Homework 3
Inference and Representation

Aodong Li
NetID: al5350
al5350@nyu.edu

June 20, 2018

1. (a) Add $(A, B), (D, E), (E, F)$ and remove all the arrows.

   (b) A,B,G,H,D,F,E,C

   (c) E,C,A,B,D,F,G,H (C,E comes first.)

   (d) Because the CDP involving the H is a valid probability distribution, summation is one is preserved. That is,

   $$\sum_h p(h|E, F) = 1.$$

   After removing H, we can safely remove F, because

   $$\sum_f p(f|C) = 1.$$

   Formally, we compute the marginal inference over B given G $p(B|G = 1)$. This marginalizes all the other variables. We use the trick to push the sum into the products,

   $$\begin{aligned} p(B, G) &= \sum_{A,C,D,E} p(A, B, C, D, E, G = 1) \sum_F p(F|C) \sum_H p(H|E, F) \\ &= \sum_{A,C,D,E} p(A, B, C, D, E, G). \end{aligned}$$

(e)

$$\begin{aligned}
p(B, G=1) &= \sum_{A,C,D,E} p(A, B, C, D, E, G=1) \\
&= \sum_{D,E,C} p(D|C)p(E|C)p(G=1|D,E) \\
&\quad \sum_{A} p(A)p(B)p(C|A,B) \\
&= \sum_{E,C} p(B,C) \sum_{D} p(D|C)p(G=1|D,E)p(E|C) \\
&= \sum_{E,C} p(B,C)p(E, G=1|C) \\
&= \sum_{C} p(B,C) \sum_{E} p(E, G=1|C) \\
&= \sum_{C} p(B,C)p(G=1|C) \\
&= p(B, G=1).
\end{aligned}$$

Once we have got the $p(B, G=1)$ we marginalize over B so to get the evidence $p(G=1) = \sum_B p(B, G=1)$. The final query $p(B=0|G=1) = \frac{p(B=0,G=1)}{p(G=1)}$ by Bayes theorem.

2. All the later states and observations can be safely marginalized and do not have influence on the earlier variables.

   (a) $p(X_2 = \text{Happy}) = p(X_2 = \text{Happy}|X_1 = \text{Happy}) = 0.9$.

   (b) Starting from this question, we hide the condition of $X_1 = \text{Happy}$ to get rid of abuse of notations.

$$\begin{aligned}
p(Y_2 = \text{Frown}) &= \sum_{X_2} p(Y_2 = \text{Frown}, X_2) \\
&= p(Y_2 = \text{Frown}|X_2 = \text{Happy})p(X_2 = \text{Happy}) + \\
&\qquad p(Y_2 = \text{Frown}|X_2 = \text{Angry})p(X_2 = \text{Angry}) \\
&= 0.1 * 0.9 + 0.6 * 0.1 \\
&= 0.15
\end{aligned}$$

(c)

$$\begin{aligned} p(X_2 = \text{Happy}|Y_2 = \text{Frown}) &= \frac{p(Y_2 = \text{Frown}|X_2 = \text{Happy})p(X_2 = \text{Happy})}{p(Y_2 = \text{Frown})} \\ &= \frac{0.09}{0.15} \\ &= 0.6. \end{aligned}$$

(d) After a long run, the Markov chain reaches a stable state where $p(X = \text{Happy}) = p(X = \text{Angry}) = 0.5$. We can assume that after 80 steps the network reaches at a stable state.

$$\begin{aligned} p(Y_{80} = \text{yell}) &= p(Y_{80} = \text{yell}|X_{80} = \text{Happy})p(X_{80} = \text{Happy}) + \\ &\qquad p(Y_{80} = \text{yell}|X_{80} = \text{Angry})p(X_{80} = \text{Angry}) \\ &= 0.5 * 0.1 + 0.5 * 0.2 \\ &= 0.15. \end{aligned}$$

(e) By employing the Markov property and probability product rule,

$$\begin{aligned} p(X_1, ..., X_n, y_1, ..., y_n) &= p(X_1, ..., X_{n-1}, y_1, ..., y_{n-1})p(X_n, y_n|X_{n-1}) \\ &= p(X_1, ..., X_{n-1}, y_1, ..., y_{n-1})p(X_n|X_{n-1})p(y_n|X_n) \end{aligned}$$

$$\begin{aligned} \arg\max_X p(X_1, ..., X_n, y_1, ..., y_n) = \arg\max_X \log p(X_1)p(y_1|X_1) + \\ \sum_{i=2}^{n} \log p(X_i|X_{i-1})p(y_i|X_i) \end{aligned}$$

where we use the notation that the lower-case letter refers to a random variable with a specific value and the capital letters are random variables.

We have seen an explicit pattern where the $\arg\max$ over X_5 depends on the what the value of variable X_4. This is a dynamic programming framework.

The first state is solid, which is X_1 =Happy. For adjacent states,

$$\begin{aligned} p(X_i = H|X_{i-1} = H)p(y_i|X_i = H) &= 0.09 \\ p(X_i = A|X_{i-1} = H)p(y_i|X_i = A) &= 0.06 \\ p(X_i = H|X_{i-1} = A)p(y_i|X_i = H) &= 0.01 \\ p(X_i = A|X_{i-1} = A)p(y_i|X_i = A) &= 0.53 \end{aligned}$$

It is obvious that taking the sequence (H, A, A, A, A) gives the maximum.

3. (a)

   (b)

   (c)