Assignment 5 Report

Saurabh Daptardar, Srikanth Kuthuru

November 3, 2017

1 Problem 1:

- Networks 1 and 2 can represent the information given in the question. In network 3, F1 and F2 are dependent on M1, M2, N which is not correct because F1, F2 are independent variables.
- Network 2 is the best network representation for the information provided. Number of stars 'N' is an independent vairable. Also, F1 and F2 are independent events, i.e. focal error is not dependent on anything. However, the measurements M1 and M2 depend on both N and Focal error events. This is precisely represented in network 2.
-

	N =1	N = 2	N = 3
M = 0	(e/2)(1-f) + f	f	f
M = 1	(1-f)(1-e)	(e/2)(1-f)	0
M = 2	(e/2)(1-f)	(1-f)(1-e)	(e/2)(1-f)
M = 3	0	(e/2)(1-f)	(1-f)(1-e)
M = 4	0	0	(e/2)(1-f)

- N = 2,4,6,7,8,9,...
- This is a maximum aposteriori (MAP) problem.

$$\begin{aligned}\hat{N} &= \underset{N}{\operatorname{argmax}} P(N/M_1 = 1, M_2 = 3) \\ &= \underset{N}{\operatorname{argmax}} \frac{P(M_1 = 1, M_2 = 3/N)P(N)}{P(M_1 = 1, M_2 = 3)} [Using Bayes Rule]\end{aligned}$$

We need to know the prior probalities of the number of stars (N), to calculate the above expression.

2 Problem 2:

All probability values taken as per Figure 14.12 of the AI textbook.

Notation for this problem $C = \text{Cloudy}, R = \text{Rain}, S = \text{Sprinkler}, W = \text{Wet Grass}$.
$x, \neg x$ implies variable X is set to `True` or `False`

$$\Pr(X|y) = \frac{\Pr(X,y)}{\Pr(y)} = \frac{\Pr(X,y)}{\sum\limits_X \Pr(X,y)} = \alpha \Pr(X,y)$$
$$\implies \alpha = \frac{1}{\sum\limits_X \Pr(X,y)}$$

where, α is normalizing constant factor and it'll be used ahead to simplify calculations

- To calculate $\Pr(R|s,w)$ using variable elimination

$$\Pr(C,R,s,w) = \Pr(C)\Pr(R|C)\Pr(s|C)\ \ \Pr(w|s,R)$$
$$\Pr(R|s,w) = \frac{\Pr(R,s,w)}{\Pr(s,w)}$$

Step 1: join and eliminate C

$$\begin{aligned}\Pr(R,s,w) &= \sum_C \Pr(C)\Pr(s|C)\Pr(R|C)\Pr(w|s,R)\\ &= \Pr(w|s,R) \times \{0.5 \times 0.1 \times \Pr(R|c) + 0.5 \times 0.5 \times \Pr(R|\neg c)\}\\ &= \langle 0.99, 0.9\rangle \times \{\langle 0.04, 0.01\rangle + \langle 0.05, 0.2\rangle\}\\ &= \langle 0.0891, 0.189\rangle\end{aligned}$$

$$\begin{aligned}\implies &\Pr(r,s,w) = 0.0891\\ &\Pr(\neg r,s,w) = 0.189\end{aligned}$$

Step 2: eliminate R:

$$\Pr(s,w) = \sum_R \Pr(R,s,w) = 0.0891 + 0.189 = 0.2781$$
$$\Pr(r|s,w) = \frac{0.0891}{0.2781} = 0.32$$
$$\Pr(\neg r|s,w) = \frac{0.189}{0.2781} = 0.68$$

- Markov Chain Monte Carlo sampling
  - We have 2 evidence variables (s,w) and we need to sample other 2 variables Rain and Cloudy which can take 2 values each hence we've 4 states to sample from which are $(c,r), (c,\neg r), (\neg c, r), (\neg c, \neg r)$. So our Markov chain will have **4 states**.

- To compute the Q matrix for this Markov chain, we need to find conditional probabilities of all variables that need to be sampled

$$\begin{aligned}
\Pr(C|r,s,w) &= \Pr(C|r,s) = \alpha \Pr(C)\Pr(r|C)\Pr(s|C) \\
&= \alpha\langle 0.04, 0.05\rangle \\
&= \langle 4/9, 5/9\rangle \qquad \text{(normalizing)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr(C|\neg r,s,w) &= \alpha \Pr(C)\Pr(\neg r|C)\Pr(s|C) \\
&= \langle 1/21, 20/21\rangle \\
\Pr(R|c,s,w) &= \Pr(R|c,s,w) = \frac{\Pr(c)\Pr(s|c)\Pr(R|c)\Pr(w|s,R)}{\Pr(c,s,w)} \\
&= \alpha \Pr(R|c)\Pr(w|s,R) \\
&= \alpha\langle 0.792, 0.18\rangle \\
&= \langle 22/27, 5/27\rangle
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr(R|\neg c,s,w) &= \alpha \Pr(R|\neg c)\Pr(w|s,R) \\
&= \alpha\langle 11/51, 40/51\rangle
\end{aligned}$$

To construct the transition matrix, we'll evaluate the probability of each transition. Note that in Gibbs sampling only one element changes in each sample, and hence probability of both elements changing $(c,r) \to (\neg c, \neg r)$ is 0. Also self looping as $(c,r) \to (c,r)$ can happen when r is sampled when c is same or the other way round and we'll weight it with probability of which variable is going to be sampled (in our case it's a uniform distribution).

$$\begin{aligned}
&q((c,r) \to (\neg c, \neg r)) = 0 \\
&q((c,r) \to (\neg c, r)) = 0.5\Pr(\neg c|r,s,w) = 5/18 \\
&q((c,r) \to (c,r)) = 0.5\Pr(c|r,s,w) + 0.5\Pr(r|c,s,w) = 17/27 \qquad \text{etc.}
\end{aligned}$$

And, so forth we compute the matrix to be:

$$Q = \begin{array}{c} \\ (c,r) \\ (c,\neg r) \\ (\neg c, r) \\ (\neg c,\neg r) \end{array} \begin{array}{c} \begin{array}{cccc} (c,r) & (c,\neg r) & (\neg c,r) & (\neg c,\neg r) \end{array} \\ \begin{pmatrix} 17/27 & 5/54 & 5/18 & 0 \\ 11/27 & 22/189 & 0 & 10/21 \\ 2/9 & 0 & 59/153 & 20/51 \\ 0 & 1/42 & 11/102 & 310/357 \end{pmatrix} \end{array}$$

- To find Q^n as $n \to \infty$ the below values have been rounded to 3 decimals

$$Q^n = \begin{array}{c} \\ (c,r) \\ (c,\neg r) \\ (\neg c, r) \\ (\neg c,\neg r) \end{array} \begin{array}{c} \begin{array}{cccc} (c,r) & (c,\neg r) & (\neg c,r) & (\neg c,\neg r) \end{array} \\ \begin{pmatrix} 0.142 & 0.032 & 0.178 & 0.648 \\ 0.142 & 0.032 & 0.178 & 0.648 \\ 0.142 & 0.032 & 0.178 & 0.648 \\ 0.142 & 0.032 & 0.178 & 0.648 \end{pmatrix} \end{array}$$

Hence, stationary distribution is: $p = \begin{pmatrix} 0.142 & 0.032 & 0.178 & 0.648 \end{pmatrix}$

$$\implies \Pr(r|s,w) = 0.142 + 0.178 = 0.32$$
$$\Pr(\neg r|s,w) = 0.032 + 0.648 = 0.68$$

which are the same probabilities that we calculated in part (a)

– Simulating the Markov chain is implemented in `mcmc.py` and the Q matrix is initialized with what we calculated above. The burn-in period is around 30 (where it stabilizes up to 3 decimals) but we'll take it to be 50 in the simulation.

  * For $N = 1000$:

$$\widehat{\Pr}(r|s,w) = 0.353$$
$$\widehat{\Pr}(\neg r|s,w) = 0.647$$

  * For $N = 5000$:

$$\widehat{\Pr}(r|s,w) = 0.333$$
$$\widehat{\Pr}(\neg r|s,w) = 0.667$$

  * For $N = 10000$:

$$\widehat{\Pr}(r|s,w) = 0.325$$
$$\widehat{\Pr}(\neg r|s,w) = 0.675$$

3 Problem 3:

Please find the attached code.

4 Problem 4:

Please find the attached code for problem 4. It is named 'prob4.py'.

5 Problem 5:

Given probabilities:

$$
\begin{aligned}
\Pr(s_0) &= 0.7 & & \\
\Pr(s_{t+1}|s_t) &= 0.8 & \Pr(s_{t+1}|\neg s_t) &= 0.3 \\
\Pr(r_t|s_t) &= 0.2 & \Pr(r_t|\neg s_t) &= 0.7 \\
\Pr(c_t|s_t) &= 0.1 & \Pr(c_t|\neg s_t) &= 0.3
\end{aligned}
$$

1. Formulating as a hidden Markov model Let the hidden state variable be $S_t \in \{s_t, \neg s_t\}$ which represent if student got enough sleep or not, and the 'single'observation variable be $E_t := (R_t, C_t)$ where $R_t \in \{r_t, \neg r_t\}$ is if the student has red eyes or not and $C_t \in \{c_t, \neg c_t\}$ is if the student is sleeping in class or not. So observation variable takes 4 values namely $E_t \in \{(r_t, c_t), (r_t, \neg c_t), (\neg r_t, c_t), (\neg r_t, \neg c_t)\}$

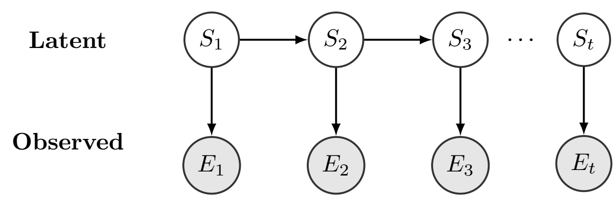

Figure 1: Hidden Markov Model diagram

The conditional transition probabilities of the model are given in the tables below

$$
\Pr(S_{t+1}|S_t) = \begin{array}{c c}
 & \begin{array}{cc} s_{t+1} & \neg s_{t+1} \end{array} \\
\begin{array}{c} s_t \\ \neg s_t \end{array} & \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}
\end{array}
$$

$$
\Pr(E_t|S_t) = \begin{array}{c c}
 & \begin{array}{cccc} (r,c) & (r,\neg c) & (\neg r, c) & (\neg r, \neg c) \end{array} \\
\begin{array}{c} s_t \\ \neg s_t \end{array} & \begin{pmatrix} 0.02 & 0.18 & 0.08 & 0.72 \\ 0.21 & 0.49 & 0.09 & 0.21 \end{pmatrix}
\end{array}
$$

2. **Filtering**: Given evidence variables: $e_1 = (\neg r, \neg c)$, $e_2 = (r, \neg c)$, $e_3 = (r, c)$, we need

to find $\Pr(S_t|e_{1:t})$ for $t = 1, 2, 3$

$$\begin{aligned}
\Pr(S_0) &= \langle 0.7, 0.3 \rangle \\
\Pr(S_1) &= \sum_{S_0} \Pr(S_1|S_0) \Pr(S_0) = \langle 0.8, 0.2 \rangle 0.7 + \langle 0.3, 0.7 \rangle 0.3 \\
&= \langle 0.65, 0.35 \rangle \\
\Pr(S_1|e_1) &= \frac{\Pr(e_1|S_1) \Pr(S_1)}{\Pr(e_1)} = \alpha \Pr(e_1|S_1) \Pr(S_1) \\
&= \alpha \langle 0.72, 0.21 \rangle \langle 0.65, 0.35 \rangle = \alpha \langle 0.468, 0.0735 \rangle \\
&= \langle \mathbf{0.8643}, \mathbf{0.1357} \rangle
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr(S_2|e_{1:2}) &= \alpha \Pr(e_2|S_2) \sum_{S_1} \Pr(S_2|S_1) \Pr(S_1|e_1) \\
&= \alpha \langle 0.18, 0.49 \rangle \ \{ \langle 0.8, 0.2 \rangle 0.8643 + \langle 0.3, 0.7 \rangle 0.1357 \} \\
&= \langle \mathbf{0.501}, \mathbf{0.499} \rangle \\
\Pr(S_3|e_{1:3}) &= \alpha \Pr(e_3|S_3) \sum_{S_2} \Pr(S_3|S_2) \Pr(S_2|e_{1:2}) \\
&= \alpha \langle 0.02, 0.21 \rangle \ \{ \langle 0.8, 0.2 \rangle 0.501 + \langle 0.3, 0.7 \rangle 0.499 \} \\
&= \langle \mathbf{0.1045}, \mathbf{0.8955} \rangle
\end{aligned}$$

3. **Smoothing**: Given evidence variables: $e_1 = (\neg r, \neg c)$, $e_2 = (r, \neg c)$, $e_3 = (r, c)$, we need to find $\Pr(S_t|e_{1:3})$ for $t = 1, 2, 3$

$$\begin{aligned}
\Pr(S_1|e_{1:3}) &= \alpha \Pr(e_{1:3}|S_1) \Pr(S_1) = \alpha \Pr(e_1|S_1) \Pr(e_{2:3}|S_1) \Pr(S_1) \\
&= \alpha \Pr(S_1|e_1) \Pr((e_{2:3}|S_1)
\end{aligned}$$

(*note: the α is not the same, it's a normalization constant factor that just absorbs other constant terms and adjusts the factors*)

As we can see from above equation, we can't directly calculate the conditional probabilities as we did above. So, we need to start backwards and then compute these probabilities.

$$\begin{aligned}
\Pr(e_3|S_3) &= \langle 0.02, 0.21 \rangle \\
\Pr(e_3|S_2) &= \sum_{S_3} \Pr(e_3|S_3) \Pr(S_3|S_2) \\
&= 0.02 \langle 0.8, 0.3 \rangle + 0.21 \langle 0.2, 0.7 \rangle \\
&= \langle 0.058, 0.153 \rangle \\
\Pr(e_{2:3}|S_1) &= \sum_{S_2} \Pr(e_3|S_2) \Pr(e_2|S_2) \Pr(S_2|S_1) \\
&= 0.058 \times 0.18 \langle 0.8, 0.3 \rangle + 0.153 \times 0.49 \langle 0.2, 0.7 \rangle \\
&= 0.01044 \langle 0.8, 0.3 \rangle + 0.07497 \langle 0.2, 0.7 \rangle \\
&= \langle 0.0233, 0.0556 \rangle
\end{aligned}$$

Now we can compute our smoothing probabilities using above values and values calculated in part 2. as follows:

$$\begin{aligned}
\Pr(S_1|e_{1:3}) &= \alpha \Pr(S_1|e_1) \Pr((e_{2:3}|S_1) \\
&= \alpha\langle 0.02, 0.0075\rangle \\
&= \langle \mathbf{0.7275}, \mathbf{0.2725}\rangle \\
\Pr(S_2|e_{1:3}) &= \alpha \Pr(S_2|e_{1:2}) \Pr((e_3|S_2) \\
&= \alpha\langle 0.029, 0.0763\rangle \\
&= \langle \mathbf{0.2757}, \mathbf{0.7243}\rangle \\
\Pr(S_3|e_{1:3}) &= \langle \mathbf{0.1045}, \mathbf{0.8955}\rangle
\end{aligned}$$

(*note:* $\Pr(S_3|e_{1:3})$ *is already calculated in part 2. of this problem*)

4. Comparing filtered probability with smoothed probability:
   In both cases we see as time progresses $(1, 2, 3)$ the probability for student not getting enough sleep is increasing. But the smoothed probabilities suggested that student isn't getting enough sleep one time step earlier than the filtered case because of more collective evidence from future data, which wasn't available while computing filtered probabilities.