Homework 1

Biswadeep Chakraborty

2/6/2020

1 QUESTION 1

1.1 Solutions to Question 1.1

Sl. No.	Independence Statement	True or False	Explanation
1	$Season \perp Chills$	**False**	Without Flu being observed, Season and Chills are not d-separated
2	$Season \perp Chills \mid Flu$	**True**	Since Flu is observed, Flu and Chills are d-separated
3	$Season \perp Headache \mid Flu$	**False**	Headache is dependent on Season following the path of Dehydration, though Flu is blocked
4	$Season \perp Headache \mid Flu, Dehydration$	**True**	Since both the parent nodes of Headache are blocked, Season and Headache are d-separated
5	$Season \perp Nausea \mid Dehydration$	**False**	Season and Nausea are connected by the path $Season \rightarrow Flu \rightarrow Headache \rightarrow Dizziness \rightarrow Nausea$. Hence they are not independent
6	$Season \perp Nausea \mid Dehydration, Headache$	**True**	Since both Headache and Dehydration are observed, all the paths to Nausea from Season are blocked. Hence they are independent
7	$Flu \perp Dehydration$	**False**	They are children of the same parent, Season. Hence, not d-separated
8	$Flu \perp Dehydration \mid Season, Headache$	**False**	Since Headache is a V-node, and Headache is observed, so Flu and Dehydration are not d-separated
9	$Flu \perp Dehydration \mid Season$	**True**	Since Headache is a V-node, and neither Headache nor its descendants are observed, so Flu and Dehydration are d-separated
10	$Flu \perp Dehydration \mid Season, Nausea$	**False**	Since Headache is a V-node, and Nausea, which is a descendent of Headache ($Headache \rightarrow Dizziness \rightarrow Nausea$), is observed, so Flu and Dehydration are not d-separated
11	$Chills \perp Nausea$	**False**	Both are the children of the same parent Season and are thus not d-separated.
12	$Chills \perp Nausea \mid Headache$	**False**	There are now two paths by which Chills and Nausea are connected. One through Season, and another through the observed V-node Headache.

1.2 Solutions to Question 1.2

1.2.1 Part 1:

$$P(S, F, D, C, H, Z, N) = P(S)P(F|S)P(D|S)P(C|F)P(H|F, D)P(Z|H)P(N|D, Z) \quad (1)$$

1.2.2 Part 2:

$$
\begin{aligned}
P(S, F, D, C, H, Z, N) = \frac{1}{Z}\phi_1(S)\phi_2(F)\phi_3(D)\phi_4(C)\phi_5(H)\phi_6(N)\phi_7(Z)\phi_8(S, F)\phi_9(S, D)\phi_{10}(F, C) \\
\times \phi_{11}(F, H)\phi_{12}(D, H)\phi_{13}(D, N)\phi_{14}(H, Z)\phi_{15}(N, Z) \qquad (2)
\end{aligned}
$$

1.3 Solutions to Question 1.3

1.3.1 Part 1:

$$
\begin{aligned}
P(F = \text{ true }) &= \sum_{s=summer,winter} P(F = 1, S = s) \qquad (3) \\
&= \sum_{s=summer,winter} P(F = 1|S = s)P(S = s) \\
&= P(F = 1|S = winter)P(S = winter) + P(F = 1|S = summer)P(S = summer) \\
&= 0.4 \cdot 0.5 + 0.1 \cdot 0.5 \\
&= 0.25
\end{aligned}
$$

1.3.2 Part 2:

$$
P(F = 1|S = winter) = 0.4 \qquad (4)
$$

1.3.3 Part 3:

$$
\begin{aligned}
&P(F = 1|S = winter, H = 1) \qquad (5) \\
&= \frac{P(F = 1, S = \text{ winter }, H = 1)}{P(S = \text{ winter }, H = 1)} \\
&= \frac{\sum_d P(F = 1, S = \text{ winter, } H = 1, D = d)}{\sum_{f,d} P(F = f, S = \text{ winter }, H = 1, D = d)}, \quad where f, d = 0, 1 \\
&= \frac{\sum_{d\in 0,1} P(H = 1|F = 1, D = d)P(F = 1|S = \text{winter})P(D = d|S = \text{ winter })P(S = \text{ winter })}{\sum_{f,d\in 0,1} P(H = 1|F = f, D = d)P(F = f|S = \text{ winter })P(D = d|S = \text{ winter })P(S = \text{ winter })} \\
&= \frac{0.9 \times 0.4 \times 0.1 \times 0.5 + 0.8 \times 0.4 \times 0.9 \times 0.5}{0.9 \times 0.4 \times 0.1 \times 0.5 + 0.8 \times 0.4 \times 0.9 \times 0.5 + 0.8 \times 0.6 \times 0.1 \times 0.5 + 0.3 \times 0.6 \times 0.9 \times 0.5} \\
&= 0.6067
\end{aligned}
$$

1.3.4 Part 4:

$$
\begin{aligned}
&P(F=1|S=winter, H=1, D=1) \qquad (6)\\
&=\frac{P(F=1, S=\text{ winter }, H=1, D=1)}{P(S=\text{ winter }, H=1, D=1)}\\
&=\frac{P(F=1, S=\text{ winter, } H=1, D=1)}{\sum_f P(F=f, S=\text{ winter }, H=1, D=1)}, \quad where f=0,1\\
&=\frac{P(H=1|F=1, D=1)P(F=1|S=\text{winter})P(D=1|S=\text{ winter })P(S=\text{ winter })}{\sum_{f\in 0,1} P(H=1|F=f, D=1)P(F=f|S=\text{ winter })P(D=1|S=\text{ winter })P(S=\text{ winter })}\\
&=\frac{0.9\times 0.4\times 0.1\times 0.5}{0.9\times 0.4\times 0.1\times 0.5+0.8\times 0.6\times 0.1\times 0.5}\\
&=0.429
\end{aligned}
$$

Part 5: If the condition of dehydration is observed, then the probability of having flu is smaller. This is in accordance with our hypothesis since dehydration is a probable cause of headache and thus the possibility of having flu decreases.

1.4 Solutions to Question 1.4

1.4.1 Part 1:

Neither the directed Bayesian network nor the undirected markov network encode any marginal independence. Hence, there is no difference in between them with respect to marginal independences.

1.4.2 Part 2:

Since there are no problems of V-structures in markov networks, there are a number of differences between the two structures with respect to the conditional independence.
Eg. For Bayesian network, $Flu \perp Dehydration|Season, Headache$ which is not true for Markov network.

2 Question 2

2.1 Solutions to Question 2.1

2.1.1 Part 1:

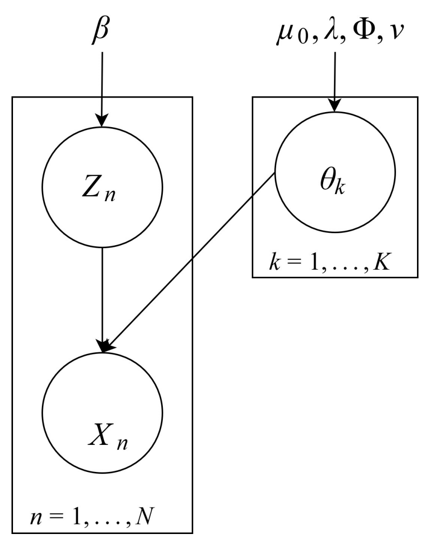

Figure 1: Plate Notation of Gaussian Mixture Model

2.1.2 Part 2:

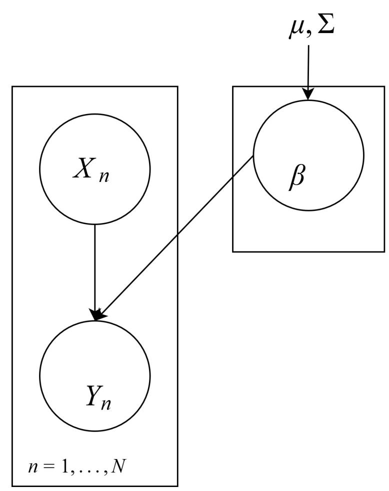

Figure 2: Plate Notation of Bayesian Logistic Regression Model

2.2 Solutions to Question 2.2

Let us assume that the p factorizes over $\mathcal{G}$. Let X_{D_i} denote the descendents of X_i and X_{N_i} denote non-descendents. Note that $\{X_1 \ldots X_n\} = \{X_i\} \cup X_{\pi_i} \cup X_{D_i} \cup X_{N_i}$. We have:

$$p\left(x_i | x_{\pi_i}, x_{N_i}\right) = \frac{p\left(x_i, x_{\pi_i}, x_{N_i}\right)}{\sum_{x_i \in \Omega_{X_i}} p\left(x_i, x_{\pi_i}, x_{N_i}\right)} \tag{7}$$

We compute the numerator:

$$p\left(x_i, x_{\pi_i}, x_{N_i}\right) = \sum_{x_{D_i}} p\left(x_i, x_{\pi_i}, x_{N_i}, x_{D_i}\right) = \sum_{x_{D_i}} \prod_{j=1}^{n} p\left(x_j | x_{\pi_j}\right) \tag{8}$$

$$= p\left(x_i | x_{\pi_i}\right) \prod_{x_j \in x_{N_i}} p\left(x_j | x_{\pi_j}\right) \prod_{x_k \in x_{\pi_i}} p\left(x_k | x_{\pi_k}\right) \sum_{x_{D_i} l \in X_{D_i}} p\left(x_l | x_{\pi_l}\right) \tag{9}$$

The last factor above is 1 The denominator of the fraction is:

$$\sum_{x_i \in \Omega_{X_i}} p\left(x_i, x_{\pi_i}, x_{N_i}\right) = \sum_{x_i \in \Omega_{X_i}} p\left(x_i | x_{\pi_i}\right) \prod_{x_j \in x_{N_i}} p\left(x_j | x_{\pi_j}\right) \prod_{x_k \in x_{\pi_i}} p\left(x_k | x_{\pi_k}\right) \tag{10}$$

$$= \prod_{x_j \in x_{N_i}} p\left(x_j | x_{\pi_j}\right) \prod_{x_k \in x_{\pi_i}} p\left(x_k | x_{\pi_k}\right) \tag{11}$$

Putting these back together in the fraction, we get:

$$p\left(x_i|x_{\pi_i}, x_{N_i}\right) = p\left(x_i|x_{\pi_i}\right) \Longrightarrow X_i \perp X_{N_i}|X_{\pi_i}$$

which means that $\mathcal{G}$ is an I-map of p.

2.3 Solutions to Question 2.3

We prove by negation. So, let us suppose, d-sep$(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$ does not hold true i.e., $(\boldsymbol{X} \not\perp \boldsymbol{Y}|\boldsymbol{Z})$ Then there is an active path in $\mathcal{G}$ between some $X \in \boldsymbol{X}$ and some $Y \in \boldsymbol{Y}$. Let us canister that the active path is formed of overlapping unidirectional segments $\mathcal{N}_1 \to \ldots \to \mathcal{N}_n$. where $\mathcal{N}_i$ belongs to either X, Y , or active V-structure. As such, either $\mathcal{N}_n$ or one of its descendants is in U, and so all nodes and edges in the segment $\mathcal{N}_1 \to \ldots \to \mathcal{N}_n$ are in the induced graph over $\texttt{Ancestors}(U) \cup U$. So the active path in $\mathcal{G}$ is a path in $\mathcal{H}$, and in $\mathcal{H}$, the path can only contain members of Z at the bases of v-structures in $\mathcal{G}$. Because $\mathcal{H}$ is the moralized graph $\mathcal{M}[\mathcal{G}]$, those members of Z are no longer there. So there is an active path between X and Y in $\mathcal{H}$, so sep$_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$ also does not hold. Thus, we ca conclude $\neg$d-sep$(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) \Rightarrow \neg$ sep$_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) \iff$ d-sep$(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) \Rightarrow$ sep$_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$

Now we need to prove the converse of it. For that, we consider $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated given $\boldsymbol{Z}$. Consider an arbitrary path in $\mathcal{G}$ between some $X \in \boldsymbol{X}$ and some $Y \in \boldsymbol{Y}$. Any path between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\boldsymbol{\mathcal{G}}$ must either be blocked by a member of Z or an inactive v-structure in $\mathcal{G}$. First, suppose the path is blocked by a member of Z. Then the path in $\mathcal{H}$ will also be blocked by that member of Z. Of course, if the path between X and Y in $\mathcal{G}$ is not blocked by a member of Z, it must be blocked by an inactive v-structure. Because the v-structure is inactive, its base and any of its children must not be in the induced graph $\mathcal{H}$. As such, the path will not exist in $\mathcal{H}$. Neither the base nor its descendants can be in Z, and they cannot be in X or Y either, because then it would have an active path from a member of X to a member of Y. Again, the edges added in moralization are necessary to create an active path in $\mathcal{H}$ only when paths would be blocked by the observed root node of a v-structure in $\mathcal{G}$. In this case, the segment would have been active in $\mathcal{G}$, so moralization edges cannot effect segments in $\mathcal{H}$ corresponding to inactive segments in $\mathcal{G}$. Because of all the above, there are no active paths in $\mathcal{H}$ between arbitrary $X \in \boldsymbol{X}$ and $Y \in \boldsymbol{Y}$, so we have sep$_{\mathcal{H}}(\boldsymbol{x};\boldsymbol{Y}|\boldsymbol{Z})$. Thus, we have sep$_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z}) \Rightarrow$ d-sep$(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{Z})$

2.4 Solutions to Question 2.4

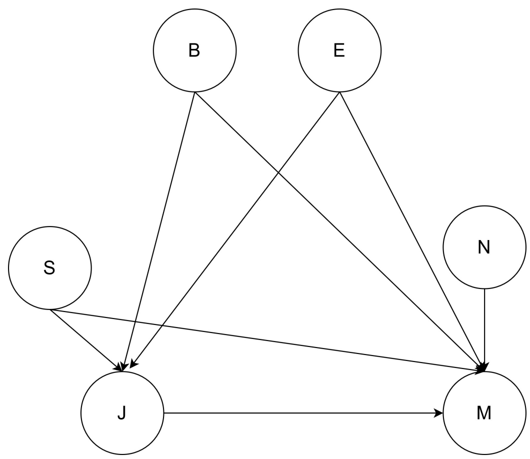

Figure 3: Bayesian Network for Burglary model without alarm

We follow the order `B` $\rightarrow$ `E` $\rightarrow$ `S` $\rightarrow$ `N` $\rightarrow$ `J` $\rightarrow$ `M`.

1. `Burglary`, `Earthquake`, `SportsOnTv`*and*`Naptime` are independent, so there is no edge between them

2. Since `Alarm` is unobserved, each of its parent `Burglary` and `Earthquake`, are connected to each of its descendents: `JohnCall` and `MaryCall`

3. `JohnCall` is an ancestor of `MaryCall` since we take it first in our ordering

4. If `JohnCall` is observed, there is no active path between `SportsOnTv` and `MaryCall`. Thus, there exists a dependency between `SportsOnTv` and `MaryCall`.

2.5 Solutions to Question 2.5

2.5.1 Part I:

To prove $(X \perp Y|Z) \implies (X \perp Y|Z,W)$, firstly, we note that the independence statements hold for all specific values of the variables X, Y, Z, and W.

So our goal is to prove that the above statement holds for all possible values of the variables.

Case I: *The probabilities of some values may be* 0*:* Let $P(W = w, Z = z) = 0, \implies (X \perp Y|Z, W)$. Again, if $P(Z = z) = 0 \implies P(W = w, Z = z) = 0$, so $(X \perp Y|Z, W)$ also holds.

Case II: $P(Z) \neq 0, P(Z, W) \neq 0$*:*

$$\begin{aligned} P(X, Y|Z, W) &= \frac{P(X, Y, W|Z)}{P(W|Z)} \quad [\text{ since } P(W|Z) \neq 0] \qquad (12) \\ &= \frac{P(X|Z) \cdot P(Y, W|Z)}{P(W|Z)} [\text{ since for a Markov Network,} (X \perp Y|Z) \iff (X \perp Y, W|Z)] \\ &= P(X|Z)P(Y|Z, W) \end{aligned}$$

We can also use the Decomposition rule $(X \perp Y, W|Z) \implies (X \perp W, Z)$ to deduce $(X \perp W, Z) \iff P(X|Z) = P(X|W, Z)$. Thus, we obtain

$$P(X, Y|Z, W) = P(X|Z, W)P(Y|Z, W)$$

which implies $(X \perp Y|Z, W)$

2.5.2 Part II:

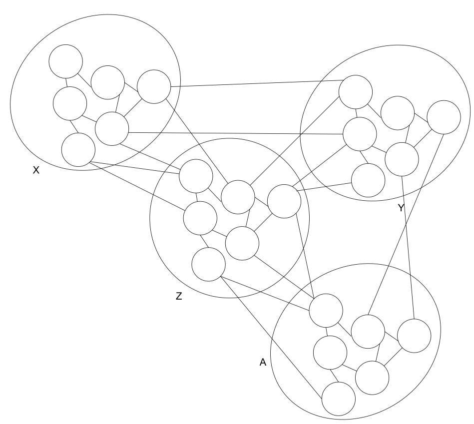

Figure 4: Markov Network Showing Transitivity Property

We need to prove that $\neg(\boldsymbol{X} \perp \boldsymbol{A}|\boldsymbol{Z}) \& \neg(\boldsymbol{A} \perp \boldsymbol{Y}|\boldsymbol{Z}) \implies \neg(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$. For $(\boldsymbol{X} \perp \boldsymbol{A}|\boldsymbol{Z})$, we see that there cannot exist any path from a node in $\boldsymbol{X}$ to the nodes of $\boldsymbol{A}$, without going through a node in $\boldsymbol{Z}$. Thus, for $\neg(\boldsymbol{X} \perp \boldsymbol{A}|\boldsymbol{Z})$ to be true, there exist a path between $\boldsymbol{X}$ and $\boldsymbol{A}$ without passing through $\boldsymbol{Z}$. Similarly, we can say for $\neg(\boldsymbol{A} \perp \boldsymbol{Y}|\boldsymbol{Z})$ to be true, there must exist some connection from $\boldsymbol{A} to \boldsymbol{Y}$ without going through $\boldsymbol{Z}$. Thus since both of them are true, we can find an active path from $\boldsymbol{X}$ to $\boldsymbol{Y}$ without going through $\boldsymbol{Z}$ as it is observed. Thus providing us with $\neg(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$. Hence the transitivity property is proved.

3 Question 3

3.1 Solutions to Question 3.1

$$
\begin{aligned}
P\left(X_i = 1 | \mathbf{X}_{-i}; \theta\right) &= \frac{P\left(X_i = 1, \mathbf{X}_{-i}\right)}{P\left(X_i = 1, \mathbf{X}_{-i}\right) + P\left(X_i = 0, \mathbf{X}_{-i}\right)} \\
&= \frac{\frac{P(X_i=1,\mathbf{X}_{-i})}{P(X_i=0,\mathbf{X}_{-i})}}{\frac{P(X_i=1,\mathbf{X}_{-i})}{P(X_i=0,\mathbf{X}_{-i})} + 1} \\
&= \frac{\exp\left(\sum_{s<i}\theta_{s,i}X_s + \sum_i \theta_i\right) - \left(-\sum_{s>i}\theta_{s,i}X_s - \sum_i \theta_i\right)}{\exp\left(\sum_{s<i}\theta_{s,i}X_s + \sum_i \theta_i\right) - \left(-\sum_{s>i}\theta_{s,i}X_s - \sum_i \theta_i\right) + 1} \\
&= \frac{\exp\left\{2 + 2\sum_{s\neq i}\theta_{s,i}X_s\right\}}{\exp\left\{2 + 2\sum_{s\neq i}\theta_{s,i}X_s\right\} + 1}
\end{aligned}
\tag{13}
$$

which is the form of the standard logistic regression.

3.2 Solutions to Question 3.2

The conditional distribution of the Ising conditional random field is given as

$$
\begin{aligned}
p(\mathbf{X}|\mathbf{W};\theta,\beta) &= \exp\left\{\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{s,t}X_sX_t + \sum_{s\in V, u\in[q]}\beta_{su}X_sW_u\right\} / Z(\mathbf{W},\theta,\beta) \\
Z(\mathbf{W},\theta,\beta) &= \sum_{\mathbf{X}}\exp\left\{\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V, u\in[q]}\beta_{su}X_sW_u\right\}
\end{aligned}
\tag{14}
$$

The conditional log-likelihood is

$$
\sum_{s\in V}\theta_s X_s + \sum_{(s,t)\in E}\theta_{st}X_sX_t + \sum_{s\in V, u\in[q]}\beta_{su}X_sW_u - \log Z(\mathbf{W},\theta,\beta)
$$

In order to minimize the log-likelihood using gradient descent, we compute the derivative of the log-likelihood on a dataset containing a single sample with respect to each of the parameters (θ_s, θ_{st}, and β_{su}). So, the derivatives are given as:

$$\frac{\partial}{\partial\theta_s}[\log p(\mathbf{X}|\mathbf{W};\theta,\beta)] = X_s - \frac{\sum_{\mathbf{X}\in\{X_s,X_t\}} X_s \exp\left\{\sum_s \theta_s X_s + \sum_{s,t}\theta_{st}X_sX_t + \sum_{s,u}\beta_{su}X_sW_u\right\}}{Z(\mathbf{W},\theta,\beta)}$$
$$= X_s - \sum_{\mathbf{X}} X_s P\left(\mathbf{X}|\mathbf{W};\theta,\beta\right) \tag{15}$$

$$\frac{\partial}{\partial\theta_{st}}[\log p(\mathbf{X}|\mathbf{W};\theta,\beta)] = X_sX_t - \frac{\sum_{\mathbf{X}} X_sX_t \exp\left\{\sum_s \theta_s X_s + \sum_{(s,t)}\theta_{st}X_sX_t + \sum_{s,u}\beta_{su}X_sW_u\right\}}{Z(\mathbf{W},\theta,\beta)}$$
$$= X_sX_t - \sum_{\mathbf{X}} X_sX_t P\left(\mathbf{X}|\mathbf{W};\theta,\beta\right) \tag{16}$$

$$\frac{\partial}{\partial\beta_{su}}[\log p(\mathbf{X}|\mathbf{W};\theta,\beta)] = W_uX_s - \frac{\sum_{\mathbf{X}} W_uX_s \exp\left\{\sum_s \theta_s X_s + \sum_{(s,t)}\theta_{st}X_sX_t + \sum_{s,u}\beta_{su}X_sW_u\right\}}{Z(\mathbf{W},\theta,\beta)}$$
$$= W_uX_s - W_u\sum_{\mathbf{X}} P\left(\mathbf{X}|\mathbf{W};\theta,\beta\right) \tag{17}$$

4 Question 4

To run the code, go to the Homework 1 folder and execute `./main.sh`
The results are stored in `Outputs`.
The codes are stored in `Codes`.

The general baseline model given in the question is given as follows: Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

Table 1: Baseline Model

	precision	recall	F1 score
GENE	0.158861	0.660436	0.256116

4.1 Part 1

In the first part, we use the counts method to change the infrequent words (whose count is less than 5) to *RARE*. This will help us to predict emission probabilities for words in the test data that do not occur in the training. We then map infrequent words in the training data to a common class and to treat unseen words as members of this class. data. Evaluating Baseline Decoding Found 2669 GENEs. Expected 642 GENEs; Correct: 424.

Table 2: Baseline Model with Rare and Emission Parameters

	precision	recall	F1 score
GENE	0.158861	0.660436	0.256116

4.2 Part 2

Running HMM with Trigram Features Evaluating HMM with Trigram Features Found 191 GENEs. Expected 642 GENEs; Correct: 104.

Table 3: Hidden Markov Model with Trigram features

	precision	recall	F1 score
GENE	0.544503	0.161994	0.249700

5 Question 5

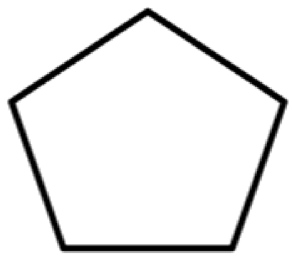

Figure 5: Markov Network

For the given Markov Network, we define the nodes x_1, x_2, x_3, x_4, x_5 counting clockwise around the pentagon with potentials $\phi(x_i, x_j)$ We need to show that the joint distribution can be written as

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

Now,

$$\begin{aligned}
p(x_1, x_2, x_5) &= \frac{1}{Z} \sum_{3,4} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51} && (18)\\
p(x_2, x_4, x_5) &= \frac{1}{Z} \sum_{1,3} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\\
p(x_2, x_3, x_4) &= \frac{1}{Z} \sum_{1,5} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\\
p(x_2, x_5) &= \frac{1}{Z} \sum_{1,3,4} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\\
p(x_2, x_4) &= \frac{1}{Z} \sum_{1,3,5} \phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
RHS &= \frac{p\left(x_1,x_2,x_5\right)p\left(x_2,x_4,x_5\right)p\left(x_2,x_3,x_4\right)}{p\left(x_2,x_5\right)p\left(x_2,x_4\right)} \qquad (19)\\
&= \frac{\left(\frac{1}{Z}\sum_{3,4}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)\left(\frac{1}{Z}\sum_{1,3}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)\left(\frac{1}{Z}\sum_{1,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}{\left(\frac{1}{Z}\sum_{1,3,4}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)\left(\frac{1}{Z}\sum_{1,3,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}\\
&= \frac{1}{Z}\frac{\left(\phi_{12}\phi_{51}\sum_{3,4}\phi_{23}\phi_{34}\phi_{45}\right)\left(\phi_{45}\phi_{51}\sum_{1,3}\phi_{12}\phi_{23}\phi_{34}\right)\left(\phi_{23}\phi_{34}\sum_{1,5}\phi_{12}\phi_{45}\phi_{51}\right)}{\left(\sum_{1,3,4,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}\\
&= \frac{1}{Z}\frac{\left(\phi_{12}\phi_{51}\phi_{45}\phi_{51}\phi_{23}\phi_{34}\right)\left(\sum_{3,4}\phi_{23}\phi_{34}\phi_{45}\right)\left(\sum_{1,3}\phi_{12}\phi_{23}\phi_{34}\right)\left(\sum_{1,5}\phi_{12}\phi_{45}\phi_{51}\right)}{\left(\sum_{1,3,4,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}\\
&= \frac{1}{Z}\frac{\left(\phi_{12}\phi_{51}\phi_{45}\phi_{51}\phi_{23}\phi_{34}\right)\left(\sum_{1,3,4,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}{\left(\sum_{1,3,4,5}\phi_{12}\phi_{23}\phi_{34}\phi_{45}\phi_{51}\right)}\\
&= \frac{1}{Z}\left(\phi_{12}\phi_{51}\phi_{45}\phi_{51}\phi_{23}\phi_{34}\right)\\
&= p\left(x_1,x_2,x_3,x_4,x_5\right)\\
&= LHS
\end{aligned}
$$