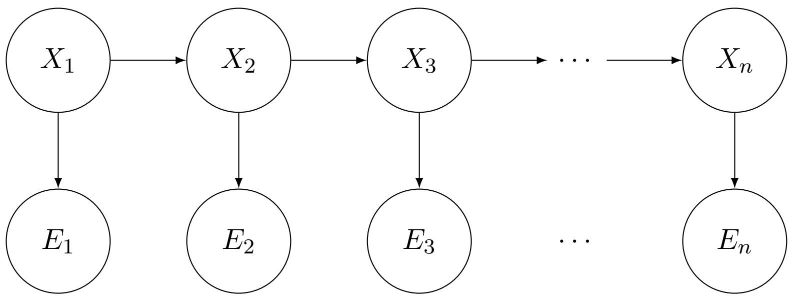
COMP 424 Assignment 4 Solutions

Ariella Smofsky

April 16, 2017

Q1 (a)



Where

$$X_t \in \{Well, Unwell\} \qquad (t = 1, ..., n)$$
$$E_t \in \{Normal, Abnormal\} \qquad (t = 1, ..., n)$$

Initial Probability

$$P(X_1 = Well) = 0.9$$
$$P(X_1 = Unwell) = 0.1$$

State Transition Probabilities

$P(X_{t+1} \mid X_t)$	$X_{t+1} = Well$	$X_{t+1} = Unwell$
$X_t = Well$	0.8	0.2
$X_t = Unwell$	0.5	0.5

Emission Probabilities

$P(E_t \mid X_t)$	$E_t = Normal$	$E_t = Abnormal$
$X_t = Well$	0.6	0.4
$X_t = Unwell$	0.3	0.7

(b) Let $x_t := (X_t = Well)$ and $\neg x_t := (X_t = Unwell)$. Similarly, let $e_t := (E_t = Normal)$ and $\neg e_t := (E_t = Abnormal)$. We want to calculate

$$P(\neg e_1, e_2, e_3)$$

There exists several methods to calculate this probability, but we choose to use the Forward Algorithm.

	$\neg e_1$	e_2	e_3
x	$\alpha_1(1)$	$\alpha_1(2)$	$\alpha_1(3)$
$\neg x$	$\alpha_2(1)$	$\alpha_2(2)$	$\alpha_2(3)$

We now calculate the cell entries.

$$\begin{aligned}
\alpha_1(1) &= P(x_1) \times P(\neg e_1|x_1) \\
&= (0.9) \times (0.4) \\
&= 0.36 \\
\alpha_2(1) &= P(\neg x_1) \times P(\neg e_1|\neg x_1) \\
&= (0.1) \times (0.7) \\
&= 0.07 \\
\alpha_1(2) &= [P(e_2|x_2) \times P(x_2|x_1) \times \alpha_1(1)] + [P(e_2|x_2) \times P(x_2|\neg x_1) \times \alpha_2(1)] \\
&= [0.6 \times 0.8 \times 0.36] + [0.6 \times 0.5 \times 0.07] \\
&= 0.1938 \\
\alpha_2(2) &= [P(e_2|\neg x_2) \times P(\neg x_2|x_1) \times \alpha_1(1)] + [P(e_2|\neg x_2) \times P(\neg x_2|\neg x_1) \times \alpha_2(1)] \\
&= [0.3 \times 0.2 \times 0.36] + [0.3 \times 0.5 \times 0.07] \\
&= 0.0321 \\
\alpha_1(3) &= [P(e_3|x_3) \times P(x_3|x_2) \times \alpha_1(2)] + [P(e_3|x_3) \times P(x_3|\neg x_2) \times \alpha_2(2)] \\
&= [0.6 \times 0.8 \times 0.1938] + [0.6 \times 0.5 \times 0.0321] \\
&= 0.102654 \\
\alpha_2(3) &= [P(e_3|\neg x_3) \times P(\neg x_3|x_2) \times \alpha_1(2)] + [P(e_3|\neg x_3) \times P(\neg x_3|\neg x_2) \times \alpha_2(2)] \\
&= [0.3 \times 0.2 \times 0.1938] + [0.3 \times 0.5 \times 0.0321] \\
&= 0.016443 \\
P(\neg e_1, e_2, e_3) &= \alpha_1(3) + \alpha_2(3) \\
&= 0.102654 + 0.016443 \\
&= 0.119097
\end{aligned}$$

Thus, the probability of observing $\{Abnormal, Normal, Normal\}$ is 0.119097.

(c) We want to calculate $P(\neg x_4|\neg e_1, e_2, e_3)$. First, apply filtering

$$P(x_3|\neg e_1, e_2, e_3) = \frac{\alpha_1(3)}{P(\neg e_1, e_2, e_3)} = \frac{0.102654}{0.119097} = 0.861936$$
$$P(\neg x_3|\neg e_1, e_2, e_3) = \frac{\alpha_2(3)}{P(\neg e_1, e_2, e_3)} = \frac{0.016443}{0.119097} = 0.138064$$

Now, use prediction

$$\begin{aligned} P(\neg x_4|\neg e_1, e_2, e_3) &= [P(\neg x_4|x_3) \times P(x_3|\neg e_1, e_2, e_3)] + [P(\neg x_4|\neg x_3) \times P(\neg x_3|\neg e_1, e_2, e_3)] \\ &= [0.2 \times 0.861936] + [0.5 \times 0.138064] \\ &= 0.241419 \end{aligned}$$

Thus, the probability of the patient being unwell after observing $\{Abnormal, Normal, Normal\}$ is 0.241419.

(d) Let us use the Viterbi Algorithm.

	$\neg e_1$	e_2	e_3
x	$\delta_1(1)$	$\delta_1(2)$	$\delta_1(3)$
$\neg x$	$\delta_2(1)$	$\delta_2(2)$	$\delta_2(3)$

$$\begin{aligned} \delta_1(1) &= \alpha_1(1) = 0.36 \\ \delta_2(1) &= \alpha_2(1) = 0.07 \\ \delta_1(2) &= \max(\delta_1(1) \times P(x_2|x_1) \times P(e_2|x_2),\ \delta_2(1) \times P(x_2|\neg x_1) \times P(e_2|x_2)) \\ &= \max(0.36 \times 0.8 \times 0.6,\ 0.07 \times 0.5 \times 0.6) \\ &= \max(0.1728,\ 0.021) \\ &= 0.1728 \text{ from } \delta_1(1) \\ \delta_2(2) &= \max(\delta_1(1) \times P(\neg x_2|x_1) \times P(e_2|\neg x_2),\ \delta_2(1) \times P(\neg x_2|\neg x_1) \times P(e_2|\neg x_2)) \\ &= \max(0.36 \times 0.2 \times 0.3,\ 0.07 \times 0.5 \times 0.3) \\ &= \max(0.0216,\ 0.0105) \\ &= 0.0216 \text{ from } \delta_1(1) \\ \delta_1(3) &= \max(\delta_1(2) \times P(x_3|x_2) \times P(e_3|x_3),\ \delta_2(2) \times P(x_3|\neg x_2) \times P(e_3|x_3)) \\ &= \max(0.1728 \times 0.8 \times 0.6,\ 0.0216 \times 0.5 \times 0.6) \\ &= \max(0.082944,\ 0.00648) \\ &= 0.082944 \text{ from } \delta_1(2) \end{aligned}$$

$$\begin{aligned}\delta_2(3) &= \max(\delta_1(2) \times P(\neg x_3|x_2) \times P(e_3|\neg x_3),\ \delta_2(2) \times P(\neg x_3|\neg x_2) \times P(e_3|\neg x_3)) \\ &= \max(0.1728 \times 0.2 \times 0.3,\ 0.0216 \times 0.5 \times 0.3) \\ &= \max(0.010368,\ 0.00324) \\ &= 0.010368 \text{ from } \delta_1(2)\end{aligned}$$

Now take the maximum of the last column

$$\begin{aligned}\max(\delta_1(3), \delta_2(3)) &= \max(0.082944, 0.010368) \\ &= 0.082944 \text{ from } \delta_1(3)\end{aligned}$$

Trace back the entries in the table to obtain $\delta_1(3)$ from $\delta_1(2)$ from $\delta_1(1)$. Thus, $\{x_1, x_2, x_3\}$ is the most likely sequence of three states to explain $\{\neg e_1, e_2, e_3\}$.

Q2 (a) Note: expected value of information = (expected value of best action given the information) - (expected value of best action without the information)

Assume the given information is whether you have chickenpox. Then your best action has utility

$$\begin{aligned}EU(info) &= P(C) \times U(treatment|C) + P(\neg C) \times U(\neg treatment|\neg C) \\ &= 0.05 \times (-100) + 0.95 \times 0 \\ &= -5\end{aligned}$$

Now assume you do not have information about whether you have chickenpox.

$$\begin{aligned}EU(treatment|no\ info) &= P(C) \times U(treatment|C) + P(\neg C) \times U(treatment|\neg C) \\ &= 0.05 \times (-100) + 0.95 \times (-20) \\ &= -24 \\ EU(\neg treatment|no\ info) &= P(C) \times U(\neg treatment|C) + P(\neg C) \times U(\neg treatment|\neg C) \\ &= 0.05 \times (-200) + 0.95 \times 0 \\ &= -10\end{aligned}$$

Thus, your best action is to not get treatment. Now we can calculate the expected value of information

$$\begin{aligned}EU(info) - EU(\neg treatment|no\ info) &= (-5) - (-10) \\ &= 5\end{aligned}$$

We can conclude you should pay 5 units to find out whether you have chickenpox.

(b) We have information about symptoms, so we calculate the probability of having chickenpox given this information.

$$\begin{aligned}
P(C|F,\neg R) &= \frac{P(F,\neg R|C)P(C)}{P(F,\neg R)} \\
&= \frac{P(F|C)P(\neg R|C)P(C)}{P(F|C)P(\neg R|C)P(C) + P(F|\neg C)P(\neg R|\neg C)P(\neg C)} \\
&= \frac{0.6 \times 0.1 \times 0.05}{0.6 \times 0.1 \times 0.05 + 0.1 \times 0.8 \times 0.95} \\
&= 0.03797
\end{aligned}$$

Now calculate the expected utility as in part (a)

$$\begin{aligned}
EU(info|F,\neg R) &= P(C|F,\neg R) \times U(treatment|C) + P(\neg C|F,\neg R) \\
&\quad \times U(\neg treatment|\neg C) \\
&= 0.03797 \times (-100) + 0.96203 \times 0 \\
&= -3.797 \\
EU(treatment|no\ info, F, \neg R) &= P(C|F,\neg R) \times U(treatment|C) + P(\neg C|F,\neg R) \\
&\quad \times U(treatment|\neg C) \\
&= 0.03797 \times (-100) + 0.96203 \times (-20) \\
&= -23.0376 \\
EU(\neg treatment|no\ info, F, \neg R) &= P(C|F,\neg R) \times U(\neg treatment|C) + P(\neg C|F,\neg R) \\
&\quad \times U(\neg treatment|\neg C) \\
&= 0.03797 \times (-200) + 0.96203 \times 0 \\
&= -7.594
\end{aligned}$$

Calculate the value of information

$$\begin{aligned}
EU(info|F,\neg R) - EU(\neg treatment|no\ info, F, \neg R) &= (-3.797) - (-7.594) \\
&= 3.797
\end{aligned}$$

Thus, you should pay 3.797 units to know whether or not you have chickenpox.

(c) We have information about symptoms, so we calculate the new probability of having chickenpox

$$\begin{aligned}
P(C|R) &= \frac{P(R|C)P(C)}{P(R)} \\
&= \frac{P(R|C)P(C)}{P(R|C)P(C) + P(R|\neg C)P(\neg C)} \\
&= \frac{0.9 \times 0.05}{0.9 \times 0.05 + 0.2 \times 0.95} \\
&= 0.1915
\end{aligned}$$

Now calculate the expected utility

$$\begin{aligned}
EU(info|R) &= P(C|R) \times U(treatment|C) + P(\neg C|R) \\
&\quad \times U(\neg treatment|\neg C) \\
&= 0.1915 \times (-100) + 0.8085 \times 0 \\
&= -19.15 \\
EU(treatment|no\ info, R) &= P(C|R) \times U(treatment|C) + P(\neg C|R) \\
&\quad \times U(treatment|\neg C) \\
&= 0.1915 \times (-100) + 0.8085 \times (-20) \\
&= -35.32 \\
EU(\neg treatment|no\ info, R) &= P(C|R) \times U(\neg treatment|C) + P(\neg C|R) \\
&\quad \times U(\neg treatment|\neg C) \\
&= 0.1915 \times (-200) + 0.8085 \times 0 \\
&= -38.3
\end{aligned}$$

Calculate the value of information

$$\begin{aligned}
EU(info|R) - EU(\neg treatment|no\ info, R) &= (-19.15) - (-35.32) \\
&= 16.17
\end{aligned}$$

Thus, you should pay 16.17 units to know whether or not you have chickenpox.

(d) Let A = knowing result of 100% accurate test
and B = knowing result of uncertain test
We want the value of perfect information of test A to equal the value of perfect

information of test B. Thus

$$EU(A) - 5 = EU(B) - c$$

where 5 is the cost of knowing whether you have chickenpox with test A and c is the cost of test B. In part (a) we determined $EU(A) = -5$. So we can re-write the above equation as

$$c = EU(B) + 10$$

Now, we calculate $EU(B)$.

$$\begin{aligned} EU(B) = {} & U(best\ action|treatment)P(treatment) \\ & + U(best\ action|\neg treatment)P(\neg treatment) \end{aligned}$$

First, considering test B, calculate the probabilities of a positive test result t_+ and a negative test result t_-

$$\begin{aligned} P(t_+) &= P(t_+|C)P(C) + P(t_+|\neg C)P(\neg C) \\ &= 0.8 \times 0.05 + 0.2 \times 0.95 \\ &= 0.23 \\ P(t_-) &= P(t_-|C)P(C) + P(t_-|\neg C)P(\neg C) \\ &= 0.2 \times 0.05 + 0.8 \times 0.95 \\ &= 0.77 \end{aligned}$$

Now calculate the probabilities of having chickenpox given the B test result

$$\begin{aligned} P(C|t_+) &= \frac{P(t_+|C)P(C)}{P(t_+)} \\ &= \frac{0.8 \times 0.05}{0.23} \\ &= 0.1739 \\ P(C|t_-) &= \frac{P(t_-|C)P(C)}{P(t_-)} \\ &= \frac{0.2 \times 0.05}{0.77} \\ &= 0.012987 \end{aligned}$$

Calculate the best action given the result of test B

$$\begin{aligned} EU(treatment|t_+) &= P(C|t_+)U(treatment|C) + P(\neg C|t_+)U(treatment|\neg C) \\ &= 0.1739 \times (-100) + 0.8261 \times (-20) \\ &= -33.912 \end{aligned}$$

$$\begin{aligned}
EU(\neg treatment|t_+) &= P(C|t_+)U(\neg treatment|C) + P(\neg C|t_+)U(\neg treatment|\neg C) \\
&= 0.1739 \times (-200) + 0.8261 \times (0) \\
&= -34.78 \\
EU(treatment|t_-) &= P(C|t_-)U(treatment|C) + P(\neg C|t_-)U(treatment|\neg C) \\
&= 0.012987 \times (-100) + 0.987013 \times (-20) \\
&= -21.03896 \\
EU(\neg treatment|t_-) &= P(C|t_-)U(\neg treatment|C) + P(\neg C|t_-)U(\neg treatment|\neg C) \\
&= 0.012987 \times (-200) + 0.987013 \times (0) \\
&= -2.5974
\end{aligned}$$

We can now conclude the best action given a positive test B is to get treatment, whereas the best action given a negative test B is to not get treatment. We are now ready to calculate

$$\begin{aligned}
EU(B) &= EU(treatment|t_+)P(t_+) + EU(\neg treatment|t_-)P(t_-) \\
&= (-33.912)(0.23) + (-2.5974)(0.77) \\
&= -9.7998
\end{aligned}$$

Finally the cost of the new test B should be

$$\begin{aligned}
c &= EU(B) + 10 \\
&= -9.7998 + 10 \\
&= 0.2002
\end{aligned}$$

Q3 (a) There are 6 states and 4 actions per state. Remember that a policy is a sequence of actions. Thus, there are 4^6 policies.

(b) The value of each state for policy $\pi(s) = Right$ can be evaluated with policy evaluation

$$V_{k+1} = \sum_a \pi(s,a) \sum_s^t P(s,a,s')[r(s,a) + \gamma V_k(s')]$$

where $\pi(s, Right) = 1$ and $\pi(s, a \neq Right) = 0$.
$P(s,a,s') = 0.7$ where s' is the desired destination and the probability of failing and staying in the same state is 0.3. We also have a discount factor of $\gamma = 0.9$.

	S_1	S_2	S_3	S_4	S_5	S_6
V^0	148.9585	172.6027	200	-60.7806	-86.3014	-100

(c) The new policy at each state for V^0

	S_1	S_2	S_3	S_4	S_5	S_6
π^0	Right	Right	Right	Up	Up	Up

(d) The optimal value function at each state:

	S_1	S_2	S_3	S_4	S_5	S_6
V	148.9585	172.6027	200	142.25188	148.9598	158.9041

(e) The optimal value function V^* is

$$V^*(s) = \max_a E_{\pi^*}\{\sum_{k=0}^{\infty} \gamma^k r_{t+1} | s_t = s, a_t = a\}$$

The optimal value function clearly depends on the optimal policy π^*, which is not generally unique. As long as the policy is optimal (regardless of whether it is unique), the optimal value function will always return the same value. Thus, there is a unique optimal value return which implies there is a unique optimal value function.

(f) An optimal policy π^* at each state is

	S_1	S_2	S_3	S_4	S_5	S_6
π^*	Right	Right	Right	Right	Up	Up

(g) No, the optimal policy is not unique in this domain. Another optimal policy π^* is

	S_1	S_2	S_3	S_4	S_5	S_6
π^*	Right	Right	Up	Right	Up	Up

(h) A possible change is to multiply all rewards by $a > 0$, which yields the same optimal policies.

Q4 Using the update equations for TD learning:

$$\begin{aligned}
\theta_0 &\leftarrow \theta_0 + \alpha(u_j(s) - \hat{U}_\theta(s)) \\
\theta_1 &\leftarrow \theta_1 + \alpha(u_j(s) - \hat{U}_\theta(s))x \\
\theta_2 &\leftarrow \theta_2 + \alpha(u_j(s) - \hat{U}_\theta(s))y \\
\theta_3 &\leftarrow \theta_3 + \alpha(u_j(s) - \hat{U}_\theta(s))\sqrt{(x - x_g)^2 + (y - y_g)^2}
\end{aligned}$$