CIS 520, Machine Learning, Fall 2020
Homework 7
Due: Monday, November 23rd, 11:59pm
Submit to Gradescope

Chun Chang

Problem 1

1. (a) Yes, the hidden state could be the weather.
   (b) Yes, we can see a sequence of a word as a Markov chain and the word as the hidden states.
   (c) No, it's hard to represent the watching history as a state-wise data and hard to depict the transition between the states.
   (d) Yes, the hidden state could include the status of the companies, market...etc.
2. (a) True, we have the state transition model, observation model and the prior, so we could estimate current hidden state with all previous observations which are embedded with previous hidden states.
   (b) False, adding too many hidden states to the HMM will force the model to overfit and therefore decrease the performance.
3. No, the transition matrix should not be a constant for different states, such as the the amount of sleeping time might effects the possibility of getting awake.
4. Recurrent neural network which stores the previous estimation and takes it to estimate current state.

Problem 2

1. (a)

$$Z_2 = ReLU(UX_2 + WZ_1)$$
$$O_2 = softmax(VZ_2)$$

(b)

$$Z_2 = RELU\{\begin{Bmatrix} 5.5 & 2.3 & 2 \\ 3.2 & 7.1 & 0.5 \end{Bmatrix} \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix} + \begin{Bmatrix} 0.5 & 0.2 \\ 2 & 0.9 \end{Bmatrix} \begin{Bmatrix} 0.5 \\ 0.5 \end{Bmatrix}\}$$

$$= \begin{Bmatrix} 2.35 \\ 1.95 \end{Bmatrix}$$

$$O_2 = softmax\{\begin{Bmatrix} 0.4 & 1.3 \\ 0.6 & 0.9 \\ 1.6 & 0.3 \end{Bmatrix} \begin{Bmatrix} 0.5 \\ 0.5 \end{Bmatrix}\}$$

$$= \begin{Bmatrix} 0.24 \\ 0.17 \\ 0.57 \end{Bmatrix}$$

(c) watch TV

2. (a)

$$Z_3 = ReLU(UX_3 + WZ_2)$$
$$O_3 = softmax(VZ_3)$$

(b)

$$Z_3 = RELU\{\begin{Bmatrix} 5.5 & 2.3 & 2 \\ 3.2 & 7.1 & 0.5 \end{Bmatrix} \begin{Bmatrix} 0.24 \\ 0.17 \\ 0.57 \end{Bmatrix} + \begin{Bmatrix} 0.5 & 0.2 \\ 2 & 0.9 \end{Bmatrix} \begin{Bmatrix} 2.35 \\ 1.95 \end{Bmatrix}\}$$

$$= \begin{Bmatrix} 4.46 \\ 8.78 \end{Bmatrix}$$

$$O_3 = softmax\{\begin{Bmatrix} 0.4 & 1.3 \\ 0.6 & 0.9 \\ 1.6 & 0.3 \end{Bmatrix} \begin{Bmatrix} 4.46 \\ 8.78 \end{Bmatrix}\}$$

$$= \begin{Bmatrix} 0.90 \\ 0.06 \\ 0.02 \end{Bmatrix}$$

(c) sing

Problem 3

1.

$$p(x_1)p(x_2)p(x_3 \mid x_1)p(x_4 \mid x_1, x_2)p(x_5 \mid x_3, x_4)p(x_6 \mid x_4)$$

2. No,

$$\frac{p(X_3)p(X_4)p(X_5 \mid X_3)p(X_6 \mid X_3)}{p(X_3 \mid X_1)p(X_4 \mid X_1, X_2)p(X_6 \mid X_4)} \neq 1$$

3.

$$p(x_1)p(x_2)p(x_3 \mid x_1)p(x_4 \mid x_1, x_2)p(x_5 \mid x_4)p(x_6 \mid x_4)$$

$$p(x_5 \mid x_3, x_4) = \frac{p(X_5, X_3 \mid X_4)}{p(X_3 \mid X_4)} = \frac{(p(X_5 \mid X_3)p(X_3)) \mid X_4)}{p(X_3 \mid X_4)}$$

If X_3 is independent with X_5, the new BN is larger.

4. (a) True, $p(X_1)$ and $p(X_2)$ are independent because there is no information from node X_1 to X_2
   (b) True, $X_i \perp Nondescendent(X_i) \mid Parent(X_i) => (X_6 \perp Nondescendent(X_6) = X_3 \mid Parent(X_6) = X_4$, $X_6 \mid X_4$ is conditionally independent to $X_3 \mid X_4$
   (c) False, given known X_6, which explained away the X_1 and X_2, so they are not conditionally independent. $X_1 \mid X_6$ and $X_2 \mid X_6$ is marginally independent given not knowing X_6.
   (d) True, given the known X_4, probability of X_5 is only dependent on X_4 and conditional independent with X_2

Problem 4

1. 4.1 Nothing to Report. Please skip.

2. Do you add the link(yes/no)?
   Steps:
   $P(B|A) = 0.7272$
   $P(B|\ A) = 0.6666$
   $P(B) = 0.7058$
   $P(B|A) \neq P(B|\ A)$ add a link from A to B

3. Do you add the link(yes/no)?
   Steps:
   $P(C|A, B) = 0.5$
   $P(C|A,\ B) = 0.6666$
   there exist 1 or more links $P(C|A) = 0.5454$
   $P(C|\ A) = 0.6666$
   Add link from A to C

4. Do you add the link(yes/no)?
   Steps:
   $P(C|B) = 0.5833$
   $P(C|\ B) = 0.6000$
   Add link from B to C

5. Do you add the link(yes/no)?
   Steps:
   $P(D|A, B, C) = 0.75$
   $P(D|\ A, B, C) = 0.3333$
   there exist 1 or more links $P(D|A) = 0.4545$
   $P(D|\ A) = 0.3333$
   Add link from A to D

6. Do you add the link(yes/no)?
   Steps:
   $P(D|B) = 0.3333$
   $P(D|\ B) = 0.6000$
   Add link from B to D

7. Do you add the link(yes/no)?
   Steps:
   $P(D|C) = 0.7777$
   $P(D|\ C) = 0$
   Add link from C to D

8. the graph: