Problem 1.1

1. **False.** Active path from Season $\rightarrow$ Flu $\rightarrow$ Chills.
2. **True.** All paths are blocked: Flu blocks all possible paths.
3. **False.** Active path from Season $\rightarrow$ Dehydration $\rightarrow$ Headache.
4. **True.** All paths are blocked: Flu blocks one path and Dehydration blocks the other path.
5. **False.** Active path from Season $\rightarrow$ Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea.
6. **True.** All paths are blocked: Headache blocks Season $\rightarrow$ Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea, Dehydration blocks Season $\rightarrow$ Dehydration $\rightarrow$ Nausea, and Dehydration blocks Season $\rightarrow$ Dehydration $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea.
7. **False.** Active path from Flu $\rightarrow$ Season $\rightarrow$ Dehydration.
8. **False.** Since Flu $\rightarrow$ Headache $\leftarrow$ Dehydration is a v-structure and Headache is observed, that path becomes activate, so Flu $\rightarrow$ Headache $\rightarrow$ Dehydration is an active path.
9. **True.** All paths are blocked: Season blocks Flu $\rightarrow$ Season $\rightarrow$ Dehydration and Headache, because it's not observed, blocks Flu $\rightarrow$ Headache $\rightarrow$ Dehydration.
10. **False.** Active path from Flu $\rightarrow$ Headache $\rightarrow$ Dizziness $\rightarrow$ Nausea $\rightarrow$ Dehydration because Dizziness $\rightarrow$ Nausea $\leftarrow$ Dehydration is a v-structure and Nausea is observed, making it active.
11. **False.** Active path from Chills $\rightarrow$ Flu $\rightarrow$ Season $\rightarrow$ Dehydration $\rightarrow$ Nausea.
12. **False.** Active path from Chills $\rightarrow$ Flu $\rightarrow$ Season $\rightarrow$ Dehydration $\rightarrow$ Nausea.

Problem 1.2

1. P(S, F, D, C, H, Z, N) = P(S)P(F | S)P(C | F)P(D | S)P(H | F, D)P(Z | H)P(N | D, Z)

2. Let K = $\phi_1(S)\phi_2(F)\phi_3(D)\phi_4(C)\phi_5(H)\phi_6(N)\phi_7(Z)\phi_8(S,F)\phi_9(F,C)\phi_{10}(N,Z)\phi_{11}(F,H)$ $\phi_{12}(F,H)\phi_{13}(D,H)\phi_{14}(D,N)\phi_{15}(H,Z)$

   The factorized form is
   $$\frac{K}{\sum_{S,F,D,C,H,N,Z} K}$$

   In case it is difficult to read, the denominator is the summation over all random variables with the value in the summation being the value K (a user would just need to substitute the correct values into K for each iteration).

Problem 1.3

1. 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(0.5) + 0.1(0.5)
   = 0.2 + 0.05
   = **0.25**

2. P(F = true | S = winter) = **0.4** (it's in the second CPD table)

3. To aid in solving this problem, I will compute 2 new probabilities given the CPDs:

   $P(S = \text{winter}|F = \text{true}) = \frac{P(F=\text{true}|S=\text{winter})P(S=\text{winter})}{P(F=\text{true})} = \frac{0.4(0.5)}{0.25} = 0.8$

   $P(S = \text{winter}|F = \text{false}) = \frac{P(F=\text{false}|S=\text{winter})P(S=\text{winter})}{P(F=\text{false})} = \frac{0.6(0.5)}{0.75} = 0.4$

   So, now we can calculate the final answer as such:

   P(F = true | S = winter, H = true)
   $= \frac{P(S=\text{winter},H=\text{true}|F=\text{true})P(F=\text{true})}{P(S=\text{winter},H=\text{true})}$
   $= \frac{P(H=\text{true}|S=\text{winter},F=\text{true})P(S=\text{winter}|F=\text{true})P(F=\text{true})}{\sum_F P(S=\text{winter},H=\text{true})P(F)}$
   $= \frac{P(H=\text{true}|D,F=\text{true})P(D|S=\text{winter})P(S=\text{winter}|F=\text{true})P(F=\text{true})}{\sum_F P(H=\text{true}|S=\text{winter})P(S=\text{winter}|F=f)P(F=f)}$
   $= \frac{\sum_D P(H=\text{true}|D=d,F=\text{true})P(D=d|S=\text{winter})P(S=\text{winter}|F=\text{true})P(F=\text{true})}{\sum_{D,F} P(H=\text{true}|D=d,F=f)P(D=d|S=\text{winter})P(S=\text{winter}|F=f)P(F=f)}$
   $= \frac{0.8(0.9)(0.8)(0.25)+0.9(0.1)(0.8)(0.25)}{0.3(0.9)(0.4)(0.75)+0.8(0.9)(0.8)(0.25)+0.8(0.1)(0.4)(0.75)+0.9(0.1)(0.8)(0.25)}$
   $= \frac{0.018+0.144}{(0.018+0.144)+(0.024+0.081)}$
   $= \frac{0.162}{0.267}$
   $=$ **0.6067**

4. P(F | S = winter, H = true, D = true)
   $= \frac{P(H=\text{true}|D=\text{true},F=\text{true})P(D=\text{true}|S=\text{winter})P(S=\text{winter}|F=\text{true})P(F=\text{true})}{\sum_F P(H=\text{true}|D=\text{true},F=f)P(D=\text{true}|S=\text{winter})P(S=\text{winter}|F=f)P(F=f)}$
   $= \frac{0.9(0.1)(0.8)(0.25)}{0.8(0.1)(0.4)(0.75)+0.9(0.1)(0.8)(0.25)}$
   $= \frac{0.018}{0.024+0.018}$
   $= \frac{0.018}{0.042}$
   $=$ **0.4286**

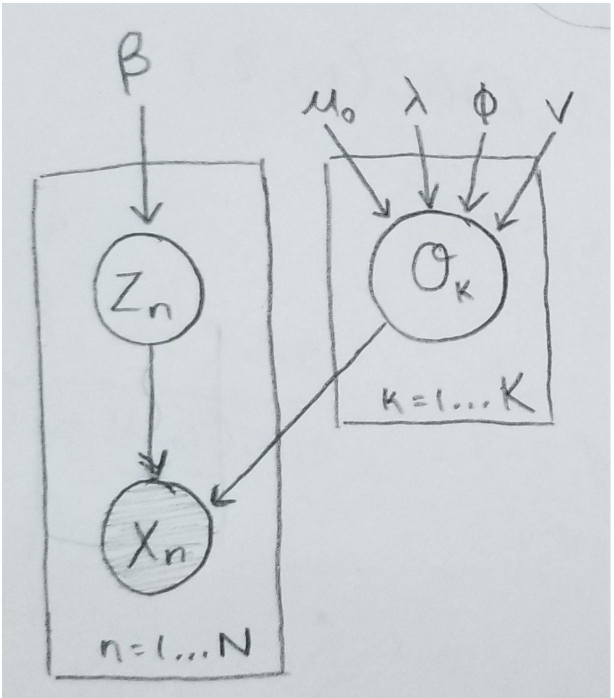
5. Dehydration causes a decrease in your likelihood of having the flu. This is an appropriate conclusion because when we know you are dehydrated, it explains away the effect of the headache on having a flu.

Problem 1.4

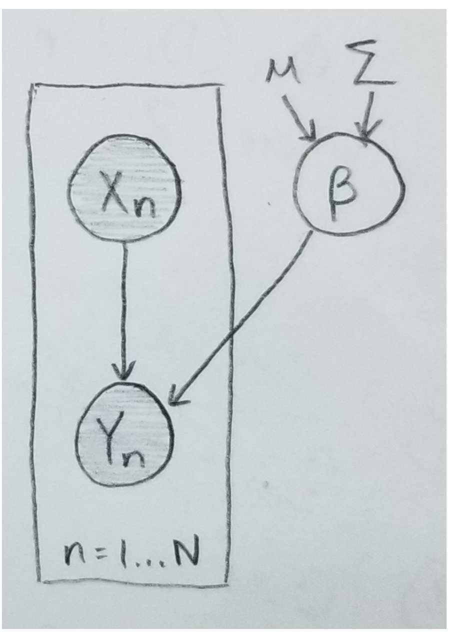
1. There are no marginal independences in Figure 1 and 2, so no there are no differences.
2. Yes. In Figure 1, Flu $\perp\!\!\!\perp$ Dehydration | Season, but this is not the case in Figure 2 because the trail Flu $\rightarrow$ Headache $\rightarrow$ Dehydration exists.

Problem 2.1

1. See the figure drawn below. Please note the slightly shaded in circle (X_n).



2. See the figure drawn below. Please note the slightly shaded in circles (X_n and Y_n).



Problem 2.2

Let us assume P factorizes according to G.

We know $\mathbf{X} = X_i \cup \mathrm{Pa}(X_i) \cup \mathrm{NonDesc}(X_i) \cup \mathrm{Desc}(X_i)$.

From that, we can show the following:

$P(X_i|\mathrm{Pa}(X_i), \mathrm{NonDesc}(X_i))$

$= \frac{P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$ (law of total probability)

$= \frac{\sum_{X_m\in\mathrm{Desc}(X_i)} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i),X_m)}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$

$= \frac{\sum_{X_m\in\mathrm{Desc}(X_i)} \prod_{j=1}^{n} P(X_j|\mathrm{Pa}(X_j))}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$ (where n = number of vertices in the graph)

$= \frac{\sum_{X_m\in\mathrm{Desc}(X_i)} \left[P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} \prod_{X_k\in\mathrm{Pa}(X_i)} \prod_{X_m\in\mathrm{Desc}(X_i)} P(X_j|\mathrm{Pa}(X_j))P(X_k|\mathrm{Pa}(X_k))P(X_m|\mathrm{Pa}(X_m))\right]}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$

$= \frac{\sum_{X_m\in\mathrm{Desc}(X_i)} \left[P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k)) \prod_{X_m\in\mathrm{Desc}(X_i)} P(X_m|\mathrm{Pa}(X_m))\right]}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$

$= \frac{P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k)) \sum_{X_m\in\mathrm{Desc}(X_i)} \prod_{X_m\in\mathrm{Desc}(X_i)} P(X_m|\mathrm{Pa}(X_m))}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$

We know $\sum_{X_m\in\mathrm{Desc}(X_i)} \prod_{X_m\in\mathrm{Desc}(X_i)} P(X_m|\mathrm{Pa}(X_m)) = 1$, so this term can be removed from the numerator. We can continue as follows:

$= \frac{P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k))}{\sum_{x\in X_i} P(X_i,\mathrm{Pa}(X_i),\mathrm{NonDesc}(X_i))}$

$= \frac{P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k))}{\sum_{x\in X_i} P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k))}$

$= \frac{P(X_i|\mathrm{Pa}(X_i)) \prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k))}{\prod_{X_j\in\mathrm{NonDesc}(X_i)} P(X_j|\mathrm{Pa}(X_j)) \prod_{X_k\in\mathrm{Pa}(X_i)} P(X_k|\mathrm{Pa}(X_k))}$ (since $\sum_{x\in X_i} P(X_i|\mathrm{Pa}(X_i)) = 1$)

$= P(X_i|\mathrm{Pa}(X_i))$

So, we have shown $P(X_i|\mathrm{Pa}(X_i), \mathrm{NonDesc}(X_i)) = P(X_i|\mathrm{Pa}(X_i))$.

This means $X_i \perp\!\!\!\perp \mathrm{NonDesc}(X_i) \mid \mathrm{Pa}(X_i)$.

$\therefore$ G is an I-map of P $\square$

Problem 2.3

We want to show d-sep$_G(X;Y|Z)$ iff sep$_H(X;Y|Z)$.

First, we will show d-sep$_G(X;Y|Z) \implies$ sep$_H(X;Y|Z)$.

Let us assume d-sep$_G(X;Y|Z)$.

Because of d-sep$_G(X;Y|Z)$, there is no active trail between any node $x \in X$ and $y \in Y$ given Z. In other words, all trails between X and Y are inactive given Z.

By the definition of an active trail in a Bayesian network, it must satisfy both of the following conditions: (1) For all V structures $m_{i-1} \rightarrow m_i \leftarrow m_{i+1}$, then $m_i \in Z$ or $\text{Desc}(m_i) \in Z$, (2) no other node in the trail exists in Z. Since all trails are inactive, all trails must violate one of the two conditions above.

As such, for each trail $T : M_i \rightarrow \cdots \rightarrow M_n$ there are two cases:

1. T violates condition 1, which means for a given V structure $m_{j-1} \rightarrow m_j \leftarrow m_{j+1}$, we have $m_j, \text{Desc}(m_j) \notin Z$. Let $P = m_j \cup \text{Desc}(m_j)$. Since m_j and its descendants are not observed, we have $P \notin U \cup \text{Ancestors}_U \implies P \notin H$. As such, that V-structure, which consists solely of inactive nodes already, is nonexistent in H, so it is an inactive trail in H by definition.
2. T violates condition 2, which means there exists at least one observed node m_j. In this case, this exact trail is inactive by definition in the Markov network H anyway.

$\therefore$ We have shown d-sep$_G(X;Y|Z) \implies$ sep$_H(X;Y|Z)$.

Now, we will show sep$_H(X;Y|Z) \implies$ d-sep$_G(X;Y|Z)$.

Let us assume sep$_H(X;Y|Z)$.

Because of sep$_H(X;Y|Z)$, there is no active trail between any node $x \in X$ and $y \in Y$ given Z. This means all trails between X and Y are inactive given Z.

By the definition of an active trail in a Markov network, it must satisfy one condition: (1) for all nodes in the trail, no node exists in Z. Since all trails are inactive, they all break this condition. There are three necessary cases to show:

1. A V structure exists in G (moralized parents in H): Two parents (m_{i-1}, m_{i+1}) in a V structure $m_{i-1} \rightarrow m_j \leftarrow m_{i+1}$ are moralized in H. This means $m_j \in Z$ or $\text{Desc}(m_j) \in Z$, so this V structure is active in G. As such, the moralized parents create a mini active trail in H (m_{i-1} — m_{i+1}), following the precedent that that V structure is active in G due to the presence of an observed m_j or descendant of m_j. So, these moralized parents, which produce an active trail in H, represent an active trail in G.
2. A V structure exists in G (non-moralized parents in H): Two parents (m_{i-1}, m_{i+1}) in a V structure $m_{i-1} \rightarrow m_j \leftarrow m_{i+1}$ are non-moralized in H. If the two parents are not moralized, then $m_j, \text{Desc}(m_j) \notin Z$, which means that V structure is not active in G, or else m_j or $\text{Desc}(m_j)$ would be in $U \cup \text{Ancestors}_U = H$. So, there is no direct path in H

between the parents (m_{i-1} — m_{i+1}), which as shown above, implies that V structure is inactive in G. This indicates this inactive path is inactive in G.

3. A causal/evidential trail or common cause exists in G: $m_i \in Z$ exists on a trail $T \in H$, so T is inactive in G due to the definition of a casual/evidential trail or common cause.

$\therefore$ We have shown $\text{sep}_H(X;Y|Z) \implies \text{d-sep}_G(X;Y|Z)$.

$\therefore$ We have shown both directions and $\text{d-sep}_G(X;Y|Z)$ iff $\text{sep}_H(X;Y|Z)$.

Problem 2.4

The joint distribution for the non-marginalized bayesian network is as follows:

P(B, E, T, N, J, M, A) = P(B)P(E)P(T)P(N)P(A | B, E)P(J | A, T)P(M | A, N)

We want to calculate the marginalized distribution P(B, E, T, N, J, M).

(*) Since we are marginalizing over A, we know P(A | B, E) = 1.

Now, we know, a marginalized joint distribution over some variable x_n means $P(x_1, \ldots, x_n) = P(x_1, \ldots, x_{n-1})$. Using this rule, we can infer the following:

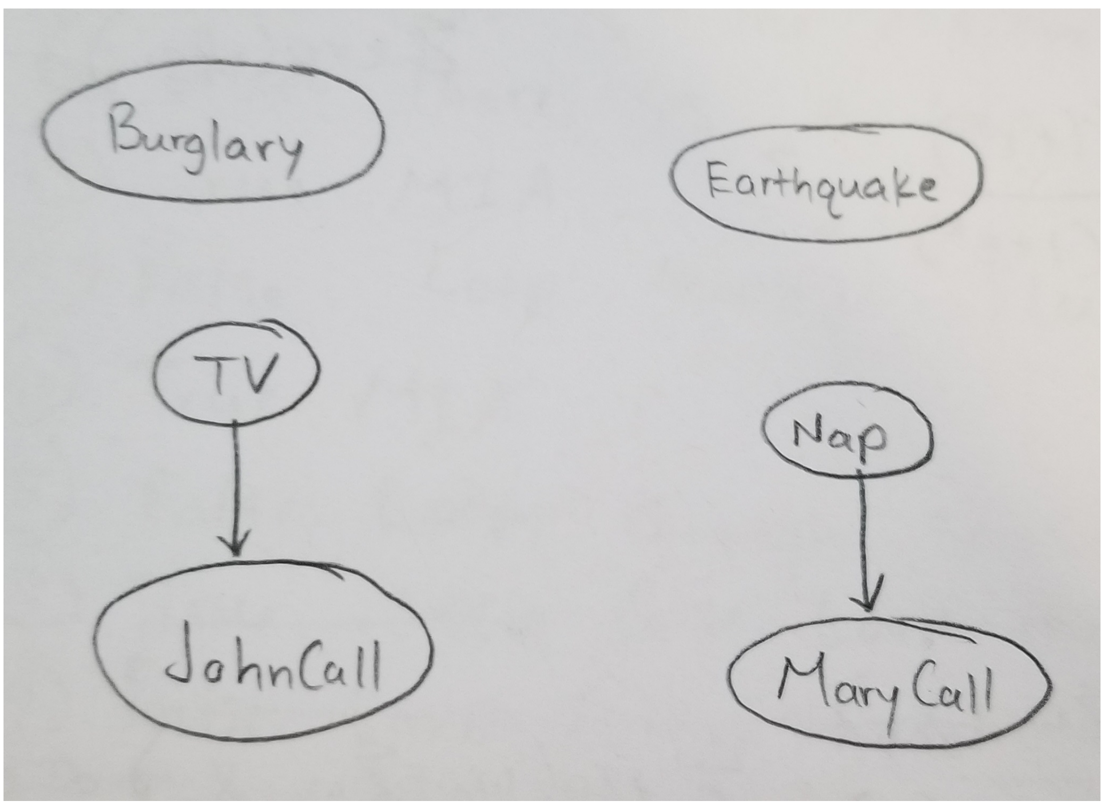
(@) $P(J|A,T) = \frac{P(J,A,T)}{P(A,T)} = \frac{P(J,T)}{P(T)} = P(J|T)$

(#) $P(M|A,N) = \frac{P(M,A,N)}{P(A,N)} = \frac{P(M,N)}{P(N)} = P(M|N)$

As such, using (*), (@), and (#), we can reduce the joint distribution to the marginalized distribution to get:

P(B, E, T, N, J, M) = P(B)P(E)P(T)P(N)P(J | T)P(M | N)

So, with this new joint distribution after marginalizing over A (Alarm) in the original joint distribution, we can see the only dependencies are (J, T) and (M, N). The resultant bayesian network showing this minimal I-map is as follows:



Problem 2.5

First, we will show I satisfies strong union ($X \perp\!\!\!\perp Y | Z \implies X \perp\!\!\!\perp Y | Z, W$).

Assume $X \perp\!\!\!\perp Y | Z$.

Since $X \perp\!\!\!\perp Y | Z$, all trails $X \to \cdots \to Y$ are blocked by an observed variable $z \in Z$ in a given three-node structure.

So, observing W will only block additional three-node structures, keeping all independences intact.

$\therefore$ I satisfies strong union.

Next, we will show I satisfies transitivity $[\neg(X \perp\!\!\!\perp A | Z) \& \neg(A \perp\!\!\!\perp Y | Z) \implies \neg(X \perp\!\!\!\perp Y | Z)]$.

Assume X depends on A ($\neg(X \perp\!\!\!\perp A | Z)$) and A depends on Y ($\neg(A \perp\!\!\!\perp Y | Z)$).

Since X depends on A (correlated), there exists trail T_1 such that $T_1 : X \to \cdots \to A$ is active (no node on T_1 is observed).

Since A depends on Y (correlated), there exists trail T_2 such that $T_2 : A \to \cdots \to Y$ is active (no node on T_2 is observed).

So, we can form a new path $T_3 : X \to \cdots \to A \to \cdots \to Y$ such that no node on T_3 is observed.

Since no node on T_3 is observed, that path is active.

So, since T_3 is active, we can simplify T_3 such that $T_3 : X \to \cdots \to Y$, which is the same active path already defined above with simplified notation.

Finally, since there is an active path between X and Y, then X and Y are correlated ($\neg(X \perp\!\!\!\perp Y | Z)$).

We have shown $\neg(X \perp\!\!\!\perp A | Z) \& \neg(A \perp\!\!\!\perp Y | Z) \implies \neg(X \perp\!\!\!\perp Y | Z)$.

$\therefore$ I satisfies transitivity.

$\therefore$ We have shown I satisfies both strong union and transitivity. $\square$

Problem 3

1. First, let us define a logistic regression model as having the form $\frac{\exp(Z)}{\exp(Z)+1}$.

   Let $P(X_i = 0, X_{-i}; \theta) \propto \exp\{\sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t\} = \exp\{0\} = 1$.

   Let $P(X_i = 1, X_{-i}; \theta) \propto \exp\{\sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t\}$
   $= \exp\{\sum_{s\in V} \theta_s + \sum_{(i,t)\in E} \theta_{i,t} X_t\}$.

   Now, by the law of total probability, we get the following:

   $P(X_i = 1|X_{-i}; \theta) \propto \frac{P(X_i=1,X_{-i})}{P(X_i=1,X_{-i})+P(X_i=0,X_{-i})}$
   $= \frac{\exp\{\sum_{s\in V} \theta_s + \sum_{(i,t)\in E} \theta_{i,t} X_t\}}{\exp\{\sum_{s\in V} \theta_s + \sum_{(i,t)\in E} \theta_{i,t} X_t\}+1}$

   If we let $Z = \exp\{\sum_{s\in V} \theta_s + \sum_{(i,t)\in E} \theta_{i,t} X_t\}$, we see $P(X_i = 1|X_{-i}; \theta) \propto \frac{\exp(Z)}{\exp(Z)+1}$, which is the exact equation representing the logistic regression model as defined above.

   $\therefore$ The conditional distribution is given by a logistic regression model. $\square$

2. The log likelihood of the probability distribution is as follows:

   $P(X|W; \theta, \beta) = \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\} / Z(W, \theta, \beta)$
   $\implies \log P(X|W; \theta, \beta) = \log\exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\} / Z(W, \theta, \beta)$
   $\implies \log P(X|W; \theta, \beta) = \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u - \log Z(W, \theta, \beta)$

   Now, to calculate the gradients:

   $\frac{\delta}{\delta\theta_s} \log P(X|W; \theta, \beta) = X_s - \frac{Z'(W,\theta,\beta)}{Z(W,\theta,\beta)}$
   $= X_s - \frac{\sum_X X_s \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\}}{Z(W,\theta,\beta)}$
   $= X_s - \sum_X X_s \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\} / Z(W, \theta, \beta)$
   $= X_s - \sum_X X_s P(X|W, \theta, \beta)$
   $= X_s - E[X_s]$

   $\frac{\delta}{\delta\theta_{s,t}} \log P(X|W; \theta, \beta) = X_s X_t - \frac{Z'(W,\theta,\beta)}{Z(W,\theta,\beta)}$
   $= X_s X_t - \frac{\sum_X X_s X_t \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\}}{Z(W,\theta,\beta)}$
   $= X_s X_t - \sum_X X_s X_t \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\} / Z(W, \theta, \beta)$
   $= X_s X_t - \sum_X X_s X_t P(X|W, \theta, \beta)$
   $= X_s X_t - E[X_s X_t]$

   $\frac{\delta}{\delta\beta_{su}} \log P(X|W; \theta, \beta) = X_s W_u - \frac{Z'(W,\theta,\beta)}{Z(W,\theta,\beta)}$
   $= X_s W_u - \frac{\sum_X X_s W_u \exp\left\{ \sum_{s\in V} \theta_s X_s + \sum_{(s,t)\in E} \theta_{s,t} X_s X_t + \sum_{s\in V, u\in[q]} \beta_{su} X_s W_u \right\}}{Z(W,\theta,\beta)}$

$= X_sW_u - \sum_X X_sW_u \exp\Big\{\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V,u\in[q]}\beta_{su}X_sW_u\Big\}/Z(W,\theta,\beta)$
$= X_sW_u - \sum_X X_sW_uP(X|W,\theta,\beta)$
$= X_sW_u - W_uE[X_s]$

Problem 4

1. **NOTE:** All code is developed in Python3, so make sure that is installed with the "python3" executable readily available. Other than that, no other dependencies are required. You can run this section of the code by running the following commands in order:

   1) Extract hmm_data.zip and navigate to the root directory.

   2) python3 rare_replace.py gene.train gene.train.rare

   3) python3 count_freqs.py gene.train.rare > gene.train.rare.counts

   4) python3 hmm.py baseline gene.train.rare.counts gene.test gene_test.p1.out

   5) python3 eval_gene_gene_tagger.py gene.key gene_test.p1.out

   The output from the results/evaluation program can be seen below:

```
Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

         precision           recall              F1-Score
GENE:    0.15886099662795056    0.660436137071651       0.25611597704620964
```

2. **NOTE:** This part assumes you have already extracted the hmm_data.zip file and ran through the steps to evaluate the baseline model. The trigram HMM can be ran and evaluated by running the following commands:

   1) python3 hmm.py trigram gene.train.rare.counts gene.test gene_test.p2.out

   2) python3 eval_gene_tagger.py gene.key gene_test.p2.out

   The output from the results/evaluation program can be seen below:

```
Found 325 GENEs. Expected 642 GENEs; Correct: 187.

         precision           recall              F1-Score
GENE:    0.5753846153846154     0.29127725856697817     0.3867631851085832
```

Problem 5

The joint distribution can be represented as:

$P(x_1, x_2, x_3, x_4, x_5)$

$= P(x_1|x_2, x_3, x_4, x_5)P(x_2, x_3, x_4, x_5)$

$= P(x_1|x_2, x_5)P(x_2, x_3, x_4, x_5)$

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

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

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

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

$= \frac{P(x_1,x_2,x_5)P(x_2,x_4,x_5)P(x_2,x_3,x_4)}{P(x_2,x_5)P(x_2,x_4)} \quad \square$

The probability distribution can then be written as:

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

We will try to reach this above factorization by marginalizing each distribution term in the numerator and denominator of the following joint distribution given to us:

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

We know the following factorizations of the marginal distributions exist below:

$P(x_1, x_2, x_5) = \frac{\phi(x_1,x_2)\phi(x_2,x_5)\phi(x_1,x_5)}{\sum_{x_1,x_2,x_5} \phi(x_1,x_2)\phi(x_2,x_5)\phi(x_1,x_5)}$

$P(x_2, x_4, x_5) = \frac{\phi(x_2,x_4)\phi(x_4,x_5)\phi(x_2,x_5)}{\sum_{x_2,x_4,x_5} \phi(x_2,x_4)\phi(x_4,x_5)\phi(x_2,x_5)}$

$P(x_2, x_3, x_4) = \frac{\phi(x_2,x_3)\phi(x_3,x_4)\phi(x_2,x_4)}{\sum_{x_2,x_3,x_4} \phi(x_2,x_3)\phi(x_3,x_4)\phi(x_2,x_4)}$

$P(x_2, x_5) = \frac{\phi(x_2,x_5)}{\sum_{x_2,x_5} \phi(x_2,x_5)}$

$P(x_2, x_4) = \frac{\phi(x_2,x_4)}{\sum_{x_2,x_4} \phi(x_2,x_4)}$

So, when we combine them, we get the following:

$P(x_1, x_2, x_3, x_4, x_5)$

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

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

We cancel out $\phi(x_2, x_4)$, $\phi(x_2, x_5)$, $\sum_{x_2,x_4}\phi(x_2, x_4)$, and $\sum_{x_2,x_5}\phi(x_2, x_5)$ from both the numerator and denominator:

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

Then, we can rearrange the factors to make them more visibly appealing:

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

And finally, since the edges (x_2, x_4) and (x_2, x_5) do not exist, those edge potentials, $\phi(x_2, x_5)$ and $\phi(x_2, x_4)$, both equal 1, so they can be removed from the numerator and denominator, giving us our final factorized form of the joint probability, shown above, based on the individual marginal probability distributions:

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

So, by combining the individual factorizations of each marginal probability distribution term in the numerator and denominator of the given equation, we can see it equals the complete pairwise potential factorization of the entire joint distribution of the markov network.