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

Yuezhong Chen, Sheil Sarda

Problem 1

1. (a) It is appropriate as causes of precipitation include many other factors besides previous day precipitation.
   (b) Not appropriate as there is nothing hidden.
   (c) Appropriate as factors other than preferences may affect the pleasure from a specific movie.
   (d) Appropriate, as many factors affect market prices besides latest prices.
2. (a) False, as Markov models depend only on the last state.
   (b) True, as it give more parameters and more freedom to fit the training data.
3. This probability may certainly depend on the time of the day which is not a random variable, so a vanilla HMM with a constant transition matrix would not work well.
4. One can take a grid of times in a 24 hours period and us an independent transition matrix for each such time.

Problem 2

1. (a) $Z_2 = ReLU(UX_2 + WZ_1); O_2 = softmax(VZ_2)$
   (b) Because $Z_2 = ReLU(UX_2 + WZ_1)$, we firstly calculate the following:

$$UX_2 + WZ_1 = \begin{pmatrix} 5.5 & 2.3 & 2.0 \\ 3.2 & 7.1 & 0.5 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.2 \\ 2.0 & 0.9 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 2.0 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.35 \\ 1.45 \end{pmatrix}$$

   Thus, $Z_2 = \begin{pmatrix} 2.35 \\ 1.95 \end{pmatrix}$. And because $O_2 = softmax(VZ_2)$

$$VZ_2 = \begin{pmatrix} 0.4 & 1.3 \\ 0.6 & 0.9 \\ 1.6 & 0.3 \end{pmatrix} \begin{pmatrix} 2.35 \\ 1.95 \end{pmatrix} = \begin{pmatrix} 3.475 \\ 3.165 \\ 4.345 \end{pmatrix}$$

   Thus, $O_2 = softmax \begin{pmatrix} 3.475 \\ 3.165 \\ 4.345 \end{pmatrix} \approx \begin{pmatrix} 0.243 \\ 0.178 \\ 0.579 \end{pmatrix}$.
   (c) It is watching TV with probability 0.5793.
2. (a) $Z_3 = ReLU(UX_3 + WZ_2) = ReLU(UO_2 + WZ_2); O_3 = softmax(VZ_3)$

(b) Following the same strategies above, we have:

$$Z_3 \approx \begin{pmatrix} 3.25 \\ 1.95 \end{pmatrix}, O_3 \approx \begin{pmatrix} 0.9047 \\ 0.0658 \\ 0.0295 \end{pmatrix}$$

(c) It is singing with probability 0.9047.

Problem 3

1. $p\left(X_1, X_2, X_3, X_4, X_5, X_6\right) = p(X_1)p(X_2)p(X_3|X_1)p(X_4|X_1, X_2)p(X_5|X_3, X_4)p(X_6|X_4)$
2. $p\left(X_1, X_2, X_3, X_4, X_5, X_6\right) = p\left(X_1\right) p\left(X_2\right) p\left(X_3\right) p\left(X_4\right) p\left(X_5 \mid X_3\right) p\left(X_6 \mid X_3\right)$

   No, it is not included since this distribution has a restriction on $X_6|X_3$ whereas the original network has a restriction on $X_6|X_4$.
3. If the edge from X_3 to X_5 is removed, will the class of joint probability distributions that can be represented by the resulting Bayesian network be smaller or larger than that associated with the original network?
   The class of joint probability distributions would be larger if the edge were removed, because we are removing the restriction on $X_5|X_3$.
4. (a) $p\left(X_1, X_2\right) = p\left(X_1\right) p\left(X_2\right)$. True, are not connected by a path.
   (b) $p\left(X_3, X_6 \mid X_4\right) = p\left(X_3 \mid X_4\right) p\left(X_6 \mid X_4\right)$. True, X_3 and X_6 are not connected.
   (c) $p\left(X_1, X_2 \mid X_6\right) = p\left(X_1 \mid X_6\right) p\left(X_2 \mid X_6\right)$. True, X_1 and X_2 are not connected.
   (d) $p\left(X_2, X_5 \mid X_4\right) = p\left(X_2 \mid X_4\right) p\left(X_5 \mid X_4\right)$. False, X_2 and X_5 are connected through X_4.

Problem 4

1. 4.1 Nothing to Report. Please skip.
2. Do you add the link(yes/no)?
   Steps:

   $\mathrm{P}(B) = 0.706$
   $\mathrm{P}(B|A) = 0.727$
   $\mathrm{P}(B|A^C) = 0.667$

   Since the difference between conditional probabilities and the probability of the event is less than 0.05, A and B are independent. Do not add the link.
3. Do you add the link(yes/no)?
   Steps:

   $\mathrm{P}(C) = 0.588$
   $\mathrm{P}(C|A) = 0.545$
   $\mathrm{P}(C|A^C) = 0.667$

   Since the difference between conditional probabilities and the probability of the event is greater than 0.05, add the link.
4. Do you add the link(yes/no)?
   Steps:

   $\mathrm{P}(C|B) = 0.583$
   $\mathrm{P}(C|B^C) = 0.600$

   Since the difference between conditional probabilities and the probability of the event is less than 0.05, do not add the link.

5. Do you add the link(yes/no)?
   Steps:

   $P(D) = 0.412$
   $P(D|A) = 0.454$
   $P(D|A^C) = 0.333$

   Since the difference between conditional probabilities and the probability of the event is greater than 0.05, add the link.

6. Do you add the link(yes/no)?
   Steps:

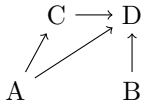
   $P(D|B) = 0.454$
   $P(D|B^C) = 0.600$

   Since the difference between conditional probabilities and the probability of the event is greater than 0.05, add the link.

7. Do you add the link(yes/no)?
   Steps:

   $P(D|C) = 0.700$

   Since the difference between conditional probabilities and the probability of the event is greater than 0.05, add the link.



8.