i) \\
0, & \text{else.}
\end{cases}
$$

- $R_T(s,a,s') : S \times A \times S \to \mathbb{R}^+$

- $$
R_T(s,a,s') = \begin{cases}
w_0, & \text{if } s = (\text{"unemployed"}, j_i), \\
 & \quad a = \text{"decline"},\ i = 1, \dots, n \\
w_i, & \text{if } s = (\text{"employed"}, j_i),\ i = 1, \dots, n \\
w_i, & \text{if } s = (\text{"unemployed"}, j_i),\ i = 1, \dots, n, \\
 & \quad a = \text{"accept"}
\end{cases}
$$

- Bellman Optimality Equation:

$$V^*(s) = \max_{a \in A} \left\{ \sum_{s' \in S} \left[P(s,a,s') \left(R_T(s,a,s') + \gamma V^*(s') \right) \right] \right\}$$