GroupAssignment3_Team10

1smat Halabi, Anudeep Battu, Muhammad Sajjad, Muhammad Hasnain Saeed, Lavy Selvaraj

December 22, 2024

Question 1

(a)

See Figure 1

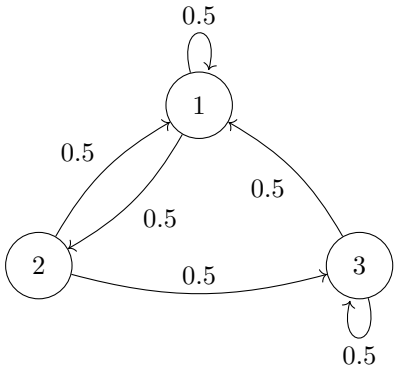

Figure 1: Markov chain for problem 1.

(b)

This Markov chain is aperiodic and irreducible, hence it has a unique and asymptotic stationary distribution. This distribution is found by solving the following system of equations:

$$\begin{cases} \Pi = \Pi \cdot P \\ \sum_{i=1}^{3} \Pi_i = 1 \end{cases} \tag{1}$$

The solution is $\pi = \begin{pmatrix} 0.5 & 0.25 & 0.25 \end{pmatrix}$

(c)

The probability is the solution to the following equation: $P_{t=4} = P_{t=1} \cdot P^3$. Thus $P_{t=4} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \end{pmatrix}$
The probability that the chain is in state 2 at time 4 is 0.25.

(d)

The mean hitting time from one state i to state k is given by the minimum non-negative solution to the following equations, where j is a transition state (`https://www.stat.auckland.ac.nz/~fewster/325/notes/ch8.pdf`):

$$\begin{cases} m_{i,k} = 0 & i = k \\ 1 + \sum_j p_{i,j} \cdot m_{j,k} & i \neq k \end{cases} \tag{2}$$

In our case, we have then:

$$\begin{cases} m_{1,3} = 1 + 0.5 \cdot m_{1,3} + 0.5 \cdot m_{2,3} + 0 \cdot m_{3,3} & = 1 + 0.5 \cdot m_{1,3} + 0.5 \cdot m_{2,3} \\ m_{2,3} = 1 + 0.5 \cdot m_{1,3} + 0.5 \cdot m_{3,3} & = 1 + 0.5 \cdot m_{1,3} \end{cases} \tag{3}$$

The solution is $m_{1,3} = 6$.

(e)

1. The period of state 1 is 1 (self loop).
2. The period of state 2 is 2 $[2, 1, 2]$ can also be 3, for ex. 2 -> 3 -> 1 -> 2, since all these jumps have pos
3. The period of state 3 is 1 (self loop).

Question 2

(a) Empirical Version of Precision and Recall

The empirical version of precision and recall is as follows: the formula is correct, but it's not i

- **Precision:**

$$\mathbb{P}(Y = 1 \mid g(X) = 1) = \frac{\mathbb{E}[\mathbf{1}(Y = 1 \text{ and } g(X) = 1)]}{\mathbb{E}[\mathbf{1}(g(X) = 1)]} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall:**

$$\mathbb{P}(g(X) = 1 \mid Y = 1) = \frac{\mathbb{E}[\mathbf{1}(Y = 1 \text{ and } g(X) = 1)]}{\mathbb{E}[\mathbf{1}(Y = 1)]} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Here:

- TP: True Positives (The model predicts positive, and it is indeed positive)
- FP: False Positives (The model predicts positive, but it is actually negative)
- FN: False Negatives (The model predicts negative, but it is actually positive)

(b) Expected Cost Formula

The cost depends on predictions and outcomes:

- If $g(X) = 1$ (model predicts deterioration):
  - Cost c when $Y = 0$ (False Positive)
  - Cost 0 when $Y = 1$ (True Positive)
- If $g(X) = 0$ (model predicts no deterioration):
  - Cost d when $Y = 1$ (False Negative)
  - Cost 0 when $Y = 0$ (True Negative)

The expected cost is:

$$\mathbb{E}[C] = c \cdot \mathbb{P}(g(X) = 1, Y = 0) + d \cdot \mathbb{P}(g(X) = 0, Y = 1)$$

Using empirical probabilities, this becomes:

$$\mathbb{E}[C] = c \cdot \frac{\text{FP}}{n} + d \cdot \frac{\text{FN}}{n}$$

Rewriting in terms of precision and recall:

$$\mathbb{E}[C] = c \cdot \mathbb{P}(g(X) = 1) \cdot (1 - \text{Precision}) + d \cdot \mathbb{P}(Y = 1) \cdot (1 - \text{Recall})$$

where:

$$\mathbb{P}(g(X) = 1) = \frac{\text{TP} + \text{FP}}{n}, \quad \mathbb{P}(Y = 1) = \frac{\text{TP} + \text{FN}}{n}$$

(c) Confidence Intervals:

Yes, confidence intervals can be produced for expected cost, precision and recall. Two common methods are listed below:

- **Normal Approximation Method:**
  - **Precision:**
$$CI_{\text{Precision}} = \text{Precision} \pm Z \cdot \sqrt{\frac{\text{Precision} \cdot (1 - \text{Precision})}{\text{TP} + \text{FP}}}$$
    where $Z = 1.96$ for 95% confidence.
  - **Recall:**
$$CI_{\text{Recall}} = \text{Recall} \pm Z \cdot \sqrt{\frac{\text{Recall} \cdot (1 - \text{Recall})}{\text{TP} + \text{FN}}}$$
  - **Cost:**
$$\mathbb{E}[C]_{\text{lower}} = c \cdot \mathbb{P}(g(X) = 1) \cdot (1 - \text{Precision}_{\text{upper}}) + d \cdot \mathbb{P}(Y = 1) \cdot (1 - \text{Recall}_{\text{upper}})$$
$$\mathbb{E}[C]_{\text{upper}} = c \cdot \mathbb{P}(g(X) = 1) \cdot (1 - \text{Precision}_{\text{lower}}) + d \cdot \mathbb{P}(Y = 1) \cdot (1 - \text{Recall}_{\text{lower}})$$
  - Suitable for large sample sizes ($\text{TP} + \text{FP} > 30$, $\text{TP} + \text{FN} > 30$).
  - Assumes precision and recall are approximately normal.
  - Provides narrower intervals.

- **DKW Inequality Method:**
  - **Precision:**
$$CI_{\text{Precision}} = \text{Precision} \pm \sqrt{\frac{\ln(2/\alpha)}{2 \cdot (\text{TP} + \text{FP})}}$$
    where $\alpha = 0.05$ for 95% confidence
  - **Recall:**
$$CI_{\text{Recall}} = \text{Recall} \pm \sqrt{\frac{\ln(2/\alpha)}{2 \cdot (\text{TP} + \text{FN})}}$$
  - **Cost:** Use the DKW bounds for precision and recall:
$$\mathbb{E}[C]_{\text{lower}} = c \cdot \mathbb{P}(g(X) = 1) \cdot (1 - \text{Precision}_{\text{upper}}) + d \cdot \mathbb{P}(Y = 1) \cdot (1 - \text{Recall}_{\text{upper}})$$
$$\mathbb{E}[C]_{\text{upper}} = c \cdot \mathbb{P}(g(X) = 1) \cdot (1 - \text{Precision}_{\text{lower}}) + d \cdot \mathbb{P}(Y = 1) \cdot (1 - \text{Recall}_{\text{lower}})$$
  - Suitable for small sample sizes or unknown distributions.
  - Provides a robust, distribution-free bound.
  - Results in wider, more conservative intervals.

Question 3

We are given that $X = (X_1, X_2, \ldots, X_d)$ and $Y = (Y_1, Y_2, \ldots, Y_d)$ are two Gaussian random vectors of zero mean and unit variance. They are orthogonal if $\mathbb{E}(X \cdot Y) = 0$.

$$\begin{aligned}
\mathbb{E}(X \cdot Y) &= \mathbb{E}\left(\sum_{i=1}^{d} X_i Y_i\right) \\
&= \sum_{i=1}^{d} \mathbb{E}(X_i Y_i) \\
&= \sum_{i=1}^{d} \mathbb{E}(X_i)\mathbb{E}(Y_i) \quad \text{(since } X_i \text{ and } Y_i \text{ are independent)} \\
&= 0.
\end{aligned}$$

Thus, $\mathbb{E}(X \cdot Y) = 0$.

$$\begin{aligned}
Var[X \cdot Y] &= \mathbb{E}[(X \cdot Y)^2] - (\mathbb{E}[X \cdot Y])^2 \\
&= \mathbb{E}[(X \cdot Y)^2] \qquad (\text{since } \mathbb{E}[X \cdot Y] = 0) \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{d} X_i Y_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{d} X_i^2 Y_i^2\right] + \mathbb{E}\left[\sum_{i=1}^{d} \sum_{j=1, i\neq j}^{d} (X_i Y_i X_j Y_j)\right] \\
&= \sum_{i=1}^{d} \mathbb{E}(X_i^2 Y_i^2) + \sum_{i=1}^{d} \sum_{j=1, i\neq j}^{d} \mathbb{E}(X_i Y_i X_j Y_j) \\
&= \sum_{i=1}^{d} \mathbb{E}(X_i^2)\mathbb{E}(Y_i^2) + \sum_{i=1}^{d} \sum_{j=1, i\neq j}^{d} \mathbb{E}(X_i)\mathbb{E}(Y_i)\mathbb{E}(X_j)\mathbb{E}(Y_i) \text{ (since } X_i \text{ and } Y_i \text{ are independent)} \\
&= \sum_{i=1}^{d} Var(X_i) Var(Y_i) \qquad (\mathbb{E}(X_i) = \mathbb{E}(Y_i) = 0) \\
&= \sum_{i=1}^{d} 1 \qquad (Var(X_i) = Var(Y_i) = 1) \\
&= d
\end{aligned}$$

Now, to find an upper bound for $\mathbb{P}(X \cdot Y \geq \epsilon)$, we proceed as follows:

$$\begin{aligned}
\mathbb{P}(X \cdot Y \geq \epsilon) &= \mathbb{P}(e^{t(X \cdot Y)} \geq e^{t\epsilon}) \quad \text{for some } t > 0 \\
&\leq \frac{\mathbb{E}(e^{t(X \cdot Y)})}{e^{t\epsilon}} \quad \text{(Markov's inequality)} \\
&= \frac{\mathbb{E}\left(e^{t\sum_{i=1}^{d} X_i Y_i}\right)}{e^{t\epsilon}}.
\end{aligned}$$

Markov inequality only works for

(e) Using the moment generating function (MGF) of X_iY_i, we find:

$$\mathbb{P}(X \cdot Y \geq \epsilon) \leq \inf_{t>0} \frac{\prod_{i=1}^{d} \mathbb{E}(e^{tX_iY_i})}{e^{t\epsilon}},$$

where $\mathbb{E}(e^{tX_iY_i}) = \frac{1}{\sqrt{1-t^2}}$ for $|t| < 1$.

Some mistakes in the period calculation, notation for empirical precision and recall, and misargumentation in the lower bound of the dot product, but apart from hat solutions look good! Great job, just have my comments in mind for the exam!

Question 4

(a)

A matrix is rank-one if it can be expressed as the outer product of two non-zero vectors. In our case, $u_iu_i^\top$ is clearly an outer product of u_i with itself. Since u_i is a unit vector ($\|u_i\| = 1$), $u_iu_i^\top$ is a rank-one matrix. For any vector $\mathbf{x} \in \mathbb{R}^n$:

$$u_iu_i^\top \mathbf{x} = u_i(u_i^\top \mathbf{x}).$$

The null space (or kernel) of a matrix A, denoted Null(A), consists of all vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. Therefore:

$$\text{Null}\left(u_iu_i^\top\right) = \left\{\mathbf{x} \in \mathbb{R}^n \mid u_i^\top \mathbf{x} = 0\right\},$$

which is the set of all vectors orthogonal to u_i.

The range of a matrix A, denoted Range(A), consists of all possible linear combinations of its columns. In the case of $u_iu_i^\top$, the range is always a scalar multiple of u_i. Therefore:

$$\text{Range}\left(u_iu_i^\top\right) = \text{span}\{u_i\}.$$

(b)

To verify that the matrix

$$U = \sum_{i=1}^{r} u_iu_i^\top$$

We know that each outer product $u_iu_i^\top$ is a rank-one matrix since it is the product of a non-zero vector u_i with itself. Given that the vectors $u_1, \ldots, u_r$ are linearly independent, these rank-one matrices contribute uniquely to the structure of U.

Specifically, for any vector $\mathbf{x} \in \mathbb{R}^n$,

$$U\mathbf{x} = \sum_{i=1}^{r} u_iu_i^\top \mathbf{x} = \sum_{i=1}^{r} u_i(u_i^\top \mathbf{x}),$$

which is a linear combination of the vectors $u_1, \ldots, u_r$. Since these vectors are linearly independent, the span has dimension r.

Therefore, the range of U is an r-dimensional subspace of $\mathbb{R}^n$, which means:

$$\text{rank}(U) = r.$$

(c)

Since each $u_iu_i^\top$ is symmetric, their sum U is also a symmetric matrix. Also, U is positive semidefinite (PSD) because for any vector $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^\top U\mathbf{x} = \sum_{i=1}^{r} (\mathbf{u}_i^\top \mathbf{x})^2 \geq 0.$$

For any real symmetric PSD matrix, the singular vectors exactly match with the eigenvectors, and the singular values are the nonnegative eigenvalues.

(e)1. **Are the Vectors $u_1, \ldots, u_r$ the Same as the Right Singular Vectors of U?**

In general, the vectors $\{u_1, \ldots, u_r\}$ are **not** the right singular vectors of U. Although each $u_i u_i^\top$ contributes to U, the eigenvectors (and hence the singular vectors) of U depend on the collective combination of these outer products. If the vectors $u_1, \ldots, u_r$ are not mutually orthogonal, the resulting matrix U has eigenvectors that are linear combinations of the u_is rather than the u_is themselves.

For example, if:

$$u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Both vectors are linearly independent but not orthogonal since $u_1^\top u_2 = 1 \neq 0$.

$$U = u_1 u_1^\top + u_2 u_2^\top = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Solving $(U - \lambda I)\mathbf{v} = \mathbf{0}$ yields eigenvectors that are linear combinations of u_1 and u_2, not identical to u_1 or u_2. Therefore, u_1 and u_2 are not the right singular vectors of U in this case.

2. **If the Vectors $u_1, \ldots, u_r$ are All Orthogonal, What Are the Singular Values of U?**

When the vectors $\{u_1, \ldots, u_r\}$ are mutually orthogonal and each u_i has unit length (i.e., they form an orthonormal set), the matrix U simplifies significantly. Each outer product $u_i u_i^\top$ acts as an orthogonal projector onto the subspace spanned by u_i. Summing these projectors yields

$$U = \sum_{i=1}^{r} u_i u_i^\top,$$

which is the orthogonal projector onto the subspace spanned by all u_i.

In this scenario:

$$U u_i = \sum_{j=1}^{r} u_j u_j^\top u_i = u_i u_i^\top u_i = u_i \cdot 1 = u_i,$$

indicating that each u_i is an eigenvector of U with eigenvalue 1. Any vector orthogonal to all u_is is in the null space of U, corresponding to eigenvalue 0. Consequently, U has:

$$\text{eigenvalues} = \underbrace{1, 1, \ldots, 1}_{r \text{ times}}, \quad \underbrace{0, 0, \ldots, 0}_{n-r \text{ times}}.$$

Since U is symmetric and PSD, its singular values are identical to its eigenvalues. Therefore, the singular values of U are:

$$\sigma_1 = \sigma_2 = \cdots = \sigma_r = 1, \quad \sigma_{r+1} = \cdots = \sigma_n = 0.$$

In this orthonormal case, the vectors $\{u_1, \ldots, u_r\}$ are same as the eigenvectors as well as the right singular vectors of U, each corresponding to a singular value of 1.

Question 5

Let $X \sim \text{Uniform}(B_1)$ and define $Y = \|X\|_2$ (the Euclidean norm).

(a) Distribution function of $Y = \|X\|_2$

Given that $X \sim \text{Uniform}(B_1)$, where B_1 is the unit ball in $\mathbb{R}^n$, we need to find the distribution of the Euclidean norm $Y = \|X\|_2$. The random vector X is uniformly distributed over the unit ball.

The cumulative distribution function of Y, denoted by :

$$F_Y(y) = P(Y \leq y),$$

is the probability that Y is less than or equal to a certain value y.

For a uniform distribution over the unit ball, the probability that $\|X\|_2 \leq y$ is proportional to the volume of the ball with radius y in n-dimensional space. Therefore, the cumulative distribution function is given by:

$$F_Y(y) = P(Y \leq y) = \frac{\text{Volume of ball of radius } y}{\text{Volume of unit ball}} = y^n \quad \text{for} \quad 0 \leq y \leq 1.$$

The probability density function , $f_Y(y)$, is the derivative of the CDF:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = ny^{n-1} \quad \text{for} \quad 0 \leq y \leq 1.$$

(b) Distribution of $\ln(1/Y)$

Now, lets assume random variable $Z = \ln(1/Y)$.

The cumulative distribution function of Z, denoted by $F_Z(z)$, is related to the cumulative distribution function of Y as follows:

$$F_Z(z) = P(Z \leq z) = P(\ln(1/Y) \leq z) = P(Y \geq e^{-z}).$$

Since the cumulative distribution function of Y is $F_Y(y) = y^n$, we have:

$$F_Z(z) = P(Y \geq e^{-z}) = 1 - P(Y \leq e^{-z}) = 1 - F_Y(e^{-z}) = 1 - (e^{-z})^n = 1 - e^{-nz}.$$

Thus, the cumulative distribution function of Z is:

$$F_Z(z) = 1 - e^{-nz} \quad \text{for} \quad z \geq 0.$$

The probability density function of Z is the derivative of the cumulative distribution function:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = ne^{-nz} \quad \text{for} \quad z \geq 0.$$

(c) Calculate $E[\ln(1/Y)]$

First we will calculate $E[\ln(1/Y)]$ using the distribution of Y. The expected value is:

$$E[\ln(1/Y)] = \int_0^1 \ln(1/y) f_Y(y)\, dy.$$

Substitute $f_Y(y) = ny^{n-1}$ into the integral:

$$E[\ln(1/Y)] = \int_0^1 \ln(1/y) ny^{n-1}\, dy.$$

We can simplify $\ln(1/y) = -\ln(y)$, so the integral becomes:

$$E[\ln(1/Y)] = -n \int_0^1 \ln(y) y^{n-1}\, dy.$$

(e) This integral is a standard result:

$$\int_0^1 \ln(y) y^{n-1}\, dy = -\frac{1}{n^2}.$$

Therefore:

$$E[\ln(1/Y)] = -n \cdot \left(-\frac{1}{n^2}\right) = \frac{1}{n}.$$

Secondly, we will compute $E[\ln(1/Y)]$ using the distribution of $Z = \ln(1/Y)$. The expected value is:

$$E[Z] = \int_0^\infty z f_Z(z)\, dz = \int_0^\infty z n e^{-nz}\, dz.$$

This is a standard exponential integral, and the result is:

$$E[Z] = \frac{1}{n}.$$