End Semester Exam - Statistical Foundations for ML

May 11, 2025

Information

- Total marks: 50
- There are two sections in the exam.
  - Section A contains Short Answer Questions and carries 10 marks. Each question carries 2 marks.
  - Section B contains Numerical questions and carries 40 marks. Each question carries 5 marks.
- Write down all your answers in the answer sheet provided to you.
- A z-table and a small t-table is provided at the end of the question paper to aid calculations.

Section A: Short Answer Questions

1. (2 points) A number of questions were posed to a random sample of visitors to a Mumbai tourist information center.
   (a) Are you staying overnight in Mumbai?
   (b) How many times have you visited Mumbai previously?
   (c) How likely are you to visit Mumbai again in the next 12 months: (1) unlikely, (2) likely, (3) very likely?

   For each of the above, describe the type of data obtained as nominal/ordinal/numerical.

Solution:

(a) Are you staying overnight in Mumbai?

- Possible responses: “Yes” or “No.”
- Type: Nominal (binary categories with no inherent order).

(b) How many times have you visited Mumbai previously?

- Possible responses: 0, 1, 2, . . .

- Type: Numerical discrete (counts).

(c) How likely are you to visit Mumbai again in the next 12 months: (1) unlikely, (2) likely, (3) very likely?

- Responses on a 3-point scale with order.
- Type: Ordinal (ranked categories).

2. (2 points) Given some population parameter θ, a point estimator $\hat{\theta}$ is said to be an *unbiased estimator* of θ if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Let $X_1, X_2, \ldots, X_n$ be a sample of values for a population with a mean μ and variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and let S^2 be an estimator of σ^2, where

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Is S^2 unbiased? Answer yes/no and briefly explain why.

Solution: Answer is **no,** because

$$E[S^2] = \frac{n-1}{n}\sigma^2 < \sigma^2$$

i.e. $E[S^2]$ underestimates the true variance. The unbiased sample variance uses $\frac{1}{n-1}$.

3. (2 points) Consider the following data:

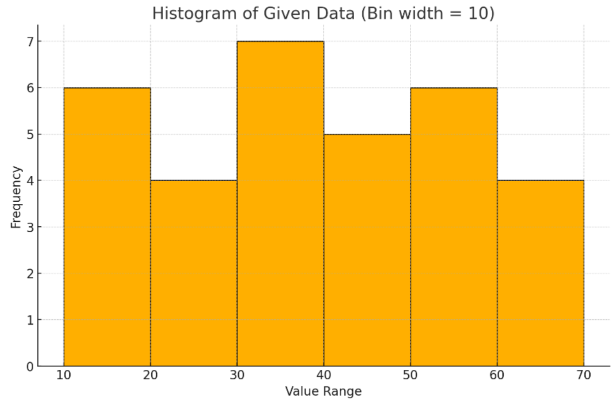
62	15	65	28	51	24	65	39
41	35	15	39	32	36	37	40
21	44	37	59	13	44	56	12
54	64	59	17	27	52	13	45

Using the above data, construct a histogram (choose an appropriate bin size).

Solution: Using bin width 10 over the range 10–70, the frequency counts are:

Bin (Value range)	**Frequency**
10-20	6
20-30	4
30-40	7
40-50	5
50-60	6
60-70	4

The histogram is as follows:



4. (2 points) Suppose u is a unit vector in $\mathbb{R}^n$, so $\|u\| = 1$. This problem is about the n by n symmetric matrix $H = I - 2uu^T$.
   (a) Show directly that $H^2 = I$. Is H orthogonal?
   (b) One eigenvector of H is u itself. Find the corresponding eigenvalue.

Solution: (a) Explicitly, we find $H^2 = (I - 2uu^T)^2 = I^2 - 4uu^T + 4uu^T uu^T$. Since $u^T u = 1$, $H^2 = I^2 - 4uu^T + 4u(u^T u)u^T = I - 4uu^T + 4u(1)u^T = I - 4uu^T + 4uu^T = I$.

Since $H = H^T$, we also have $H^T H = I$, implying that H is an **orthogonal** (or **unitary**) matrix. The blank should be filled with: **orthogonal** (or unitary).

(b) Since $Hu = (I - 2uu^T)u = Iu - 2uu^T u = u - 2u(u^T u) = u - 2u(1) = u - 2u = -u$. The corresponding eigenvalue is $\lambda = -1$.

5. (2 points) A random sample of data has a mean of 75 and a variance of 25.
   (a) Use Chebyshev's theorem to determine the percent of observations between 65 and 85.
   (b) If the data is normally distributed, use the empirical rule to find the approximate percent of observations between 65 and 85. Is this estimate less/more informative than the estimate obtained in part (a)?

Solution: A sample has mean 75 and variance 25 ($\sigma = 5$).

(a) **Chebyshev's theorem:** For any distribution, at least $1 - \frac{1}{k^2}$ of observations lie within k standard deviations of the mean. Here the interval 65–85 is $75 \pm 2\sigma$. Thus $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 0.75$. So at least 75% of observations lie between 65 and 85.

(b) **Empirical (68–95–99.7) rule:** If the data are approximately normally distributed, then about 95% of observations lie within two standard deviations of the mean. So, under the normality assumption, $\approx 95\%$ lie between 65 and 85.

The estimate obtained using the empirical rule is more informative because it uses the information that the data is normally distributed, unlike Chebyshev's theorem. However, Chebyshev's theorem is more useful when the data distribution is not known.

Section B: Numerical Questions

6. The following data give X, the price charged per piece of plywood (in ₹), and Y, the quantity sold (in thousands).

Price per piece (in ₹), X	Thousands of pieces sold, Y
24.49	80
26.46	60
28.28	70
30.00	40
31.62	0

(a) (2 points) Let the line $\hat{y} = b_0 + b_1 x^2$ be the regression line fit on the given data. Compute b_0 and b_1, and interpret the value of b_1.

(b) (2 points) Let the relation between X and Y be $Y = \beta_0 + \beta_1 X^2 + \epsilon$. Test if $\beta_1 = 0$ at the significance level of 0.95.

(c) (1 point) What quantity of plywood would you expect to sell if the price were ₹26.46 per piece?

Solution: *(Regression)*
Part (a): Compute b_0 and b_1: The regression equation is given by

$$\hat{y} = b_0 + b_1 X^2.$$

We need to calculate b_0 (intercept) and b_1 (slope). Note that the regression line describes the linear relationship between Y and X^2, not Y and X. So we construct a new variable Z using the transformation $Z = X^2$. The regression equation is given by

$$\hat{y} = b_0 + b_1 Z.$$

From the given data, we have

Price per piece (X in Rs.)	Quantity sold (Y in thousands)	$Z = X^2$
24.49	80	$599.5001 \approx 600$
26.46	60	$699.1716 \approx 700$
28.28	70	$800.0384 \approx 800$
30.00	40	$900.0000 \approx 900$
31.62	0	$999.7444 \approx 1000$

We now calculate the necessary sums:

$$\bar{Z} = \frac{\sum_i Z_i}{5} = \frac{600 + 700 + 800 + 900 + 1000}{5} = 800$$
$$\bar{Y} = \frac{\sum_i Y_i}{5} = \frac{80 + 60 + 70 + 40 + 0}{5} = 50$$

$$\sum_i Z_i Y_i = (600 \times 80) + (700 \times 60) + (800 \times 70) + (900 \times 40) + (1000 \times 0) = 182000$$
$$\sum Z_i^2 = 600^2 + 700^2 + 800^2 + 900^2 + 1000^2 = 3300000$$
$$\sum Y_i^2 = 80^2 + 60^2 + 70^2 + 40^2 + 0^2 = 16500$$

From the above, we obtain

$$S_{ZY} = \sum_i Z_i Y_i - 5\bar{Z}\bar{Y} = 182000 - 5(800)(50) = -18000$$
$$S_{ZZ} = \sum_i Z_i^2 - 5\bar{Z}^2 = 3300000 - 5(800)^2 = 100000$$
$$S_{YY} = \sum_i Y_i^2 - 5\bar{Y}^2 = 16500 - 5(50)^2 = 4000$$

We can calculate b_0 and b_1 using the above.

$$b_1 = \frac{S_{ZY}}{S_{ZZ}} = \frac{-18000}{100000} = -0.18$$
$$b_0 = \bar{Y} - b_1\bar{Z} = 50 - (-0.18)800 = 50 + 144 = 194$$

Thus, the regression equation is:

$$\hat{y} = 194 - 0.18x^2$$

Part (b): Test if $\beta_1 = 0$

We need to test the null hypothesis $H_0 : \beta_1 = 0$ at the 0.05 significance level. The number of samples $n = 5$. The test statistic is given by

$$T = \sqrt{\frac{(n-2)S_{ZZ}}{SS_R}}|b_1|$$

We reject the null hypothesis if $T > t_{\alpha/2,n-2} = t_{0.025,3}$ (because the significance level $\alpha = 0.05$). We first calculate SS_R.

$$SS_R = \frac{S_{ZZ}S_{YY} - S_{ZY}^2}{S_{ZZ}} = \frac{(100000)(4000) - (-18000)^2}{100000} = \frac{400000000 - 324000000}{10000} = 760.$$

Using this, we get

$$T = \sqrt{\frac{(n-2)S_{ZZ}}{SS_R}}|b_1| = \sqrt{\frac{(3)(100000)}{760}}(0.18) = 3.58$$

From the t-table, we obtain that $t_{0.025,3}$ is approximately 3.182. Because $T = 3.58$, we can see that $T > t_{0.025,3}$, and hence the null hypothesis is rejected.

Conclusion: β_1 is significantly different from zero.

Part (c): Predict the quantity sold when $X = 26.46$

We need to compute X^2 when $X = 26.46$:

$$X^2 = 26.46^2 = 699.1716 \approx 700$$

Substitute this into the regression equation:

$$\hat{y} = 194 - 0.18 \times 700$$
$$\hat{y} = 194 - 126 = 68$$

Thus, the expected quantity sold when the price per piece is Rs. 26.46 is approximately 68 thousand pieces.

7. Let X be a non-negative random variable i.e. $P(X \geq 0) = 1$. Then, for some $\epsilon > 0$,

$$P(X \geq \epsilon) \leq \frac{E[X]}{\epsilon}.$$

This inequality is called the *Markov inequality.* When ϵ is much larger than $E[X]$, the quantity on the left, $P(X \geq \epsilon)$, can be interpreted as the probability that X deviates too much from the expected value $E[X]$.

(a) (4 points) Prove the Markov inequality assuming X is a discrete random variable. (Hint: Start with the definition of expectation)

(b) (1 point) Let Y denote the number of customers that arrive at a particular restaurant in a day. It is known from past data that 500 customers visit the restaunt every day on average. What is the maximum possible probability that more than 1200 customers visit the restaurant today?

Solution: *(Random variables and expectations)*

(a) Let X be a discrete non-negative random variable. We start from the definition of expectation:

$$\begin{aligned} E[X] &= \sum_{x=-\infty}^{\infty} xP(X = x) \\ &= \sum_{x=0}^{\infty} xP(X = x) \quad \text{(because } X \text{ is non-negative)} \\ &= \sum_{0 \leq x < \epsilon} xP(X = x) + \sum_{x \geq \epsilon} xP(X = x) \end{aligned}$$

Notice that the first term on the RHS is non-negative. We can drop it and introduce an inequality i.e.

$$\begin{aligned}
E[X] &\geq \sum_{x \geq \epsilon} xP(X = x) \quad \text{(because the first term above is non-negative)} \\
&\geq \sum_{x \geq \epsilon} \epsilon P(X = x) \quad \text{(because each } x \text{ is greater than } \epsilon\text{)} \\
&\geq \epsilon \sum_{x \geq \epsilon} P(X = x) \\
&\geq \epsilon P(X \geq \epsilon)
\end{aligned}$$

Dividing by ϵ on both sides gives the final inequality

$$\frac{E[X]}{\epsilon} \geq P(X \geq \epsilon).$$

(b) Given $E[Y] = 500$, where Y is the number of customers, we need $P(Y > 1200)$. Apply the Markov inequality with $\epsilon = 1200$:

$$P(Y \geq 1200) \leq \frac{E[Y]}{1200} = \frac{500}{1200} = \frac{5}{12} \approx 0.4167.$$

The maximum possible probability is $\frac{5}{12}$.

8. A consultant knows that it will cost him ₹10,00,000 to fulfill a particular contract. The contract is to be put out for bids, and he believes that the lowest bid, excluding his own, can be represented by a distribution that is uniform between ₹8,00,000 and ₹20,00,000. Therefore, if the random variable X denotes the lowest of all other bids (in lakhs of rupees), its probability density function is as follows:

$$f(x) = \begin{cases} c & \text{for } 8 \leq x \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

Answer the following questions:

(a) (1 point) What is the value of c?

(b) (1 point) What is the probability that the lowest of the other bids will be less than the consultant's cost estimate of ₹10,00,000?

(c) (1 point) If the consultant submits a bid of ₹12,00,000, what is the probability that he will secure the contract?

(d) (1 point) The consultant decides to submit a bid of ₹12,00,000. What is his expected profit from this strategy?

(e) (1 point) If the consultant wants to submit a bid so that his expected profit is as high as possible, discuss how he should go about making this choice.

Solution: *(Uniform random variable - CDF, PDF, Expectation)*
The lowest bid X (in lakhs) is uniform over $[8, 20]$, with density:

$$f(x) = \begin{cases} c & \text{for } 8 \leq x \leq 20, \\ 0 & \text{otherwise.} \end{cases}$$

8(a):

$$\int_8^{20} c\,dx = c(20 - 8) = 12c = 1 \implies c = \frac{1}{12}.$$

8(b):

$$P(X < 10) = \int_8^{10} \frac{1}{12}\,dx = \frac{1}{12}(10 - 8) = \frac{1}{6} \approx 0.1667.$$

8(c): The consultant wins if $X > 12$:

$$P(X > 12) = \int_{12}^{20} \frac{1}{12}\,dx = \frac{1}{12}(20 - 12) = \frac{2}{3} \approx 0.6667.$$

8(d): Profit is ₹2 lakhs if $X > 12$.

Expected profit:

$$E[\text{Profit}] = P(X > 12) \cdot 2 = \frac{2}{3} \cdot 2 = \frac{4}{3}$$

8(e): Say the consultant bids b lakhs. We know that the consultant incurs a cost of 10 lakhs. The profit if consultant wins is $b - 10$. Probability of winning is given by:

$$P(X > b) = \int_b^{20} \frac{1}{12}\,dx = \frac{20 - b}{12}.$$

The expected profit, denoted by $g(b)$, is a function of the bid b. The consultant gets no profit when the bid loses. Hence, the expected profit is given by:

$$g(b) = \text{Profit} \times \text{Probability of winning} = \frac{(b - 10)(20 - b)}{12}.$$

$$g'(b) = \frac{-b + 15}{6} = 0 \implies b = 15.$$

$$g''(b) = \frac{-1}{6} < 0 \text{ (maximum).}$$

At $b = 15$:

$$P(X > 15) = \frac{5}{12}, \quad E[\text{Profit}] = \frac{5}{12} \cdot 5 = \frac{25}{12}.$$

Since $b = 15$ maximizes expected profit, bid ₹15,00,000.

9. The fuel consumption, in kilometers per litre, of all cars of a particular model has a mean of 25 and

a standard deviation of 2. The population distribution can be assumed to be normal. A random sample of these cars is taken.

(a) (3 points) Find the probability that sample mean fuel consumption will be fewer than 24 kilometers per litre if

i. a sample of 1 observation is taken.

ii. a sample of 4 observations is taken.

iii. a sample of 16 observations is taken.

(b) (2 points) Explain why the three answers in part (a) differ in the way they do. Draw a graph to illustrate your reasoning.

Solution: *(Central Limit Theorem)*
The population fuel consumption (km per litre) is $X \sim N(\mu = 25, \sigma = 2)$. For an i.i.d. sample of size n, the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is distributed as

$$\bar{X}_n \sim N\Big(\mu_{\bar{X}} = 25,\ \sigma^2_{\bar{X}} = \frac{2^2}{n}\Big), \quad \sigma_{\bar{X}} = \frac{2}{\sqrt{n}}.$$

We seek

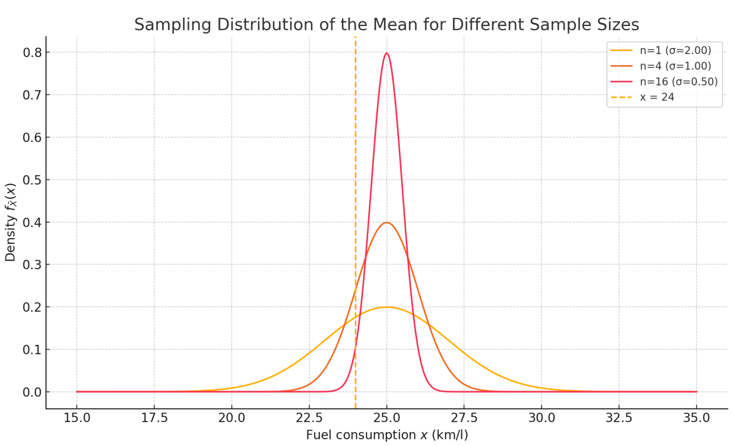
$$P\big(\bar{X}_n < 24\big) = P\Big(Z < \frac{24 - 25}{2/\sqrt{n}}\Big) = \Phi\Big(-\tfrac{1}{2}\sqrt{n}\Big), \quad Z \sim N(0, 1).$$

where Z is a standard normal random variable, and Φ is the CDF of Z.

1. For each n:
   - $n = 1$: $P(\bar{X}_1 < 24) = \Phi(-\frac{\sqrt{1}}{2}) = \Phi(-0.5) \approx 0.3085$.
   - $n = 4$: $P(\bar{X}_4 < 24) = \Phi(-\frac{\sqrt{4}}{2}) = \Phi(-1) \approx 0.1587$.
   - $n = 16$: $P(\bar{X}_{16} < 24) = \Phi(-\frac{\sqrt{16}}{2}) = \Phi(-2) \approx 0.0228$.

2. As n increases, the standard error $\sigma_{\bar{X}} = 2/\sqrt{n}$ decreases, so the sampling distribution of $\bar{X}$ becomes more concentrated around 25. A fixed deviation below the mean $(24 = 25 - 1)$ thus has smaller probability for larger n.

   The following figure displays the densities of the sampling distribution for $n = 1, 4, 16$ together with a vertical line at $x = 24$.



10. A college admissions officer for an MBA program has determined that historically applicants have undergraduate grade point averages that are normally distributed with standard deviation 0.45. From a random sample of 25 applications from the current year, the sample mean grade point average is 2.90.

   (a) (3 points) Find a 95% confidence interval for the population mean.

   (b) (2 points) Based on these sample results, a statistician computes for the population mean a confidence interval extending from 2.81 to 2.99. Find the confidence level associated with this interval.

Solution: *(Confidence Intervals)*

(a) Confidence interval for the population mean μ when the population standard deviation σ is known is

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where $\bar{x} = 2.90$, $\sigma = 0.45$ and $z_{\alpha/2}$ is the critical value from the standard normal distribution for a $100(1-\alpha)\%$ confidence level.

For a 95% confidence level, the confidence coefficient is 0.95, so $\alpha = 1 - 0.95 = 0.05$, and $\alpha/2 = 0.025$. From the z-table, we get $z_{0.025} \approx 1.96$.

$$\frac{\sigma}{\sqrt{n}} = \frac{0.45}{\sqrt{25}} = \frac{0.45}{5} = 0.09$$

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot 0.09 = 0.1764$$

Thus, the 95% confidence interval is:

$$(2.90 - 0.1764, 2.90 + 0.1764) = (2.7236, 3.0764)$$

(b) Given a confidence interval from 2.81 to 2.99, we need to find the associated confidence level. $\bar{x} = 2.90$, The interval is centered around the mean, so the margin of error E is:

$$E = (2.99 - 2.81)/2 = 0.09$$

The margin of error is given by:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Using the known values $E = 0.09$, $\sigma = 0.45$, $n = 25$, we get

$$0.09 = z_{\alpha/2} \cdot \frac{0.45}{5} = z_{\alpha/2} \cdot 0.09$$

$$\implies z_{\alpha/2} = \frac{0.09}{0.09} = 1$$

The value $z_{\alpha/2} = 1$ corresponds to the z-score where $P(Z > 1) = \alpha/2$. From standard normal tables, $P(Z \leq 1) \approx 0.8413$, so:

$$P(Z > 1) = 1 - 0.8413 = 0.1587$$

Thus, $\alpha/2 = 0.1587$, and $\alpha = 2 \cdot 0.1587 = 0.3174$. The confidence level is:

$$1 - \alpha = 1 - 0.3174 = 0.6826 \quad \text{or} \quad 68.26\%$$

11. A specialized laboratory monitors the occurrence of rare particle decays in a controlled environment, where the number of decays per hour is modeled as a Poisson process. Historical data from 10 independent 1-hour observation periods, conducted under identical conditions, recorded the following decay counts: 3, 5, 2, 4, 6, 3, 5, 4, 3, 2. The lab's theoretical model predicts an average decay rate of 4.5 decays per hour, but recent equipment upgrades may have altered the underlying rate.

    (a) (1 point) Using the provided data, determine the most appropriate statistical measures to characterize the decay rate.

    (b) (2 points) Evaluate whether the observed data align with expectations for stable performance.

    (c) (2 points) Comment on whether the performance of the equipment used to measure the decay rates is sufficiently stable. Is the observed data sufficient to make a conclusion?

    Table 1 (available at the end) provides critical values for the t-distribution for some degrees of freedom (df):

Solution: *(Maximum likelihood estimation and Hypothesis testing)*
Part (a): For $X_i \sim \text{Poisson}(\lambda)$, the MLE is:

$$\hat{\lambda} = \bar{X} = \frac{3+5+2+4+6+3+5+4+3+2}{10} = 3.7.$$

The estimated mean rate is 3.7 decays/hour.

Sample variance:

$$s^2 = \frac{1}{9}\sum(x_i - 3.7)^2 \approx 1.8, \quad s \approx 1.342.$$

Parts (b) and (c) can be answered using a two-sided test or a one-sided test.

For Two-sided Test:

Part (b): Test $H_0 : \lambda = 4.5$ vs. $H_1 : \lambda \neq 4.5$ using a t-test:
Standard error:

$$\text{SE} = \frac{s}{\sqrt{10}} \approx 0.424.$$

$$t = \frac{3.7 - 4.5}{0.424} \approx -1.887.$$

For $df = n - 1 = 9$ $t_{0.025,9} \approx -2.262$. Since $|t| < |-2.262|$, fail to reject H_0.

Part (c): For a 95% CI ($df = 9$, $t_{0.025,9} \approx -2.262$):

$$3.7 \pm 2.262 \cdot 0.424 \approx 3.7 \pm 0.959 \approx (2.741, 4.659).$$

The estimated rate of 3.7 is below 4.5, but the CI includes 4.5. The equipment appears stable, but the wide CI suggests collecting more data for precision.

For One-sided Test:

Part (b): Test $H_0 : \lambda = 4.5$ vs. $H_1 : \lambda < 4.5$ using a t-test:
Standard error:

$$\text{SE} = \frac{s}{\sqrt{10}} \approx 0.424.$$

$$t = \frac{3.7 - 4.5}{0.424} \approx -1.887.$$

For $df = n - 1 = 9$, $t_{0.05,9} \approx -1.833$. Since $t < -1.833$, we reject H_0.

Part (c): For a 95% CI ($df = 9$, $t_{0.025,9} \approx 2.262$):

$$3.7 \pm 2.262 \cdot 0.424 \approx 3.7 \pm 0.959 \approx (2.741, 4.659).$$

The interval includes 4.5, but we rejected H_0, indicating that the equipment may not be stable. Further data are needed for a robust conclusion.

12. (5 points) Let $A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$. This problem is concerned with solving the differential equation $\frac{d}{dt}u(t) = Au(t)$, where $u(0) = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$. The solution to the differential equation is given by

$$u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2,$$

where $c_1, c_2, \lambda_1, \lambda_2$ are scalars, while x_1, x_2 are vectors with two components. Find $c_1, c_2, \lambda_1, \lambda_2, x_1, x_2$.

Solution: *(Eigenvalues, eigenvectors, system of equations)*
We need to solve the differential equation $\frac{d}{dt}u(t) = Au(t)$, where:

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}, \quad u(0) = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

The solution is given by:

$$u(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2$$

We need to find the eigenvalues λ_1, λ_2 and corresponding eigenvectors x_1, x_2 and coefficients c_1, c_2. (Refer to Pg. 19 of slides on eigenvalues and eigenvectors.)

Step 1: Find eigenvalues of A
The eigenvalues of A are found by solving the characteristic equation $\det(A - \lambda I) = 0$, where I is the 2×2 identity matrix. Compute:

$$A - \lambda I = \begin{bmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{bmatrix}$$

The determinant is:

$$\det(A - \lambda I) = (1-\lambda)(4-\lambda) - (-1)\cdot 2 = (1-\lambda)(4-\lambda) + 2$$

$$= 4 - \lambda - 4\lambda + \lambda^2 + 2 = \lambda^2 - 5\lambda + 6$$

Solve the quadratic equation:

$$\lambda^2 - 5\lambda + 6 = 0$$

$$\lambda = \frac{5 \pm \sqrt{25-24}}{2} = \frac{5 \pm 1}{2}$$

$$\lambda_1 = 3, \quad \lambda_2 = 2$$

Step 2: Find eigenvectors of A

1. For $\lambda_1 = 3$: Solve $(A - 3I)x_1 = 0$:

$$A - 3I = \begin{bmatrix} 1-3 & -1 \\ 2 & 4-3 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives the system:

$$-2x - y = 0 \quad \Rightarrow \quad y = -2x$$

$$2x + y = 0 \quad \Rightarrow \quad 2x + y = 0 \quad \Rightarrow \quad y = -2x$$

The equations are consistent. Choose $x = 1$, then $y = -2$, so the eigenvector is:

$$x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

The eigenvector is not unique, any choice satisfying $y = -2x$ is correct

2. For $\lambda_2 = 2$: Solve $(A - 2I)x_2 = 0$:

$$A - 2I = \begin{bmatrix} 1-2 & -1 \\ 2 & 4-2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives:

$$-x - y = 0 \quad \Rightarrow \quad y = -x$$

$$2x + 2y = 0 \quad \Rightarrow \quad x + y = 0 \quad \Rightarrow \quad y = -x$$

Choose $x = 1$, then $y = -1$, so the eigenvector is:

$$x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Eigenvector is not unique, any choice satisfying $y = -x$ is correct.

Step 3: Finding coefficients c_1 and c_2
The general solution is:

$$u(t) = c_1 e^{3t} \begin{bmatrix} 1 \\ -2 \end{bmatrix} + c_2 e^{2t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

To find c_1 and c_2, use the initial condition $u(0) = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$:

$$u(0) = c_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ -2c_1 - c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

This gives the system of equations:

$$c_1 + c_2 = 0$$
$$-2c_1 - c_2 = 6$$

From the first equation, $c_2 = -c_1$. Substitute into the second:

$$-2c_1 - (-c_1) = 6 \quad \Rightarrow \quad -2c_1 + c_1 = 6 \quad \Rightarrow \quad -c_1 = 6 \quad \Rightarrow \quad c_1 = -6$$
$$c_2 = -c_1 = 6$$

Thus, the solution is:

$$u(t) = -6e^{3t} \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 6e^{2t} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -6e^{3t} + 6e^{2t} \\ 12e^{3t} - 6e^{2t} \end{bmatrix}$$

$$\lambda_1 = 3, \quad \lambda_2 = 2, \quad x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$c_1 = -6, c_2 = 6$$

Note, the coefficients c_1, c_2 depend on the choice of eigenvectors x_1 and x_2 and are not unique. But the equation $u(t)$ is unique as the product of c_1, x_1 and c_2, x_2 are unique.

13. (5 points) This question is about a basic text classification algorithm A that maps a given text document d to a single topic $i \in \{1, 2, 3, 4\}$ i.e. $A(d) = i$. Each possible output of the algorithm corresponds to a specific topic: 1-Education, 2-Sports, 3-Entertainment, 4-Commercial. The algorithm works in three steps:

    i. A text encoder ϕ maps a given document d to an encoding x which is a 2-dimensional vector. Mathematically, $\phi(d) = x$.

    ii. Given a topic i and an encoding x, a similarity finding function s calculates the cosine similarity between the document encoding x and the encoding of the topic ϕ_i i.e.

    $$g_2(\phi_i, x) = s_i = \cos(\theta_i)$$

    where θ_i is the angle between ϕ_i and x. A sequence of similarities s_1, s_2, s_3, s_4 are calculated with respect to the encoding of all the words w_1, w_2, w_3, w_4 that are available to the algorithm. The following table contains the list of all the topics and their encodings available to the algorithm.

Topic i	Topic name	Encoding ϕ_i
1	Education	$\begin{bmatrix} 4 & 3 \end{bmatrix}$
2	Sports	$\begin{bmatrix} -1 & -1 \end{bmatrix}$
3	Entertainment	$\begin{bmatrix} 2 & -2 \end{bmatrix}$
4	Commercial	$\begin{bmatrix} -6 & 0 \end{bmatrix}$

    iii. Finally, the topic with the highest similarity is given as the output k i.e.

    $$k = \arg\max_{i \in [4]} s_i$$

    where $[4] = \{1, 2, 3, 4\}$, and the "arg max" function is such that $s_k = \max\{s_1, s_2, s_3, s_4\}$.

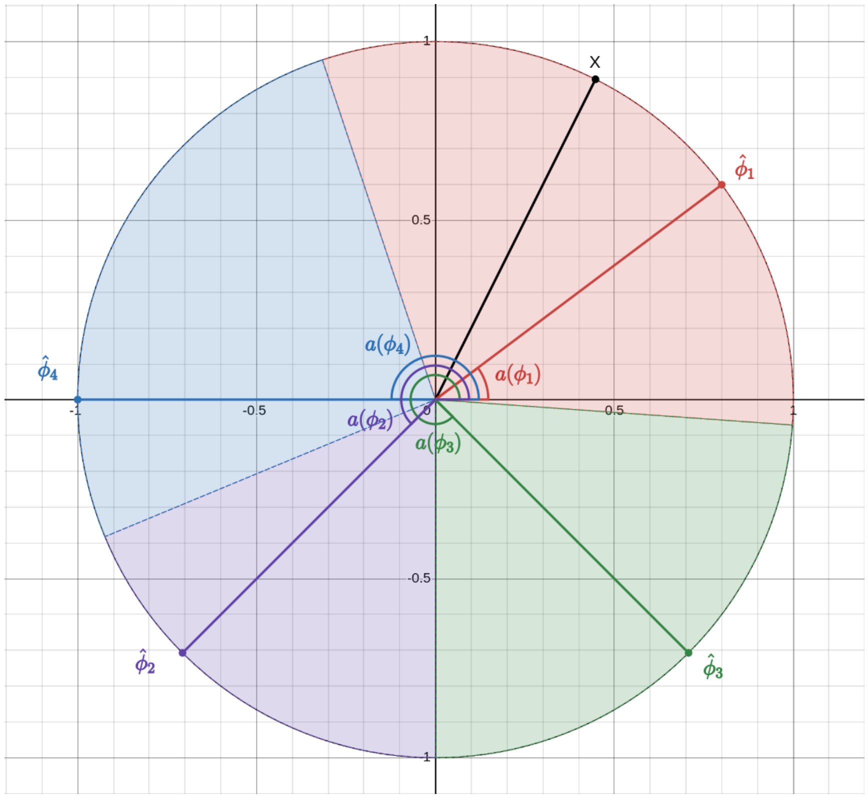
    Say the document given to the text summarization algorithm is random, and the encoding of the document is a random variable $X = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \end{bmatrix}$, where α is a Uniform random variable such that $\alpha \sim \text{Uniform}(0, 2\pi)$. Let the output of the algorithm be I. Find the PMF of I.

Solution: *(Random Variables and Vectors)*
Observe that X is uniformly distributed on the circle of radius 1. This is because $\|X\| = \sqrt{\cos^2(\alpha) + \sin^2(\alpha)} = 1$, and the fact that α is uniformly distributed over the interval $[0, 2\pi]$. Moreover, for each $i \in \{1, 2, 3, 4\}$, observe that

$$s_i = \cos(\theta_i).$$

Because s_i only depends on the angle between X and encoding ϕ_i, we infer that s_i does not depend on the norm of ϕ_i. We use this to obtain the PMF.

$$\begin{aligned} P(I = i) &= P(s_i = \max\{s_1, s_2, s_3, s_4\}) \\ &= P\left(\cos(\theta_i) = \max\{\cos(\theta_1), \cos(\theta_2), \cos(\theta_3), \cos(\theta_4)\}\right) \end{aligned}$$



This probability is hard to compute, so we try to use conditional probabilities to simplify the calculation. First, we define $a(v)$ as the counter-clockwise angle made by the 2-dimensional vector v with the X-axis (see figure above, showing the unit vectors and angles corresponding to each encoding). Observe that the domain of α can be partitioned into 4 regions:

1. $R_{1,4} = (a(\phi_1), a(\phi_4)) \implies P(\alpha \in R_{1,4}) = \frac{a(\phi_4)-a(\phi_1)}{2\pi} = 0.398$

2. $R_{4,2} = (a(\phi_4), a(\phi_2)) \implies P(\alpha \in R_{4,2}) = \frac{a(\phi_2)-a(\phi_4)}{2\pi} = 0.125$

3. $R_{2,3} = (a(\phi_3), a(\phi_3)) \implies P(\alpha \in R_{2,3}) = \frac{a(\phi_3)-a(\phi_2)}{2\pi} = 0.25$

4. $R_{3,1} = (a(\phi_3), 2\pi + a(\phi_1)) \implies P(\alpha \in R_{3,1}) = \frac{2\pi+a(\phi_1)-a(\phi_3)}{2\pi} = 0.227$

Since ϕ_i for all $i = 1, 2, 3, 4$ are known, $a(\phi_i)$ for all i are calculated as above. Now, observe that if $\alpha \in R_{i,j}$, then $I = i$ or $I = j$ (all other possible values have probability 0). Using this, we get

$$
\begin{aligned}
P(I=1) &= P(I=1 \text{ and } \alpha \in R_{1,4}) + P(I=1 \text{ and } \alpha \in R_{4,2}) \\
&\quad + P(I=1 \text{ and } \alpha \in R_{2,3}) + P(I=1 \text{ and } \alpha \in R_{3,1}) \\
&= P(I=1 \text{ and } \alpha \in R_{1,4}) + P(I=1 \text{ and } \alpha \in R_{3,1}) \\
&= P(I=1|\alpha \in R_{1,4})P(\alpha \in R_{1,4}) + P(I=1|\alpha \in R_{3,1})P(\alpha \in R_{3,1}) \\
&= \frac{1}{2}P(\alpha \in R_{1,4}) + \frac{1}{2}P(\alpha \in R_{3,1}) = \frac{0.398+0.227}{2} = 0.3125
\end{aligned}
$$

Similarly, we get

$$P(I=2) = \frac{1}{2}P(\alpha \in R_{4,2}) + \frac{1}{2}P(\alpha \in R_{2,3}) = \frac{0.125+0.25}{2} = 0.1875$$
$$P(I=3) = \frac{1}{2}P(\alpha \in R_{2,3}) + \frac{1}{2}P(\alpha \in R_{3,1}) = \frac{0.25+0.227}{2} = 0.2385$$
$$P(I=4) = \frac{1}{2}P(\alpha \in R_{1,4}) + \frac{1}{2}P(\alpha \in R_{4,2}) = \frac{0.398+0.125}{2} = 0.2615$$

df	$P(T < t) = 0.975$	$P(T < t) = 0.95$
1	12.706	6.314
2	4.303	2.920
3	3.182	2.353
4	2.776	2.132
5	2.571	2.015
6	2.447	1.943
7	2.365	1.895
8	2.306	1.860
9	2.262	1.833
10	2.228	1.812
11	2.201	1.796

Table 1: T-critical values for specified probabilities and degrees of freedom.