Probabilistic Graphical Models, HW1

Haekyu Park

Feb 6, 2020

1 Fundamentals

Problem 1.1

1. Season $\perp$ Chills
   $\boxed{\text{False}}$: There is a path Season $\rightarrow$ Flu $\rightarrow$ Chills.

2. Season $\perp$ Chills | Flu
   $\boxed{\text{True}}$: There is no path from Season to Chills, since Flu is observed and it blocks all paths.

3. Season $\perp$ Headache | Flu
   $\boxed{\text{False}}$: There is a path Season $\rightarrow$ Dehydration $\rightarrow$ Chills.

4. Season $\perp$ Headache | Flu, Dehydration
   $\boxed{\text{True}}$: There is no path from Season to Headache, since Flu and Dehydration are observed and they block all paths.

5. Season $\perp$ Nausea | Dehydration
   $\boxed{\text{False}}$: There is a path Season $\rightarrow$ Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea.

6. Season $\perp$ Nausea | Dehydration, Headache
   $\boxed{\text{True}}$: There is no path from Season to Nausea, since Dehydration and Headache are observed and they block all paths.

7. Flu $\perp$ Dehydration
   $\boxed{\text{False}}$: There is a path Flu $\leftarrow$ Season $\rightarrow$ Dehydration.

8. Flu $\perp$ Dehydration | Season, Headache
   $\boxed{\text{False}}$: There is a path Flu $\rightarrow$ Headache $\leftarrow$ Dehydration.

9. Flu $\perp$ Dehydration | Season
   $\boxed{\text{True}}$: There is no path from Flu and Dehydration.
   - Flu $\leftarrow$ Season $\rightarrow$ Dehydration is blocked since Season is observed.
   - Flu $\rightarrow$ Headache $\leftarrow$ Dehydration is blocked since Headache is not observed.

10. Flu $\perp$ Dehydration | Season, Nausea
    $\boxed{\text{False}}$: There is a path Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea $\leftarrow$ Dehydration. Dizziness $\rightarrow$ Nausea $\leftarrow$ Dehydration is connected, since it is a v-structure and Nausea is observed.

11. Chills $\perp$ Nausea
    $\boxed{\text{False}}$: There is a path Chills $\leftarrow$ Flu $\leftarrow$ Season $\rightarrow$ Dehydration $\rightarrow$ Nausea.

12. Chills $\perp$ Nausea | Headache
    $\boxed{\text{False}}$: There is a path Chills $\leftarrow$ Flu $\leftarrow$ Season $\rightarrow$ Dehydration $\rightarrow$ Nausea.

Problem 1.2

1.

$$\boxed{P(S,F,D,C,H,N,Z) = P(S)P(F|S)P(D|S)P(C|F)P(H|F,D)P(N|D,Z)P(Z|H)}$$

2.

$$\boxed{\begin{aligned}&\frac{1}{Z}\cdot\phi_1(S)\phi_2(F)\phi_3(D)\phi_4(C)\phi_5(H)\phi_6(N)\phi_7(Z)\\ &\quad\cdot\phi_8(S,F)\phi_9(S,D)\phi_{10}(F,C)\phi_{11}(F,H)\phi_{12}(D,H)\phi_{13}(D,N)\phi_{14}(H,Z)\phi_{15}(N,Z)\end{aligned}}$$

Problem 1.3

1.

$$\begin{aligned}&P(F = true)\\ &= \sum_s P(F = true, S = s)\\ &= \sum_s P(F = true|S = s)P(S = s)\\ &= P(F = true|S = winter)P(S = winter) + P(F = true|S = summer)P(S = summer)\\ &= (0.4\cdot 0.5) + (0.1\cdot 0.5)\\ &= \boxed{0.25}\end{aligned}$$

2.

$$P(F = true|S = winter) = \boxed{0.4}$$

3.

$$\begin{aligned}&P(F = true|S = winter, H = true)\\ &= \frac{P(F = true, S = winter, H = true)}{P(S = winter, H = true)}\\ &= \frac{\sum_d P(F = true, S = winter, H = true, D = d)}{\sum_{f,d} P(F = f, S = winter, H = true, D = d)}\\ &= \frac{\sum_d P(S = winter)P(F = true|S = winter)P(D = d|S = winter)P(H = true|F = true, D = d)}{\sum_{f,d} P(S = winter)P(F = f|S = winter)P(D = d|S = winter)P(H = true|F = f, D = d)}\\ &= \frac{A}{B}\end{aligned}$$

$$
\begin{aligned}
A &= \sum_d P(S = winter)P(F = true|S = winter)P(D = d|S = winter)P(H = true|F = true, D = d) \\
&= P(S = winter)P(F = true|S = winter) \sum_d P(D = d|S = winter)P(H = true|F = true, D = d) \\
&= 0.5 \cdot 0.4 \sum_d P(D = d|S = winter)P(H = true|F = true, D = d) \\
&= 0.2 \sum_d P(D = d|S = winter)P(H = true|F = true, D = d) \\
&= 0.2 \Big(P(D = true|S = winter)P(H = true|F = true, D = true) \\
&\qquad + P(D = false|S = winter)P(H = true|F = true, D = false) \Big) \\
&= 0.2 \cdot (0.1 \cdot 0.9 + 0.9 \cdot 0.8) \\
&= 0.162
\end{aligned}
$$

$$
\begin{aligned}
B &= \sum_{f,d} P(S = winter)P(F = f|S = winter)P(D = d|S = winter)P(H = true|F = f, D = d) \\
&= P(S = winter) \sum_{f,d} P(F = f|S = winter)P(D = d|S = winter)P(H = true|F = f, D = d) \\
&= 0.5 \sum_{f,d} P(F = f|S = winter)P(D = d|S = winter)P(H = true|F = f, D = d) \\
&= 0.5 \Big(P(F = true|S = winter)P(D = true|S = winter)P(H = true|F = true, D = true) \\
&\qquad + P(F = true|S = winter)P(D = false|S = winter)P(H = true|F = true, D = false) \\
&\qquad + P(F = false|S = winter)P(D = true|S = winter)P(H = true|F = false, D = true) \\
&\qquad + P(F = false|S = winter)P(D = false|S = winter)P(H = true|F = false, D = false) \Big) \\
&= 0.5 \Big(0.4 \cdot 0.1 \cdot 0.9 \\
&\qquad + 0.4 \cdot 0.9 \cdot 0.8 \\
&\qquad + 0.6 \cdot 0.1 \cdot 0.8 \\
&\qquad + 0.6 \cdot 0.9 \cdot 0.3 \Big) \\
&= 0.267
\end{aligned}
$$

$$
P(F = true|S = winter, H = true) = \frac{A}{B} = \frac{0.162}{0.267} = \boxed{0.6067}
$$

4.

$$
\begin{aligned}
&P(F = true|S = winter, H = true, D = true)\\
&= \frac{P(F = true, S = winter, H = true, D = true)}{P(S = winter, H = true, D = true)}\\
&= \frac{P(S = winter)P(F = true|S = winter)P(D = true|S = winter)P(H = true|F = true, D = true)}{\sum_f P(S = winter, H = true, D = true, F = f)}\\
&= \frac{0.5 \cdot 0.4 \cdot 0.1 \cdot 0.9}{\sum_f P(S = winter)P(F = f|S = winter)P(D = true|S = winter)P(H = true|F = f, D = true)}\\
&= \frac{0.5 \cdot 0.4 \cdot 0.1 \cdot 0.9}{P(S = winter)P(D = true|S = winter)\sum_f P(F = f|S = winter)P(H = true|F = f, D = true)}\\
&= \frac{0.5 \cdot 0.4 \cdot 0.1 \cdot 0.9}{0.5 \cdot 0.1 \sum_f P(F = f|S = winter)P(H = true|F = f, D = true)}\\
&= \frac{0.4 \cdot 0.9}{P(F = true|S = w)P(H = t|F = true, D = t) + P(F = false|S = w)P(H = t|F = false, D = t)}\\
&= \frac{0.4 \cdot 0.9}{0.4 \cdot 0.9 + 0.6 \cdot 0.8}\\
&= \boxed{0.4286}
\end{aligned}
$$

5.

$$
\begin{aligned}
&P(F = true|D = true)\\
&= \frac{P(F = true, D = true)}{P(D = true)}\\
&= \frac{\sum_s P(F = true, D = true, S = s)}{\sum_s P(D = true, s = s)}\\
&= \frac{\sum_s P(F = true, D = true, S = s)}{\sum_s P(D = true, s = s)}\\
&= \frac{\sum_s P(S = s)P(F = true|S = s)P(D = true|S = s)}{\sum_s P(S = s)P(D = true|s = s)}\\
&= \frac{0.5 \cdot 0.4 \cdot 0.1 + 0.5 \cdot 0.1 \cdot 0.3}{0.5 \cdot 0.1 + 0.5 \cdot 0.3}\\
&= \frac{0.4 \cdot 0.1 + 0.1 \cdot 0.3}{0.1 + 0.3}\\
&= 0.175
\end{aligned}
$$

$P(F = true) = 0.25$ by Problem 1.3-1. $P(F = true|D = true) \leq P(F = true)$; thus, knowing you are dehydrated decreases the likelihood of having the flu.

It makes sense intuitively. Generally, the symptoms of flu and dehydration are similar. If people who have the symptoms know that they are having dehydration, they can exclude flu as the reason of the symptoms.

Problem 1.4

1. There aren't any differences between the set of marginal independencies encoded by the directed and undirected versions of this model. This is because neither model encodes any marginal independencies at all.

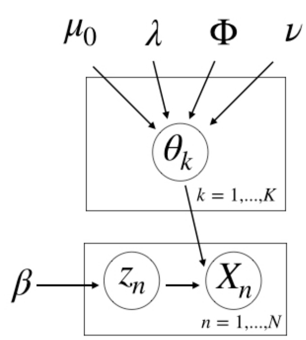
2. There are some differences between the set of conditional independencies encoded by the directed and undirected versions of this model. For example, in the directed version of the model, $\neg(F \perp D)|S, H$, since there is a path $F \rightarrow H \leftarrow D$. In the undirected version of the model, $F \perp D|S, H$, since F and D are blocked by S and H.

2 Bayesian Networks

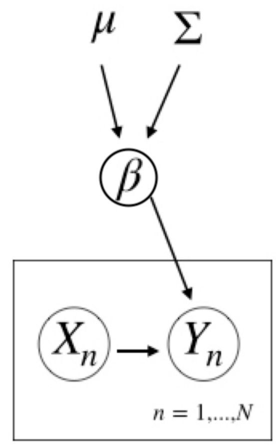
Problem 2.1

1. (Gaussian Mixture Model)
   BN of the given Gaussian Mixture Model can be build as in $\boxed{\text{Figure 1a}}$. The relation between θ and X which is represented by $\theta_k \to X_n$ includes μ and Σ between them; $\theta_k \to (\mu, \Sigma) \to X_n$.

2. (Bayesian Logistic Regression)
   BN of the given Bayesian Logistic Regression can be build as in $\boxed{\text{Figure 1b}}$.



(a) The plate diagram of the given Gaussian Mixture Model



(b) The plate diagram of the given Gaussian Mixture Model

Problem 2.2 Without loss of generality, we can assume that $X_1, ..., X_n$ are sorted in a topological ordering of the graph $\mathcal{G}$. Let's use the following notations.

- $Pa(X_i)$: the parents of X_i
- $Nd(X_i)$: the non-descendants of X_i
- $D(X_i)$: the descendants of X_i

Given that P factorizes according to $\mathcal{G}$, we know the following equation.

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i)) \tag{1}$$

To show that $\mathcal{G}$ is an I-map for P, we need to show that the following equation.

$$P(X_i | Nd(X_i)) = P(X_i | Pa(X_i)) \tag{2}$$

Let's show (2).

$$
\begin{aligned}
P(X_i|Nd(X_i)) &= P(X_i|Pa(X_i)) \\
&= \frac{P(X_i, Nd(X_i))}{P(Nd(X_i)} \\
&= \frac{P(X_i, Nd(X_i))}{\sum_{X_i} P(Nd(X_i), X_i)} \\
&= \frac{\prod_{X_j \in \{X_i\} \cup Nd(X_i)} P(X_j|Pa(X_j))}{\sum_{X_i} \left(\prod_{X_j \in \{X_i\} \cup Nd(X_i)} P(X_j|Pa(X_j)) \right)} \\
&= \frac{P(X_i|Pa(X_i)) \cdot \prod_{X_j \in Nd(X_i)} P(X_j|Pa(X_j))}{\sum_{X_i} \left(P(X_i|Pa(X_i)) \cdot \prod_{X_j \in Nd(X_i)} P(X_j|Pa(X_j)) \right)} \\
&= \frac{P(X_i|Pa(X_i)) \cdot \cancel{\left(\prod_{X_j \in Nd(X_i)} P(X_j|Pa(X_j)) \right)}}{\cancel{\left(\prod_{X_j \in Nd(X_i)} P(X_j|Pa(X_j)) \right)} \cdot \sum_{X_i} P(X_i|Pa(X_i))} \\
&= \frac{P(X_i|Pa(X_i))}{1} = P(X_i|Pa(X_i))
\end{aligned} \tag{3}
$$

Equation (2) holds; thus, $\mathcal{G}$ is an I-map for P.

Problem 2.3

1. Let's show that d-sep$_{\mathcal{G}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \Rightarrow$ sep$_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$.

   If d-sep$_{\mathcal{G}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$, every path from any $X \in \mathbf{X}$ to any $Y \in \mathbf{Y}$ should include a node W that makes the path inactive. W should be one of the following cases:

   - Case1: $W \in \mathbf{Z}$.
     In case1, W should make edges with its adjacent nodes in the path as in one of the following ways.
     - ▷ ← W →
     - ▷ → W →
     - ▷ ← W ←

     In $\mathcal{H}$, the nodes adjacent to W do not make additional connection between them, because they are not common parents to W. If W is observed in $\mathcal{H}$, the adjacent nodes are independent to each other, thus the path is inactive.
   - Case2: $W \notin \mathbf{Z}$.
     In case2, W should make a v-structure (i.e., → W ←). Because the v-structure is inactive, its base (and any of its descendants) must not be in $\mathcal{H}$; thus; the path does not exist in $\mathcal{H}$.

   In both cases, sep$_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$.

2. Let's show that $\text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \Rightarrow \text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$.

   Let's assume that $\text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ is false. With this assumption, there is an active path from $X \in \mathbf{X}$ to $Y \in \mathbf{Y}$ given $\mathbf{Z}$. Let's represent the active path as $X, W_1, ..., W_n, Y$. Every three sequential nodes A, B, C in the active path should satisfy the following cases.

   - Case1: the three nodes construct a v-structure (i.e., $A \rightarrow B \leftarrow C$).
     B should be in $\mathbf{Z}$ to make the sequence A, B, C be a part of an active path. In $\mathcal{H}$, (A, C) is newly connected, and (A, B) and (B, C) stay connected. This means that A, B, C are not independent to each other in $\mathcal{H}$. Thus, there is an active path $A - B - C$ in $\mathcal{H}$.
   - Case2: the three nodes do not construct a v-structure.
     In this case, the connections A, B and B, C stay connected in $\mathcal{H}$. Thus there is an active path $A - B - C$ in $\mathcal{H}$.

   By the case 1 and 2, there is an active path $X, W_1, ..., W_n, Y$ in $\mathcal{H}$; thus, $\text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ is false. By the proof of conflict, $\text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \Rightarrow \text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$.

We showed $\text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \Rightarrow \text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ in 1, and $\text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z}) \Leftarrow \text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ in 2. Thus, $\text{d-sep}_{\mathcal{G}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$ if and only if $\text{sep}_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$.

Problem 2.4 To ensure that all dependencies remain from the original network, we need to check all active paths assuming that Alarm is always unobserved.

- Alarm is unobserved, so there are active paths between Alarm's direct ancestors and children. Thus, direct edges that should be added are $B \rightarrow J$, $B \rightarrow M$, $E \rightarrow J$, and $E \rightarrow J$.
- Also, any descendants of Alarm are connected by active paths. Thus, there should be a direct edge between J and M. Let's add $J \leftarrow M$.
- In the original graph, there are direct edges $T \rightarrow J$ and $N \rightarrow M$. They should be added too.
- In the original graph, if M is observed and A is unobserved, there is an active path J, A, M, N. We need to preserve the independency that $\neg(J \perp N)|M$. For this independency, we need $N \rightarrow J$ in the new graph.

Figure 2 shows the result minimal I-map.

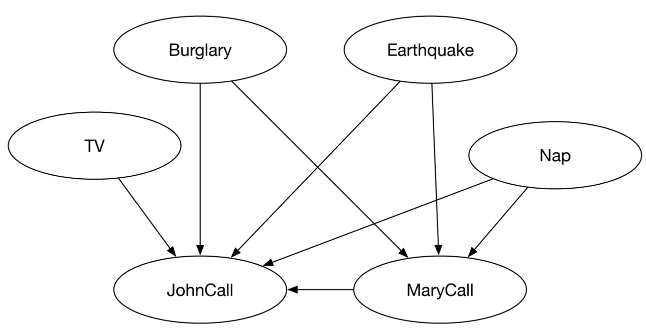


Figure 2: A minimal I-map except for Alarm.

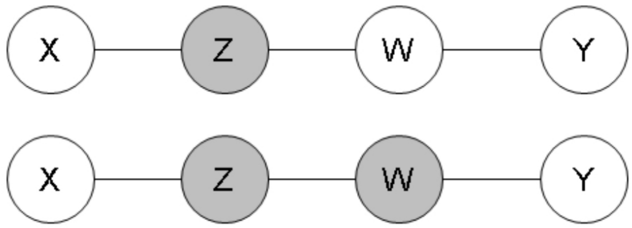


Figure 3: A Markov network where $X \perp Y | Z$.

Problem 2.5

1. **Strong Union**
   If $\mathbf{X}$ and $\mathbf{Y}$ are blocked by $\mathbf{Z}$ as seen in Figure 3, adding $\mathbf{W}$ does not unblock the path. Therefore, $\mathcal{I}$ satisfies strong union.

2. **Transitivity**
   $\neg(\mathbf{X} \perp A | \mathbf{Z})$ means that there is a path from $X \in \mathbf{X}$ to A that does not go through $\mathbf{Z}$. $\neg(A \perp \mathbf{Y} | \mathbf{Z})$ means that there is a path from A to $X \in \mathbf{X}$ that does not go through $\mathbf{Z}$. Combining these, we know that there is a path from $X \to \cdots \to A \to \cdots \to Y$ to that does not go through $\mathbf{Z}$, meaning $\mathbf{X}$ and $\mathbf{Y}$ are not independent given $\mathbf{Z}$. Thus, the transitivity holds; $\neg(\mathbf{X} \perp A | \mathbf{Z})$ and $\neg(A \perp \mathbf{Y} | \mathbf{Z}) \Rightarrow \neg(\mathbf{X} \perp \mathbf{Y} | \mathbf{Z})$.

3 Ising graphical model

Problem 3

1. • First, let's compute $\frac{P(X_i=1,X_{-i};\theta)}{P(X_i=0,X_{-i};\theta)}$.

$$\begin{aligned}
\frac{P(X_i=1,X_{-i};\theta)}{P(X_i=0,X_{-i};\theta)} &= \frac{P(X_i=1,\mathbf{X}_{-i};\theta)}{\cancel{P(\mathbf{X}_{-i};\theta)}} \cdot \frac{\cancel{P(\mathbf{X}_{-i};\theta)}}{P(X_i=0,\mathbf{X}_{-i};\theta)} \\
&= \frac{P(X_i=1,\mathbf{X}_{-i};\theta)}{P(X_i=0,\mathbf{X}_{-i};\theta)} \\
&= \frac{exp\left(\theta_i + \cancel{\sum_{s\in V, s\neq i} \theta_s X_s} + \sum_{(s,i)\in E} \theta_{s,i} X_s + \cancel{\sum_{(s,t)\in E, t\neq i} \theta_{s,t} X_s X_t}\right)}{exp\left(0 + \cancel{\sum_{s\in V, s\neq i} \theta_s X_s} + 0 + \cancel{\sum_{(s,t)\in E, t\neq i} \theta_{s,t} X_s X_t}\right)} \\
&= exp\left(\theta_i + \sum_{(s,i)\in E} \theta_{s,i} X_s\right)
\end{aligned}$$

• Let's compute $P(X_i=1|\mathbf{X}_{-i};\theta)$.

$$\begin{aligned}
P(X_i=1|\mathbf{X}_{-i};\theta) &= \frac{P(X_i=1|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)+P(X_i=1|-\mathbf{i};\theta)} \\
&= \frac{\frac{P(X_i=1|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)}}{\frac{P(X_i=0|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)} + \frac{P(X_i=1|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)}} \\
&= \frac{\frac{P(X_i=1|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)}}{1 + \frac{P(X_i=1|-\mathbf{i};\theta)}{P(X_i=0|-\mathbf{i};\theta)}} \\
&= \frac{exp\left(\theta_i + \sum_{(s,i)\in E} \theta_{s,i} X_s\right)}{1 + exp\left(\theta_i + \sum_{(s,i)\in E} \theta_{s,i} X_s\right)}
\end{aligned}$$

So we get the logistic regression model, where

$$exp\left(\theta_i + \sum_{(s,i)\in E} \theta_{s,i} X_s\right)$$

is the log-odds.

2. •

$$logP(\mathbf{X}|\mathbf{W};\theta,\beta) = \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{st} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u - logZ(\mathbf{W},\theta,\beta)$$

- $$\begin{aligned}
\frac{\partial logP(\mathbf{X}|\mathbf{W};\theta,\beta)}{\partial\theta_s} &= \frac{\partial}{\partial\theta_s}\Bigg(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u - logZ(\mathbf{W},\theta,\beta)\Bigg)\\
&= X_s - \frac{\partial}{\partial\theta_s}logZ(\mathbf{W},\theta,\beta)\\
&= X_s - \frac{\frac{\partial}{\partial\theta_s}Z(\mathbf{W},\theta,\beta)}{Z(\mathbf{W},\theta,\beta)}\\
&= X_s - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\frac{\partial}{\partial\theta_s}\sum_{\mathbf{X}}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_s - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}}\frac{\partial}{\partial\theta_s}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_s - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}}X_sexp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_s - \sum_{\mathbf{X}}X_s\frac{1}{Z(\mathbf{W},\theta,\beta)}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_s - \sum_{\mathbf{X}}X_sP(\mathbf{X}|\mathbf{W};\theta,\beta)\\
&= \boxed{X_s - \mathbb{E}_{P(\mathbf{X}|\mathbf{W};\theta,\beta)}[X_s]}
\end{aligned}$$

- $$\begin{aligned}
\frac{\partial logP(\mathbf{X}|\mathbf{W};\theta,\beta)}{\partial\theta_{st}} &= \frac{\partial}{\partial\theta_{st}}\Bigg(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u - logZ(\mathbf{W},\theta,\beta)\Bigg)\\
&= X_sX_t - \frac{\partial}{\partial\theta_{st}}logZ(\mathbf{W},\theta,\beta)\\
&= X_sX_t - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\frac{\partial}{\partial\theta_{st}}\sum_{\mathbf{X}}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_sX_t - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}}\frac{\partial}{\partial\theta_{st}}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_sX_t - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}}X_sX_texp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_sX_t - \sum_{\mathbf{X}}X_sX_t\frac{1}{Z(\mathbf{W},\theta,\beta)}exp\Bigg(\sum_{s\in V}\theta_sX_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Bigg)\\
&= X_sX_t - \sum_{\mathbf{X}}X_sX_tP(\mathbf{X}|\mathbf{W};\theta,\beta)\\
&= \boxed{X_sX_t - \mathbb{E}_{P(\mathbf{X}|\mathbf{W};\theta,\beta)}[X_sX_t]}
\end{aligned}$$

-

$$
\begin{aligned}
\frac{\partial log P(\mathbf{X}|\mathbf{W};\theta,\beta)}{\partial \beta_{su}} &= \frac{\partial}{\partial \beta_{su}}\left(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u - logZ(\mathbf{W},\theta,\beta)\right)\\
&= X_sW_u - \frac{\partial}{\partial \beta_{su}} logZ(\mathbf{W},\theta,\beta)\\
&= X_sW_u - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\frac{\partial}{\partial \beta_{su}}\sum_{\mathbf{X}} exp\left(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\right)\\
&= X_sW_u - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}}\frac{\partial}{\partial \beta_{su}} exp\left(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\right)\\
&= X_sW_u - \frac{1}{Z(\mathbf{W},\theta,\beta)}\cdot\sum_{\mathbf{X}} X_sW_u exp\left(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\right)\\
&= X_sW_u - \sum_{\mathbf{X}} X_sW_u \frac{1}{Z(\mathbf{W},\theta,\beta)} exp\left(\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\right)\\
&= X_sW_u - \sum_{\mathbf{X}} X_sW_u P(\mathbf{X}|\mathbf{W};\theta,\beta)\\
&= X_sW_u - W_u\sum_{\mathbf{X}} X_s P(\mathbf{X}|\mathbf{W};\theta,\beta)\\
&= \boxed{X_sW_u - W_u\mathbb{E}_{P(\mathbf{X}|\mathbf{W};\theta,\beta)}[X_s]}
\end{aligned}
$$

4 Hidden Markov Model

Problem 4 The codes, results, and `README.md` are in `HWK_1_Codes.zip`. Please read `README.md` for the detailed instruction.

1. Baseline
   Figure 4 shows the evaluation result of 4.1.

```
(base)  20-02-08  19:05  ~/GoogleDrive/Gatech/pgm/HW/hw1/
coding-prob4/hmm-sol-haekyu/code  python baseline.py
Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

         precision      recall          F1-Score
GENE:    0.158861       0.660436        0.256116
```

Figure 4: The evaluation result of 4.1.

2. HMM with trigram features
   Figure 5 shows the evaluation result of 4.2.

```
(base)  20-02-08  19:05  ~/GoogleDrive/Gatech/pgm/HW/hw1/
coding-prob4/hmm-sol-haekyu/code  python hmm_trigram.py
Found 373 GENEs. Expected 642 GENEs; Correct: 202.

         precision      recall          F1-Score
GENE:    0.541555       0.314642        0.398030
```

Figure 5: The evaluation result of 4.2.

5 Markov Network

Problem 5 With the potentials $\phi(x_i, x_j)$, we can factorize the joint distribution as follows, where Z is the normalization value.

$$p(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z}\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\phi(x_5, x_1)$$

By summing over x_3 and x_4, we get

$$p(x_1, x_2, x_5) = \frac{1}{Z}\phi(x_1, x_2)\phi(x_5, x_1)\sum_{x_3, x_4}\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)$$

By summing over x_1 and x_3, we get

$$p(x_2, x_4, x_5) = \frac{1}{Z}\phi(x_4, x_5)\sum_{x_1, x_3}\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_5, x_1)$$

By summing over x_1 and x_5, we get

$$p(x_2, x_3, x_4) = \frac{1}{Z}\phi(x_2, x_3)\phi(x_3, x_4)\sum_{x_1, x_5}\phi(x_1, x_2)\phi(x_4, x_5)\phi(x_5, x_1)$$

By summing over x_1, x_3, and x_4, we get

$$p(x_2, x_5) = \frac{1}{Z}\sum_{x_1, x_3, x_4}\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\phi(x_5, x_1)$$

By summing over x_1, x_3, and x_5, we get

$$p(x_2, x_4) = \frac{1}{Z}\sum_{x_1, x_3, x_5}\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\phi(x_5, x_1)$$

Let's compute $p(x_1, x_2, x_5)p(x_2, x_4, x_5)p(x_2, x_3, x_4)$.

$$\begin{aligned}
p(x_1, x_2, x_5)p(x_2, x_4, x_5)p(x_2, x_3, x_4) &= \frac{1}{Z^3}\phi(x_1, x_2)\phi(x_5, x_1)\phi(x_4, x_5)\phi(x_2, x_3)\phi(x_3, x_4) \\
&\quad \cdot \left(\sum_{x_3, x_4}\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\right) \\
&\quad \cdot \left(\sum_{x_1, x_3}\phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_5, x_1)\right) \\
&\quad \cdot \left(\sum_{x_1, x_5}\phi(x_1, x_2)\phi(x_4, x_5)\phi(x_5, x_1)\right) \\
&= \frac{1}{Z^2} \cdot p(x_1, x_2, x_3, x_4, x_5) \\
&\quad \cdot \left(\sum_{x_3, x_4}\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\right) \\
&\quad \cdot \left(\sum_{x_1}\phi(x_1, x_2)\phi(x_5, x_1)\right)\left(\sum_{x_3}\phi(x_2, x_3)\phi(x_3, x_4)\right) \\
&\quad \cdot \left(\sum_{x_1, x_5}\phi(x_1, x_2)\phi(x_4, x_5)\phi(x_5, x_1)\right)
\end{aligned}$$

Let's compute $p(x_2, x_5)p(x_2, x_4)$.

$$
\begin{aligned}
p(x_2, x_5)p(x_2, x_4) &= \frac{1}{Z^2} \\
&\quad \cdot \Bigg(\sum_{x_1, x_3, x_4} \phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\phi(x_5, x_1) \Bigg) \\
&\quad \cdot \Bigg(\sum_{x_1, x_3, x_5} \phi(x_1, x_2)\phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5)\phi(x_5, x_1) \Bigg) \\
&= \frac{1}{Z^2} \\
&\quad \cdot \Bigg(\sum_{x_1} \phi(x_1, x_2)\phi(x_5, x_1) \sum_{x_3, x_4} \phi(x_2, x_3)\phi(x_3, x_4)\phi(x_4, x_5) \Bigg) \\
&\quad \cdot \Bigg(\sum_{x_3} \phi(x_2, x_3)\phi(x_3, x_4) \sum_{x_1, x_5} \phi(x_1, x_2)\phi(x_4, x_5)\phi(x_5, x_1) \Bigg)
\end{aligned}
$$

Thus,

$$
\frac{p(x_1, x_2, x_5)p(x_2, x_4, x_5)p(x_2, x_3, x_4)}{p(x_2, x_5)p(x_2, x_4)} = p(x_1, x_2, x_3, x_4, x_5)
$$

References