CSCI 567, Spring'18
Michael Shindler

Problem Set #5

Due: 11:59 pm, April 25, 2018

Instructions

Submission: Assignment submission will be via courses.uscden.net. By the submission date, there will be a folder set up in which you can submit your files. Please be sure to follow all directions outlined here.

You can submit multiple files, but only the last one submitted counts. That means if you finish some problems and want to submit something now, and then a later file when you finish, that's fine. If I were taking the class, that's what I'd do: that way, if I forget to finish the homework or something happens (remember Murphy's Law), I still get credit for what I finished and turned in. Remember, there are no grace days on problem sets, just on programming assignments!

Problem sets must be typewritten or neatly handwritten when submitted; if the grader cannot read your handwriting on a problem, they may elect to grade it as a zero. If it is handwritten, your submission must still be submitted as a PDF.

It is strongly recommended that you typeset with LaTeX and use that to generate a PDF file.

- The file should be named as `firstname_lastname_USCID.pdf` (e.g., Jenny_Tutone_8675309.pdf).
- Do not have any spaces in your file name when uploading it.
- Please include your name and USCID in the header of the report as well.

There are many free integrated LaTeX editors that are convenient to use. Choose the one(s) you like the most. This http://www.andy-roberts.net/writing/latex seems like a good tutorial.

Collaboration: You may discuss with your classmates. However, you need to write your own solutions and submit separately. Also in your report, you need to list with whom you have discussed for each problem. Please consult the syllabus for what is and is not acceptable collaboration. Review the rules on academic conduct in the syllabus: a single instance of plagiarism can adversely affect you significantly more than you could stand to gain.

Notes on notation:

- Unless stated otherwise, scalars are denoted by small letter in normal font, vectors are denoted by small letters in bold font and matrices are denoted by capital letters in bold font.
- The bias term is subsumed in the input vector, so the input vector is actually $x = [x', 1]^T$, unless mentioned otherwise.
- $\|.\|$ means L2-norm unless specified otherwise i.e. $\|.\| = \|.\|_2$

Problem 1 Principal Component Analysis (10 points)

For the following problem, we have N *zero-mean* data points $\mathbf{x}_i \in \mathbb{R}^{D\times 1}$ and $\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^T \in \mathbb{R}^{D\times D}$ is the sample covariance matrix of the dataset.

In a study, a simple random sample of 100 bird species is collected. Three factors were measured: length (inches), wingspan (inches), and weight (ounces). Thus N = 100, D = 3. The covariance matrix **S** is calculated:

$$S = \begin{bmatrix} 91.43 & 171.92 & 297.99 \\ & 373.92 & 545.21 \\ & & 1297.26 \end{bmatrix} \tag{1}$$

The entries below the diagonal were omitted, since the matrix is symmetric. Also, **S** was computed without dividing by $N-1$ (also a common practice).

1.1 Compute the eigenvalues and orthonormal eigenvectors.

The eigenvalues are as follows:

$$\lambda_1 = 1626.526, \qquad \lambda_2 = 128.986, \qquad \lambda_3 = 7.097 \tag{2}$$

The orthonormal eigenvectors are as follows:

$$\mathbf{v}_1 = \begin{bmatrix} 0.218 \\ 0.414 \\ 0.884 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 0.247 \\ 0.853 \\ -0.461 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 0.944 \\ -0.318 \\ -0.084 \end{bmatrix} \tag{3}$$

1.2 Is there any of the orthonormal directions that can be omitted without losing lot of information? If yes which one(s) and why? *You may explain in your own words*

The first principal component accounts for $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = 0.923$; the second principal component accounts for $\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = 0.073$; the second principal component accounts for $\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} = 0.004$.

Now, we can see that compared with first two principal components, the third principal component is negligible, and the first principal component contains the most of information regarding this data. So, the last orthonormal direction can be omitted without losing lot of information. However, if we only want to keep one orthonormal direction, we can just keep the first orthonormal direction and the last two orthonormal directions can be omitted without losing lot of information.

If we only want to keep one orthonormal direction, we can just keep the first orthonormal direction and the last two orthonormal directions can be omitted without losing lot of information.

1.3 How do you interpret the eigenvector(s) that contain(s) the most of information regarding this data? (in term of three factors measured) *You may explain in your own words*

All three entries in $\mathbf{v}_1$ represent how much this orthonormal eigenvector "points to" three directions including "length", "wingspan" and "weight. Since the third entry is the largest entry in $\mathbf{v}_1$, we can say that "weight" is the most important factor for this data, "wingspan" (0.414) is the second important factor for this data and "length" (0.218) is the third important factor for this data.

Problem 2 Hidden Markov Model (20 points)

In this problem, we want to fit DNA sequence data with a generative model. In particular, we assume that they are generated by a hidden Markov model (HMM). Let $\mathbf{X}_{1:N} = [X_1 X_2 \dots X_N]$ be random variables corresponding to a DNA sequence of length N, controlled by hidden states $\mathbf{Z}_{1:N} = [Z_1 Z_2 \dots Z_N]$. Each X_n takes a value in $\{A, C, G, T\}$ and each Z_n takes one of the two possible states $\{s_1, s_2\}$. This HMM has the following parameters $\mathbf{\Theta} = \{\pi_i, a_{ij}, b_{ik}\}$ for $i \in \{1, 2\}$, $j \in \{1, 2\}$, and $k \in \{A, C, G, T\}$:

- Initial state distribution π_i for $i = 1, 2$:

$$\pi_1 = P(Z_1 = s_1) = 0.7; \quad \pi_2 = P(Z_1 = s_2) = 0.3. \tag{4}$$

- Transition probabilities $a_{ij} = P(Z_{n+1} = s_j | Z_n = s_i)$ for any $n \in \mathbb{N}^+$, $i = 1, 2$ and $j = 1, 2$:

$$a_{11} = 0.8, \quad a_{12} = 0.2, \quad a_{21} = 0.4, \quad a_{22} = 0.6. \tag{5}$$

- Emission probabilities $b_{ik} = P(X_n = k | Z_n = s_i)$ for any $n \in \mathbb{N}^+$, $i = 1, 2$ and $k \in \{A, C, G, T\}$:

$$b_{1A} = 0.4, \quad b_{1C} = 0.1, \quad b_{1G} = 0.4, \quad b_{1T} = 0.1; \tag{6}$$
$$b_{2A} = 0.2, \quad b_{2C} = 0.3, \quad b_{2G} = 0.2, \quad b_{2T} = 0.3. \tag{7}$$

We observe a sequence $\mathbf{O}_{1:6} = [o_1 o_2 \dots o_6] = [AGCGTA]$, please answer the following questions with step-by-step computations.

2.1 Probability of an observed sequence Compute $P(\mathbf{X}_{1:6} = \mathbf{O}_{1:6}; \mathbf{\Theta})$.

$$\begin{aligned}\alpha(Z_1 = s_1) &= P(X_1 = A, Z_1 = s_1) = P(X_1 = A | Z_1 = s_1) P(Z_1 = s_1) \\ &= b_{1A} \cdot \pi_1 = 0.4 \cdot 0.7 = 0.28\end{aligned}$$

$$\begin{aligned}\alpha(Z_1 = s_2) &= P(X_1 = A, Z_1 = s_2) = P(X_1 = A | Z_1 = s_2) P(Z_1 = s_2) \\ &= b_{2A} \cdot \pi_2 = 0.2 \cdot 0.3 = 0.06\end{aligned}$$

$$\begin{aligned}\alpha(Z_2 = s_1) &= P(X_2 = G, Z_2 = s_1) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2 | Z_1) = b_{1G} \cdot [\alpha(Z_1 = s_1) \cdot a_{11} + \alpha(Z_1 = s_2) \cdot a_{21}] \\ &= 0.4 \cdot [0.28 \cdot 0.8 + 0.06 \cdot 0.4] = 0.0992\end{aligned}$$

$$\begin{aligned}\alpha(Z_2 = s_2) &= P(X_2 = G, Z_2 = s_2) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2 | Z_1) = b_{2G} \cdot [\alpha(Z_1 = s_1) \cdot a_{12} + \alpha(Z_1 = s_2) \cdot a_{22}] \\ &= 0.2 \cdot [0.28 \cdot 0.2 + 0.06 \cdot 0.6] = 0.0184\end{aligned}$$

$$\begin{aligned}\alpha(Z_3 = s_1) &= P(X_3 = C, Z_3 = s_1) \cdot \sum_{Z_2} \alpha(Z_2) P(Z_3 | Z_2) = b_{1C} \cdot [\alpha(Z_2 = s_1) \cdot a_{11} + \alpha(Z_2 = s_2) \cdot a_{21}] \\ &= 0.1 \cdot [0.0992 \cdot 0.8 + 0.0184 \cdot 0.4] = 8.672 \times 10^{-3}\end{aligned}$$

$$\begin{aligned}\alpha(Z_3 = s_2) &= P(X_3 = C, Z_3 = s_2) \cdot \sum_{Z_2} \alpha(Z_2) P(Z_3 | Z_2) = b_{2C} \cdot [\alpha(Z_2 = s_1) \cdot a_{12} + \alpha(Z_2 = s_2) \cdot a_{22}] \\ &= 0.3 \cdot [0.0992 \cdot 0.2 + 0.0184 \cdot 0.6] = 9.364 \times 10^{-3}\end{aligned}$$

$$\begin{aligned}\alpha(Z_4 = s_1) &= P(X_2 = G, Z_2 = s_1) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{1G} \cdot [\alpha(Z_1 = s_1) \cdot a_{11} + \alpha(Z_1 = s_2) \cdot a_{21}] \\ &= 0.4 \cdot [8.672 \times 10^{-3} \cdot 0.8 + 9.364 \times 10^{-3} \cdot 0.4] = 4.257 \times 10^{-3}\end{aligned}$$

$$\begin{aligned}\alpha(Z_4 = s_2) &= P(X_2 = G, Z_2 = s_2) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{2G} \cdot [\alpha(Z_1 = s_1) \cdot a_{12} + \alpha(Z_1 = s_2) \cdot a_{22}] \\ &= 0.2 \cdot [8.672 \times 10^{-3} \cdot 0.2 + 9.364 \times 10^{-3} \cdot 0.6] = 1.459 \times 10^{-3}\end{aligned}$$

$$\begin{aligned}\alpha(Z_5 = s_1) &= P(X_2 = G, Z_2 = s_1) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{1G} \cdot [\alpha(Z_1 = s_1) \cdot a_{11} + \alpha(Z_1 = s_2) \cdot a_{21}] \\ &= 0.1 \cdot [4.257 \times 10^{-3} \cdot 0.8 + 1.459 \times 10^{-3} \cdot 0.4] = 3.9892 \times 10^{-4}\end{aligned}$$

$$\begin{aligned}\alpha(Z_5 = s_2) &= P(X_2 = G, Z_2 = s_2) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{2G} \cdot [\alpha(Z_1 = s_1) \cdot a_{12} + \alpha(Z_1 = s_2) \cdot a_{22}] \\ &= 0.3 \cdot [4.257 \times 10^{-3} \cdot 0.2 + 1.459 \times 10^{-3} \cdot 0.6] = 5.1798 \times 10^{-4}\end{aligned}$$

$$\begin{aligned}\alpha(Z_6 = s_1) &= P(X_2 = G, Z_2 = s_1) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{1G} \cdot [\alpha(Z_1 = s_1) \cdot a_{11} + \alpha(Z_1 = s_2) \cdot a_{21}] \\ &= 0.4 \cdot [3.9892 \times 10^{-4} \cdot 0.8 + 5.1798 \times 10^{-4} \cdot 0.4] = 2.1053 \times 10^{-4}\end{aligned}$$

$$\begin{aligned}\alpha(Z_6 = s_2) &= P(X_2 = G, Z_2 = s_2) \cdot \sum_{Z_1} \alpha(Z_1) P(Z_2|Z_1) = b_{2G} \cdot [\alpha(Z_1 = s_1) \cdot a_{12} + \alpha(Z_1 = s_2) \cdot a_{22}] \\ &= 0.2 \cdot [3.9892 \times 10^{-4} \cdot 0.2 + 5.1798 \times 10^{-4} \cdot 0.6] = 7.8114 \times 10^{-5}\end{aligned}$$

$$\begin{aligned}P(\mathbf{X}; \Theta) &= \sum_{\mathbf{Z}_6} \alpha(\mathbf{Z}_6) = \alpha(Z_6 = s_1) + \alpha(Z_6 = s_2) \\ &= 2.1053 \times 10^{-4} + 7.8114 \times 10^{-5} = 2.8865 \times 10^{-4}\end{aligned}$$

$$\log P(\mathbf{X}; \Theta) = -8.1503 \qquad \square$$

2.2 Most likely explanation

Compute $\mathbf{z}^*{}_{1:6} = [z_1^* z_2^* \ldots z_6^*] = \arg\max_{\mathbf{z}_{1:6}} P(\mathbf{Z}_{1:6} = \mathbf{z}_{1:6} | \mathbf{X}_{1:6} = \mathbf{O}_{1:6}; \boldsymbol{\Theta})$.

$$
\begin{aligned}
\log(P(X_1, Z_1 = s_1)) &= \log(P(X_1 = A|Z_1 = s_1)P(Z_1 = s_1)) \\
&= \log(b_{1A}) + \log(\pi_1) = \log(0.4) + \log(0.7) = -1.2730 \\
\log(P(X_1, Z_1 = s_2)) &= \log(P(X_1 = A|Z_1 = s_2)P(Z_1 = s_2)) \\
&= \log(b_{2A}) + \log(\pi_2) = \log(0.2) + \log(0.3) = -2.8134
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_2, X_1, \mathbf{Z_2} = \mathbf{s_1}, \mathbf{Z_1} = \mathbf{s_1})) &= \log(P(X_2, Z_2 = s_1|Z_1 = s_1)P(X_1, Z_1 = s_1)) \\
&= \log(P(X_2 = G|Z_2 = s_1)P(Z_2 = s_1|Z_1 = s_1)P(X_1, Z_1 = s_1)) \\
&= \log(b_{1G}) + \log(a_{11}) + \log(P(X_1, Z_1 = s_1)) \\
&= \log(0.4) + \log(0.8) - 1.2730 = -\mathbf{2.4124} \\
\log(P(X_2, X_1, Z_2 = s_1, Z_1 = s_2)) &= \log(P(X_2, Z_2 = s_1|Z_1 = s_2)P(X_1, Z_1 = s_2)) \\
&= \log(P(X_2 = G|Z_2 = s_1)P(Z_2 = s_1|Z_1 = s_2)P(X_1, Z_1 = s_2)) \\
&= \log(b_{1G}) + \log(a_{21}) + \log(P(X_1, Z_1 = s_2)) \\
&= \log(0.4) + \log(0.4) - 2.8134 = -4.6460
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_2, X_1, \mathbf{Z_2} = \mathbf{s_2}, \mathbf{Z_1} = \mathbf{s_1})) &= \log(P(X_2, Z_2 = s_2|Z_1 = s_1)P(X_1, Z_1 = s_1)) \\
&= \log(P(X_2 = G|Z_2 = s_2)P(Z_2 = s_2|Z_1 = s_1)P(X_1, Z_1 = s_1)) \\
&= \log(b_{2G}) + \log(a_{12}) + \log(P(X_1, Z_1 = s_1)) \\
&= \log(0.2) + \log(0.2) - 1.2730 = -\mathbf{4.4918} \\
\log(P(X_2, X_1, Z_2 = s_2, Z_1 = s_2)) &= \log(P(X_2, Z_2 = s_2|Z_1 = s_2)P(X_1, Z_1 = s_2)) \\
&= \log(P(X_2 = G|Z_2 = s_2)P(Z_2 = s_2|Z_1 = s_2)P(X_1, Z_1 = s_2)) \\
&= \log(b_{2G}) + \log(a_{22}) + \log(P(X_1, Z_1 = s_2)) \\
&= \log(0.2) + \log(0.6) - 2.8134 = -4.9337
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_3, X_2, \mathbf{Z_3} = \mathbf{s_1}, \mathbf{Z_2} = \mathbf{s_1})) &= \log(P(X_3, Z_3 = s_1 | Z_2 = s_1) P(X_2, Z_2 = s_1)) \\
&= \log(P(X_3 = C | Z_3 = s_1) P(Z_3 = s_1 | Z_2 = s_1) P(X_2, Z_2 = s_1)) \\
&= \log(b_{1C}) + \log(a_{11}) + \log(P(X_2, Z_2 = s_1)) \\
&= \log(0.1) + \log(0.8) - 2.4124 = -\mathbf{4.9381} \\
\log(P(X_3, X_2, Z_3 = s_1, Z_2 = s_2)) &= \log(P(X_3, Z_3 = s_1 | Z_2 = s_2) P(X_2, Z_2 = s_2)) \\
&= \log(P(X_3 = C | Z_3 = s_1) P(Z_3 = s_1 | Z_2 = s_2) P(X_2, Z_2 = s_2)) \\
&= \log(b_{1C}) + \log(a_{21}) + \log(P(X_2, Z_2 = s_2)) \\
&= \log(0.1) + \log(0.4) - 4.4918 = -7.7107 \\
\log(P(X_3, X_2, \mathbf{Z_3} = \mathbf{s_2}, \mathbf{Z_2} = \mathbf{s_1})) &= \log(P(X_3, Z_3 = s_2 | Z_2 = s_1) P(X_2, Z_2 = s_1)) \\
&= \log(P(X_3 = C | Z_3 = s_2) P(Z_3 = s_2 | Z_2 = s_1) P(X_2, Z_2 = s_1)) \\
&= \log(b_{2C}) + \log(a_{12}) + \log(P(X_2, Z_2 = s_1)) \\
&= \log(0.3) + \log(0.2) - 2.4124 = -\mathbf{5.2258} \\
\log(P(X_3, X_2, Z_3 = s_2, Z_2 = s_2)) &= \log(P(X_3, Z_3 = s_2 | Z_2 = s_2) P(X_2, Z_2 = s_2)) \\
&= \log(P(X_3 = C | Z_3 = s_2) P(Z_3 = s_2 | Z_2 = s_2) P(X_2, Z_2 = s_2)) \\
&= \log(b_{2C}) + \log(a_{22}) + \log(P(X_2, Z_2 = s_2)) \\
&= \log(0.3) + \log(0.6) - 4.4918 = -6.2066
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_4, X_3, \mathbf{Z_4} = \mathbf{s_1}, \mathbf{Z_3} = \mathbf{s_1})) &= \log(P(X_4, Z_4 = s_1 | Z_3 = s_1) P(X_3, Z_3 = s_1)) \\
&= \log(P(X_4 = G | Z_4 = s_1) P(Z_4 = s_1 | Z_3 = s_1) P(X_3, Z_3 = s_1)) \\
&= \log(b_{1G}) + \log(a_{11}) + \log(P(X_3, Z_3 = s_1)) \\
&= \log(0.4) + \log(0.8) - 4.9381 = -\mathbf{6.0776} \\
\log(P(X_4, X_3, Z_4 = s_1, Z_3 = s_2)) &= \log(P(X_4, Z_4 = s_1 | Z_3 = s_2) P(X_3, Z_3 = s_2)) \\
&= \log(P(X_4 = G | Z_4 = s_1) P(Z_4 = s_1 | Z_3 = s_2) P(X_3, Z_3 = s_2)) \\
&= \log(b_{1G}) + \log(a_{21}) + \log(P(X_3, Z_3 = s_2)) \\
&= \log(0.4) + \log(0.4) - 5.2258 = -7.0584 \\
\log(P(X_4, X_3, Z_4 = s_2, Z_3 = s_1)) &= \log(P(X_4, Z_4 = s_2 | Z_3 = s_1) P(X_3, Z_3 = s_1)) \\
&= \log(P(X_4 = G | Z_4 = s_2) P(Z_4 = s_2 | Z_3 = s_1) P(X_3, Z_3 = s_1)) \\
&= \log(b_{2G}) + \log(a_{12}) + \log(P(X_3, Z_3 = s_1)) \\
&= \log(0.2) + \log(0.2) - 4.9381 = -8.1570 \\
\log(P(X_4, X_3, \mathbf{Z_4} = \mathbf{s_2}, \mathbf{Z_3} = \mathbf{s_2})) &= \log(P(X_4, Z_4 = s_2 | Z_3 = s_2) P(X_3, Z_3 = s_2)) \\
&= \log(P(X_4 = G | Z_4 = s_2) P(Z_4 = s_2 | Z_3 = s_2) P(X_3, Z_3 = s_2)) \\
&= \log(b_{2G}) + \log(a_{22}) + \log(P(X_3, Z_3 = s_2)) \\
&= \log(0.2) + \log(0.6) - 5.2258 = -\mathbf{7.3461}
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_5, X_4, \mathbf{Z_5} = \mathbf{s_1}, \mathbf{Z_4} = \mathbf{s_1})) &= \log(P(X_5, Z_5 = s_1 | Z_4 = s_1) P(X_4, Z_4 = s_1)) \\
&= \log(P(X_5 = T | Z_5 = s_1) P(Z_5 = s_1 | Z_4 = s_1) P(X_4, Z_4 = s_1)) \\
&= \log(b_{1T}) + \log(a_{11}) + \log(P(X_4, Z_4 = s_1)) \\
&= \log(0.1) + \log(0.8) - 6.0776 = -\mathbf{8.6033} \\
\log(P(X_5, X_4, Z_5 = s_1, Z_4 = s_2)) &= \log(P(X_5, Z_5 = s_1 | Z_4 = s_2) P(X_4, Z_4 = s_2)) \\
&= \log(P(X_5 = T | Z_5 = s_1) P(Z_5 = s_1 | Z_4 = s_2) P(X_4, Z_4 = s_2)) \\
&= \log(b_{1T}) + \log(a_{21}) + \log(P(X_4, Z_4 = s_2)) \\
&= \log(0.1) + \log(0.4) - 7.3461 = -10.5650 \\
\log(P(X_5, X_4, \mathbf{Z_5} = \mathbf{s_2}, \mathbf{Z_4} = \mathbf{s_1})) &= \log(P(X_5, Z_5 = s_2 | Z_4 = s_1) P(X_4, Z_4 = s_1)) \\
&= \log(P(X_5 = T | Z_5 = s_2) P(Z_5 = s_2 | Z_4 = s_1) P(X_4, Z_4 = s_1)) \\
&= \log(b_{2T}) + \log(a_{12}) + \log(P(X_4, Z_4 = s_1)) \\
&= \log(0.3) + \log(0.2) - 6.0776 = -\mathbf{8.8910} \\
\log(P(X_5, X_4, Z_5 = s_2, Z_4 = s_2)) &= \log(P(X_5, Z_5 = s_2 | Z_4 = s_2) P(X_4, Z_4 = s_2)) \\
&= \log(P(X_5 = T | Z_5 = s_2) P(Z_5 = s_2 | Z_4 = s_2) P(X_4, Z_4 = s_2)) \\
&= \log(b_{2T}) + \log(a_{22}) + \log(P(X_4, Z_4 = s_2)) \\
&= \log(0.3) + \log(0.6) - 7.3461 = -9.0609
\end{aligned}
$$

$$
\begin{aligned}
\log(P(X_6, X_5, \mathbf{Z_6} = \mathbf{s_1}, \mathbf{Z_5} = \mathbf{s_1})) &= \log(P(X_6, Z_6 = s_1 | Z_5 = s_1) P(X_5, Z_5 = s_1)) \\
&= \log(P(X_6 = A | Z_6 = s_1) P(Z_6 = s_1 | Z_5 = s_1) P(X_5, Z_5 = s_1)) \\
&= \log(b_{1A}) + \log(a_{11}) + \log(P(X_5, Z_5 = s_1)) \\
&= \log(0.4) + \log(0.8) - 8.6033 = -\mathbf{9.7427} \\
\log(P(X_6, X_5, Z_6 = s_1, Z_5 = s_2)) &= \log(P(X_6, Z_6 = s_1 | Z_5 = s_2) P(X_5, Z_5 = s_2)) \\
&= \log(P(X_6 = A | Z_6 = s_1) P(Z_6 = s_1 | Z_5 = s_2) P(X_5, Z_5 = s_2)) \\
&= \log(b_{1A}) + \log(a_{21}) + \log(P(X_5, Z_5 = s_2)) \\
&= \log(0.4) + \log(0.4) - 8.8910 = -10.7236 \\
\log(P(X_6, X_5, Z_6 = s_2, Z_5 = s_1)) &= \log(P(X_6, Z_6 = s_2 | Z_5 = s_1) P(X_5, Z_5 = s_1)) \\
&= \log(P(X_6 = A | Z_6 = s_2) P(Z_6 = s_2 | Z_5 = s_1) P(X_5, Z_5 = s_1)) \\
&= \log(b_{2A}) + \log(a_{12}) + \log(P(X_5, Z_5 = s_1)) \\
&= \log(0.2) + \log(0.2) - 8.6033 = -11.8222 \\
\log(P(X_6, X_5, \mathbf{Z_6} = \mathbf{s_2}, \mathbf{Z_5} = \mathbf{s_2})) &= \log(P(X_6, Z_6 = s_2 | Z_5 = s_2) P(X_5, Z_5 = s_2)) \\
&= \log(P(X_6 = A | Z_6 = s_2) P(Z_6 = s_2 | Z_5 = s_2) P(X_5, Z_5 = s_2)) \\
&= \log(b_{2A}) + \log(a_{22}) + \log(P(X_5, Z_5 = s_2)) \\
&= \log(0.2) + \log(0.6) - 8.8910 = -\mathbf{11.0112}
\end{aligned}
$$

so the most likely hidden states are $\mathbf{Z} = s_1 \to s_1 \to s_1 \to s_1 \to s_1 \to s_1$ □

2.3 Prediction Compute $x^* = \arg\max_x P(X_7 = x | \mathbf{X}_{1:6} = \mathbf{O}_{1:6}; \mathbf{\Theta})$.

$$
\begin{aligned}
P(X_7|\mathbf{X};\Theta) &= \sum_{Z_7} P(X_7, Z_7|\mathbf{X};\Theta) \\
&= \sum_{Z_7} P(X_7|Z_7)P(Z_7|\mathbf{X};\Theta) \\
&= \sum_{Z_7} P(X_7|Z_7) \sum_{Z_6} P(Z_7, Z_6|\mathbf{X};\Theta) \\
&= \sum_{Z_7} P(X_7|Z_7) \sum_{Z_6} P(Z_7|Z_6)P(Z_6|\mathbf{X};\Theta) \\
&= \frac{1}{P(\mathbf{X};\Theta)} \sum_{Z_7} P(X_7|Z_7) \sum_{Z_6} P(Z_7|Z_6)\alpha(Z_6) \\
&= \frac{1}{P(\mathbf{X};\Theta)} [P(X_7|Z_7 = s_1) \cdot (P(Z_7 = s_1|Z_6 = s_1)\alpha(Z_6 = s_1) + P(Z_7 = s_1|Z_6 = s_2)\alpha(Z_6 = s_2)) \\
&\quad + P(X_7|Z_7 = s_2) \cdot (P(Z_7 = s_2|Z_6 = s_1)\alpha(Z_6 = s_1) + P(Z_7 = s_2|Z_6 = s_2)\alpha(Z_6 = s_2))] \\
&= \frac{1}{P(\mathbf{X};\Theta)} [P(X_7|Z_7 = s_1) \cdot (a_{11} \cdot \alpha(Z_6 = s_1) + a_{21} \cdot \alpha(Z_6 = s_2)) \\
&\quad + P(X_7|Z_7 = s_2) \cdot (a_{12} \cdot \alpha(Z_6 = s_1) + a_{22} \cdot \alpha(Z_6 = s_2))]
\end{aligned}
$$

$$
\begin{aligned}
P(X_7 = A|\mathbf{X};\Theta) &= \frac{1}{P(\mathbf{X};\Theta)} [b_{1A} \cdot (a_{11} \cdot \alpha(Z_6 = s_1) + a_{21} \cdot \alpha(Z_6 = s_2)) \\
&\quad + b_{2A} \cdot (a_{12} \cdot \alpha(Z_6 = s_1) + a_{22} \cdot \alpha(Z_6 = s_2))] \\
&= \frac{0.4 \cdot (0.8 \cdot 2.1053 \times 10^{-4} + 0.4 \cdot 7.8114 \times 10^{-5}) + 0.2 \cdot (0.2 \cdot 2.1053 \times 10^{-4} + 0.6 \cdot 7.8114 \times 10^{-5})}{2.8865 \times 10^{-4}} \\
&= 0.3383
\end{aligned}
$$

$$
\begin{aligned}
P(X_7 = C|\mathbf{X};\Theta) &= \frac{1}{P(\mathbf{X};\Theta)} [b_{1C} \cdot (a_{11} \cdot \alpha(Z_6 = s_1) + a_{21} \cdot \alpha(Z_6 = s_2)) \\
&\quad + b_{2C} \cdot (a_{12} \cdot \alpha(Z_6 = s_1) + a_{22} \cdot \alpha(Z_6 = s_2))] \\
&= \frac{0.1 \cdot (0.8 \cdot 2.1053 \times 10^{-4} + 0.4 \cdot 7.8114 \times 10^{-5}) + 0.3 \cdot (0.2 \cdot 2.1053 \times 10^{-4} + 0.6 \cdot 7.8114 \times 10^{-5})}{2.8865 \times 10^{-4}} \\
&= 0.1617
\end{aligned}
$$

$$
\begin{aligned}
P(X_7 = G|\mathbf{X};\Theta) &= \frac{1}{P(\mathbf{X};\Theta)} [b_{1G} \cdot (a_{11} \cdot \alpha(Z_6 = s_1) + a_{21} \cdot \alpha(Z_6 = s_2)) \\
&\quad + b_{2G} \cdot (a_{12} \cdot \alpha(Z_6 = s_1) + a_{22} \cdot \alpha(Z_6 = s_2))] \\
&= \frac{0.4 \cdot (0.8 \cdot 2.1053 \times 10^{-4} + 0.4 \cdot 7.8114 \times 10^{-5}) + 0.2 \cdot (0.2 \cdot 2.1053 \times 10^{-4} + 0.6 \cdot 7.8114 \times 10^{-5})}{2.8865 \times 10^{-4}} \\
&= 0.3383
\end{aligned}
$$

$$
\begin{aligned}
P(X_7 = T|\mathbf{X};\Theta) &= \frac{1}{P(\mathbf{X};\Theta)}[b_{1T} \cdot (a_{11} \cdot \alpha(Z_6 = s_1) + a_{21} \cdot \alpha(Z_6 = s_2)) \\
&\quad + b_{2T} \cdot (a_{12} \cdot \alpha(Z_6 = s_1) + a_{22} \cdot \alpha(Z_6 = s_2))] \\
&= \frac{0.1 \cdot (0.8 \cdot 2.1053 \times 10^{-4} + 0.4 \cdot 7.8114 \times 10^{-5}) + 0.3 \cdot (0.2 \cdot 2.1053 \times 10^{-4} + 0.6 \cdot 7.8114 \times 10^{-5})}{2.8865 \times 10^{-4}} \\
&= 0.1617
\end{aligned}
$$

so $\quad X_7 = A$ or $G \qquad \square$

Problem 3 Probabilistic graphical models (PGM) (10 points)

In this problem, we will practice with the key concept—(conditional) independence—used in probabilistic graphical models (PGM).

3.1 Conditional independence Given three discrete random variables $X_1 \in \{-1, 1\}$ and $X_2 \in \{-1, 1\}$ and $X_3 \in \{-1, 1\}$ and their joint probability in Table 1, please verify if $X_1 \perp\!\!\!\perp X_2 | X_3$ or not, together with no more than 5 lines of derivation or explanation.

Table 1: The joint probability $P(X_1, X_2, X_3)$.

X_1, X_2 / X_3	-1	1
-1, -1	0.1	0.01
1, -1	0.05	0.1
-1, 1	0.2	0.04
1, 1	0.1	0.4

$P(X_1 = -1 | X_3 = -1) = \frac{2}{3} = P(X_1 = -1 | X_3 = -1, X_2 = 1) = P(X_1 = -1 | X_3 = -1, X_2 = -1)$
$P(X_1 = -1 | X_3 = 1) = \frac{1}{11} = P(X_1 = -1 | X_3 = 1, X_2 = 1) = P(X_1 = -1 | X_3 = 1, X_2 = -1)$
$P(X_2 = -1 | X_3 = -1) = \frac{1}{3} = P(X_2 = -1 | X_3 = -1, X_1 = 1) = P(X_2 = -1 | X_3 = -1, X_1 = -1)$
$P(X_2 = -1 | X_3 = 1) = \frac{1}{5} = P(X_2 = -1 | X_3 = 1, X_1 = 1) = P(X_2 = -1 | X_3 = 1, X_1 = -1)$
Therefore, $X_1 \perp\!\!\!\perp X_2 | X_3$.

3.2 Independence Vs. uncorrelated Given two random variables X_1 and X_2, if they are independent with each other, then they are uncorrelated. That is,

$$\mathbb{E}_{X_1,X_2}[(x_1 - \mathbb{E}_{X_1}[x_1])(x_2 - \mathbb{E}_{X_2}[x_2])] = 0, \tag{8}$$

where $\mathbb{E}$ corresponds to taking expectation. However, the reverse may not be correct. Please show that X_1 and X_2 are uncorrelated given the joint probability in Table 2 (no more than 3 lines of derivation), but they are not independent (no more than 3 lines of derivation or explanation).

Table 2: The joint probability $P(X_1, X_2)$.

X_1 / X_2	-1	0	1
-1	0	0.25	0
0	0.25	0	0.25
1	0	0.25	0

$$0.25 \times 0 \times -1 + 0.25 \times -1 \times 0 + 0.25 \times 0 \times 1 + 0.25 \times 1 \times 0 = 0$$

$$P(X_1 = -1 | X_2 = 0) = 0.5$$
$$P(X_1 = -1 | X_2 = -1) = 0.0$$

3.3 Explaining away Given a Bayesian network in Fig. 1 with three discrete random variables $X \in \{-1, 1\}$ and $Y \in \{-1, 1\}$ and $Z \in \{-1, 1\}$, together with the probability defined as follows,

$$\begin{aligned} &P(X = 1) = 0.1 \\ &P(Z = 1) = 0.2 \\ &P(Y = 1|X = -1, Z = -1) = 0.01 \\ &P(Y = 1|X = 1, Z = -1) = 0.7 \\ &P(Y = 1|X = -1, Z = 1) = 0.85 \\ &P(Y = 1|X = 1, Z = 1) = 0.95 \end{aligned} \tag{9}$$

please derive the following two terms $P(Z = 1|Y = 1)$ and $P(Z = 1|Y = 1, X = 1)$ (within 5 lines for each term). You should see the explaining away effect that the second term is small than the first one. This also shows that when the value of Y is given, X and Z are not conditionally independent.

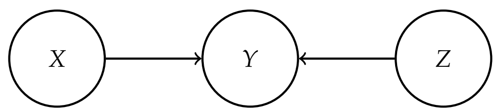

Figure 1: A Bayesian network with three discrete random variables.

$$P(Z = 1|Y = 1) = \frac{P(Z = 1, Y = 1)}{P(Y = 1)} = 0.731$$
$$P(Z = 1|Y = 1, X = 1) = \frac{P(Z = 1, Y = 1, X = 1)}{P(Y = 1, X = 1)} = 0.253$$