Comp 424 Assignment 4

Jonathan Pearce, 260672004

April 14, 2018

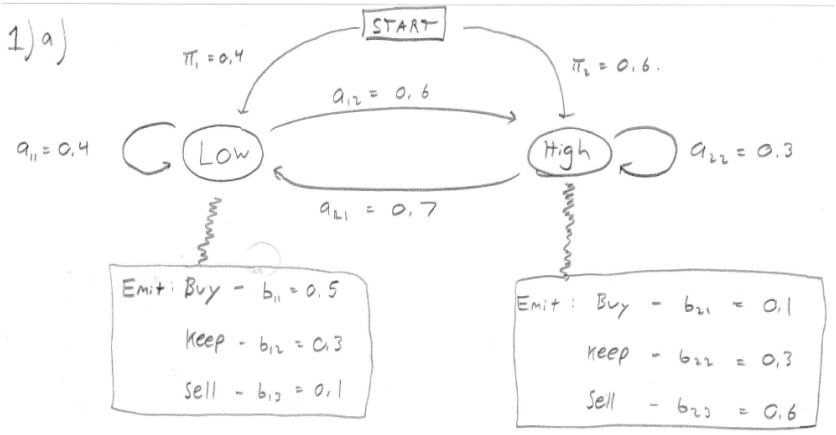
Problem 1a.

Initial Probabilities $\pi_i = Pr(X_1 = i)$	
Low	0.4
High	0.6

Transition Probabilities $a_{ij} = Pr(X_{t+1} = i \mid X_t = j)$		
	Low	High
Low	0.4	0.6
High	0.7	0.3

Emission Probabilities $b_{ik} = Pr(E_t = k \mid X_t = i)$			
	Buy	Keep	Sell
Low	0.5	0.3	0.1
High	0.1	0.3	0.6

Given these tables and values, this is the HMM,



Problem 1b. Using the Forward Algorithm,

$$\alpha_1(1) = \pi_1 \cdot b_{11}$$
$$= 0.4 \cdot 0.5$$
$$= 0.2$$

$$\alpha_2(1) = \pi_2 \cdot b_{21}$$
$$= 0.6 \cdot 0.1$$
$$= 0.06$$

$$\alpha_1(2) = b_{12}(\alpha_1(1) \cdot a_{11} + \alpha_2(1) \cdot a_{21})$$
$$= 0.3(0.2 \cdot 0.4 + 0.06 \cdot 0.7)$$
$$= 0.0366$$

$$\alpha_2(2) = b_{22}(\alpha_1(1) \cdot a_{12} + \alpha_2(1) \cdot a_{22})$$
$$= 0.3(0.2 \cdot 0.6 + 0.06 \cdot 0.3)$$
$$= 0.0414$$

$$\alpha_1(3) = b_{13}(\alpha_1(2) \cdot a_{11} + \alpha_2(2) \cdot a_{21})$$
$$= 0.2(0.0366 \cdot 0.4 + 0.0414 \cdot 0.7)$$
$$= 0.00872$$

$$\alpha_2(3) = b_{23}(\alpha_1(2) \cdot a_{12} + \alpha_2(2) \cdot a_{22})$$
$$= 0.6(0.0366 \cdot 0.6 + 0.0414 \cdot 0.3)$$
$$= 0.0206$$

Therefore,

$$P(E_{1:3} = \{\text{Buy,Keep,Sell}\}) = \alpha_1(3) + \alpha_2(3) = 0.0293$$

Problem 1c.

$$P(X_3 = \text{High}|E_{1:3} = \{\text{Buy,Keep,Sell}\}) = \frac{P(X_3 = \text{High}, E_{1:3} = \{\text{Buy,Keep,Sell}\})}{P(E_{1:3} = \{\text{Buy,Keep,Sell}\})}$$

$$= \frac{\alpha_2(3)}{\alpha_2(3) + \alpha_1(3)}$$
$$= 0.703$$

Problem 1d. Using the Viterbi Algorithm,

$$\begin{aligned}\delta_1(1) &= \pi_1 \cdot b_1(1) \\ &= 0.4 \cdot 0.5 \\ &= 0.2\end{aligned}$$

$$\begin{aligned}\delta_2(1) &= \pi_2 \cdot b_2(1) \\ &= 0.6 \cdot 0.1 \\ &= 0.06\end{aligned}$$

$$\begin{aligned}\delta_1(2) &= Max(b_1(2) \cdot \delta_1(1) \cdot a_{11}, b_1(2) \cdot \delta_2(1) \cdot a_{21}) \\ &= Max(0.3 \cdot 0.2 \cdot 0.4, 0.3 \cdot 0.06 \cdot 0.7) \\ &= 0.024\end{aligned}$$

(backpointer: Low)

$$\begin{aligned}\delta_2(2) &= Max(b_2(2) \cdot \delta_1(1) \cdot a_{12}, b_2(2) \cdot \delta_2(1) \cdot a_{22}) \\ &= Max(0.3 \cdot 0.2 \cdot 0.6, 0.3 \cdot 0.06 \cdot 0.3) \\ &= 0.036\end{aligned}$$

(backpointer: Low)

$$\begin{aligned}\delta_1(3) &= Max(b_1(3) \cdot \delta_1(2) \cdot a_{11}, b_1(3) \cdot \delta_2(2) \cdot a_{21}) \\ &= Max(0.2 \cdot 0.024 \cdot 0.4, 0.2 \cdot 0.036 \cdot 0.7) \\ &= 0.00504\end{aligned}$$

(backpointer: High)

$$\begin{aligned}\delta_2(3) &= Max(b_2(3) \cdot \delta_1(2) \cdot a_{12}, b_2(3) \cdot \delta_2(2) \cdot a_{22}) \\ &= Max(0.6 \cdot 0.024 \cdot 0.6, 0.6 \cdot 0.036 \cdot 0.3) \\ &= 0.00864\end{aligned}$$

(backpointer: Low)

Since $\delta_2(3) > \delta_1(3)$, the third state in the most likely sequence is High. Using the backpointers we find that the most likely sequence of states is,

{Low,Low,High}

Problem 2a.

$$P(\text{accident}) = \sum_{d \in D, s \in S} P(\text{accident} \mid D = d, S = s) P(D = d) P(S = s)$$

$$= 0.8 \cdot 0.9 \cdot 0.1 + 0.8 \cdot 0.1 \cdot 0.2 + 0.2 \cdot 0.9 \cdot 0.3 + 0.2 \cdot 0.1 \cdot 0.5$$

$$= 0.152$$

It follows,

$$EU(\text{insured}) = U(\text{accident} \mid \text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \text{insured}) P(\neg\text{accident})$$

$$= -200 \cdot 0.152 + -10 \cdot (1 - 0.152)$$

$$= -38.88$$

$$EU(\neg\text{insured}) = U(\text{accident} \mid \neg\text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \neg\text{insured}) P(\neg\text{accident})$$

$$= -400 \cdot 0.152 + 0 \cdot (1 - 0.152)$$

$$= -60.8$$

Finally,

$$EU(\text{insured}) - EU(\neg\text{insured}) = 21.92$$

Therefore they should pay 21.92 for insurance.

Problem 2b.

$$P(\text{accident}) = P(\text{accident} \mid S, \neg D)$$

$$= 0.3$$

It follows,

$$EU(\text{insured}) = U(\text{accident} \mid \text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \text{insured}) P(\neg\text{accident})$$

$$= -200 \cdot 0.3 + -10 \cdot (1 - 0.3)$$

$$= -67$$

$$EU(\neg\text{insured}) = U(\text{accident} \mid \neg\text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \neg\text{insured}) P(\neg\text{accident})$$

$$= -400 \cdot 0.3 + 0 \cdot (1 - 0.3)$$

$$= -120$$

Finally,

$$EU(\text{insured}) - EU(\neg\text{insured}) = 53$$

Therefore they should pay 53 for insurance.

Problem 2c.

$$P(\text{accident}) = \sum_{s \in S} P(\text{accident} \mid D, S = s) P(S = s)$$
$$= 0.2 \cdot 0.8 + 0.5 \cdot 0.2$$
$$= 0.26$$

It follows,

$$EU(\text{insured}) = U(\text{accident} \mid \text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \text{insured}) P(\neg\text{accident})$$
$$= -200 \cdot 0.26 + -10 \cdot (1 - 0.26)$$
$$= -59.4$$
$$EU(\neg\text{insured}) = U(\text{accident} \mid \neg\text{insured}) P(\text{accident}) + U(\neg\text{accident} \mid \neg\text{insured}) P(\neg\text{accident})$$
$$= -400 \cdot 0.26 + 0 \cdot (1 - 0.26)$$
$$= -104$$

Finally,

$$EU(\text{insured}) - EU(\neg\text{insured}) = 44.6$$

Therefore they should pay 44.6 for insurance.

Problem 2d. from part a we have $P(\text{accident}) = 0.152$,

$$EU(\text{insured}) = U(\text{accident} \mid \text{insured,covered}) P(\text{accident}) P(\text{covered})$$
$$+U(\text{accident} \mid \text{insured,not covered}) P(\text{accident}) P(\text{not covered}) + U(\neg\text{accident} \mid \text{insured}) P(\neg\text{accident})$$
$$= -200 \cdot 0.152 \cdot 0.80 + -400 \cdot 0.152 \cdot 0.20 + -10 \cdot (1 - 0.152)$$
$$= -44.96$$

The expected utility if we are uninsured does not change from part a, therefore

$$EU(\neg\text{insured}) = -60.8$$

Finally,

$$EU(\text{insured}) - EU(\neg\text{insured}) = 15.84$$

Therefore they should charge 15.84 for insurance.

Problem 3a.
A policy for this MDP would consist of 6 actions, 1 action for each state, $(a_1, a_2, a_3, a_4, a_5, a_6)$. there are 4 options for each action, therefore the total number of policies for this MDP is $6^4 = 1296$.

Problem 3b. From the Bellman equations we have,

$$V^0(s) = R(s, \pi^0(s)) + \gamma \sum_{s\prime \in S} T(s, \pi^0(s), s')V^0(s')$$

Therefore,

$$V^0(s_1) = 0 + 0.9(0.2 \cdot V^0(s_1) + 0.8 \cdot V^0(s_2))$$
$$V^0(s_2) = 0 + 0.9(0.2 \cdot V^0(s_2) + 0.8 \cdot V^0(s_3))$$
$$V^0(s_3) = 20 + 0.9(V^0(s_3))$$
$$V^0(s_4) = 10 + 0.9(0.2 \cdot V^0(s_4) + 0.8 \cdot V^0(s_5))$$
$$V^0(s_5) = 0 + 0.9(0.2 \cdot V^0(s_5) + 0.8 \cdot V^0(s_6))$$
$$V^0(s_6) = -10 + 0.9(V^0(s_6))$$

Solving this system of equations,

$$V^0(s_1) = 154.19$$
$$V^0(s_2) = 175.61$$
$$V^0(s_3) = 200$$
$$V^0(s_4) = -64.9$$
$$V^0(s_5) = -87.8$$
$$V^0(s_6) = -100$$

Problem 3c. Using the policy iteration equation we have,

$$\pi'(s) = argmax_{a \in A}(R(s, a) + \gamma \sum_{s\prime \in S} T(s, a, s')V^0(s'))$$

Therefore,

$$\pi_1'(s_1) = argmax_{a \in A} \begin{cases} 0 + 0.9(0.2 \cdot 154.19 + 0.8 \cdot (-64.9)) \text{ (a = Down)} \\ 0 + 0.9(0.2 \cdot 154.19 + 0.8 \cdot 175.61) \text{ (a = Right)} \\ 0 + 0.9(154.19) \text{ (a = Left or Up)} \end{cases}$$

$$\Rightarrow \pi_1'(s_1) = \text{Right}$$

$$\pi_1'(s_2) = argmax_{a \in A} \begin{cases} 0 + 0.9(0.2 \cdot 175.61 + 0.8 \cdot (-87.8)) \text{ (a = Down)} \\ 0 + 0.9(0.2 \cdot 175.61 + 0.8 \cdot 200) \text{ (a = Right)} \\ 0 + 0.9(0.2 \cdot 175.61 + 0.8 \cdot 154.19) \text{ (a = Left)} \\ 0 + 0.9(175.61) \text{ (a = Up)} \end{cases}$$

$$\Rightarrow \pi_1'(s_2) = \text{Right}$$

$$\pi_1'(s_3) = argmax_{a\in A} \begin{cases} 20 + 0.9(0.2 \cdot 200 + 0.8 \cdot 175.61) \text{ (a = Left)} \\ 20 + 0.9(200) \text{ (a = Right, Down or Up)} \end{cases}$$
$$\Rightarrow \pi_1'(s_3) = \text{Down}$$

$$\pi_1'(s_4) = argmax_{a\in A} \begin{cases} 10 + 0.9(0.2 \cdot (-64.9) + 0.8 \cdot (-87.8)) \text{ (a = Right)} \\ 10 + 0.9(-64.9) \text{ (a = Left, Down or Up)} \end{cases}$$
$$\Rightarrow \pi_1'(s_4) = \text{Down}$$

$$\pi_1'(s_5) = argmax_{a\in A} \begin{cases} 0 + 0.9(0.2 \cdot (-87.8) + 0.8 \cdot (-64.9)) \text{ (a = Left)} \\ 0 + 0.9(0.2 \cdot (-87.8) + 0.8 \cdot (-100)) \text{ (a = Right)} \\ 0 + 0.9(0.2 \cdot (-87.8) + 0.8 \cdot 175.61) \text{ (a = Up)} \\ 0 + 0.9(-87.8) \text{ (a = Down)} \end{cases}$$
$$\Rightarrow \pi_1'(s_5) = \text{Up}$$

$$\pi_1'(s_6) = argmax_{a\in A} \begin{cases} -10 + 0.9(0.2 \cdot (-100) + 0.8 \cdot (-87.8)) \text{ (a = Left)} \\ -10 + 0.9(0.2 \cdot (-100) + 0.8 \cdot 200) \text{ (a = Up)} \\ -10 + 0.9(-100) \text{ (a = Down or Right)} \end{cases}$$
$$\Rightarrow \pi_1'(s_6) = \text{Up}$$

Finally,

$$\pi_1'(s) = \{\text{Right,Right,Down,Down,Up,Up}\}$$

visually,

$\Rightarrow$	$\Rightarrow$	$\Downarrow$
$\Downarrow$	$\Uparrow$	$\Uparrow$

Problem 3d. Iteration 2:

$$V^1(s_1) = 0 + 0.9(0.2 \cdot V^1(s_1) + 0.8 \cdot V^1(s_2))$$
$$V^1(s_2) = 0 + 0.9(0.2 \cdot V^1(s_2) + 0.8 \cdot V^1(s_3))$$
$$V^1(s_3) = 20 + 0.9(V^1(s_3))$$
$$V^1(s_4) = 10 + 0.9(V^1(s_4))$$
$$V^1(s_5) = 0 + 0.9(0.2 \cdot V^1(s_5) + 0.8 \cdot V^1(s_2))$$
$$V^1(s_6) = -10 + 0.9(0.2 \cdot V^1(s_6) + 0.8 \cdot V^1(s_3))$$

Solving this system of equations,

$$V^1(s_1) = 154.19$$
$$V^1(s_2) = 175.61$$

$$V^1(s_3) = 200$$
$$V^1(s_4) = 100$$
$$V^1(s_5) = 154.19$$
$$V^1(s_6) = 163.41$$

Further,

$$\pi_2'(s) = \{\text{Right,Right,Down,Right,Up,Up}\}$$

visually,

$\Rightarrow$	$\Rightarrow$	$\Downarrow$
$\Rightarrow$	$\Uparrow$	$\Uparrow$

Iteration 3:

$$V^2(s_1) = 0 + 0.9(0.2 \cdot V^2(s_1) + 0.8 \cdot V^2(s_2))$$
$$V^2(s_2) = 0 + 0.9(0.2 \cdot V^2(s_2) + 0.8 \cdot V^2(s_3))$$
$$V^2(s_3) = 20 + 0.9(V^2(s_3))$$
$$V^2(s_4) = 10 + 0.9(0.2 \cdot V^2(s_4) + 0.8 \cdot V^2(s_5))$$
$$V^2(s_5) = 0 + 0.9(0.2 \cdot V^2(s_5) + 0.8 \cdot V^2(s_2))$$
$$V^2(s_6) = -10 + 0.9(0.2 \cdot V^2(s_6) + 0.8 \cdot V^2(s_3))$$

Solving this system of equations,

$$V^2(s_1) = 154.19$$
$$V^2(s_2) = 175.61$$
$$V^2(s_3) = 200$$
$$V^2(s_4) = 147.58$$
$$V^2(s_5) = 154.19$$
$$V^2(s_6) = 163.41$$

Further,

$$\pi_2'(s) = \{\text{Right,Right,Down,Right,Up,Up}\}$$

visually,

$\Rightarrow$	$\Rightarrow$	$\Downarrow$
$\Rightarrow$	$\Uparrow$	$\Uparrow$

Iteration 4:

$$V^3(s_1) = 0 + 0.9(0.2 \cdot V^3(s_1) + 0.8 \cdot V^3(s_2))$$
$$V^3(s_2) = 0 + 0.9(0.2 \cdot V^3(s_2) + 0.8 \cdot V^3(s_3))$$
$$V^3(s_3) = 20 + 0.9(V^3(s_3))$$
$$V^3(s_4) = 10 + 0.9(0.2 \cdot V^3(s_4) + 0.8 \cdot V^3(s_5))$$
$$V^3(s_5) = 0 + 0.9(0.2 \cdot V^3(s_5) + 0.8 \cdot V^3(s_2))$$
$$V^3(s_6) = -10 + 0.9(0.2 \cdot V^3(s_6) + 0.8 \cdot V^3(s_3))$$

Solving this system of equations,

$$V^3(s_1) = 154.19$$
$$V^3(s_2) = 175.61$$
$$V^3(s_3) = 200$$
$$V^3(s_4) = 147.58$$
$$V^3(s_5) = 154.19$$
$$V^3(s_6) = 163.41$$

Therefore the policy values for iteration 3 and iteration 4 are the same, thus we have reached convergence and the optimal value function is,

$$V^\star(s) = \{154.19, 175.61, 200, 147.58, 154.19, 163.41\}$$

Problem 3e. This optimal value function is unique because the MDP is finite

Problem 3f. from part d, the optimal policy is

$$\pi^\star(s) = \{\text{Right,Right,Down,Right,Up,Up}\}$$

visually,

$\Rightarrow$	$\Rightarrow$	$\Downarrow$
$\Rightarrow$	$\Uparrow$	$\Uparrow$

Problem 3g. Consider the following policy

$$\pi^\star(s) = \{\text{Right,Right,Right,Right,Up,Up}\}$$

visually,

$\Rightarrow$	$\Rightarrow$	$\Rightarrow$
$\Rightarrow$	$\Uparrow$	$\Uparrow$

This new policy is still optimal because in state 3 the action Down and Right are equivalent since both guarantee the agent stays in state 3. Therefore the optimal policy is not unique.

Problem 3h. Increasing the reward at state 3 would effect the optimal value function, however it would not change the policy as we would still want to reach state 3 as quick as possible.

Problem 4a.

$t = 1$,

$$Q_1(3) = 3 + \frac{(2-3)}{1} = 2$$
$$\rightarrow Q_1 = \{1, 1, 2, 1, 1\}$$

$t = 2$,

$$Q_2(5) = 1 + \frac{(0-1)}{1} = 0$$
$$\rightarrow Q_2 = \{1, 1, 2, 1, 0\}$$

$t = 3$,

$$Q_3(3) = 2 + \frac{(1-2)}{2} = 1.5$$
$$\rightarrow Q_3 = \{1, 1, 1.5, 1, 0\}$$

$t = 4$,

$$Q_4(1) = 1 + \frac{(0-1)}{1} = 0$$
$$\rightarrow Q_4 = \{0, 1, 1.5, 1, 0\}$$

$t = 5$,

$$Q_5(3) = 1.5 + \frac{(0-1.5)}{3} = 1$$
$$\rightarrow Q_5 = \{0, 1, 1, 1, 0\}$$

Problem 4b. If an optimal action was taken, then either the player choose to pull the best arm or the best arm was choose randomly. The optimal action was selected at $t = 1$,$t = 3$ and $t = 5$. Therefore we can only conclude that a random action was taken when the player selects a non optimal arm, this occurs at $t = 2$ and $t = 4$.