LUMS
A Not-for-Profit University

School of Science and Engineering

AI 501 Mathematics for Artificial Intelligence

ASSIGNMENT 1

Due Date: 11 am, Saturday, October 5, 2024.
Format: 11 problems for a total of 100 marks
Instructions:

- You are allowed to collaborate with your peers but copying your colleague's solution is strictly prohibited. This is not a group assignment. Each student must submit his/her own assignment.
- Solve the assignment on blank A4 sheets and staple them before submitting.
- Submit in-class or in the dropbox labeled AI-501 outside the instructor's office.
- **Write your name and roll no. on the first page.**
- Feel free to contact the instructor or the teaching assistants if you have any concerns.

- You represent the most competent individuals in the country, do not let plagiarism come in between your learning. In case any instance of plagiarism is detected, the disciplinary case will be dealt with according to the university's rules and regulations.
- We require you to acknowledge any use or contributions from generative AI tools. Include the following statement to acknowledge the use of AI where applicable.

  I have used [insert Tool Name] to [write, generate, plot or compute; explain specific use of generative AI] [number of times].

Problem 1 (6 marks)

Farmer's Dilemma - Predicting Crop Yield

A farmer in a rural village has been struggling to predict his wheat crop yield every season. He believes that the amount of rainfall and the number of sunny days during the growing season significantly affect his harvest. To better understand this relationship, he collects data over 8 seasons, recording the amount of rainfall (in inches) and the number of sunny days, along with the corresponding wheat yield (in tons).

Season	Rainfall (in inches)	Sunny Days	Wheat Yield (in tons)
1	12	50	6.0
2	15	55	7.0
3	10	48	5.0
4	18	60	8.0
5	20	65	9.0
6	14	53	6.5
7	16	57	7.5
8	17	62	8.2

(a) Calculate the Pearson correlation coefficient between the amount of rainfall and wheat yield. [5 marks]

(b) Calculate the Pearson correlation coefficient between the number of sunny days and wheat yield. Based on your findings, which factor seems to have a stronger linear relationship with wheat yield? What recommendations would you give the farmer? [5 marks]

Problem 2 (8 marks)

Vector Space and Gram-Schmidt Orthogonalization in Engineering Stability

Imagine you are part of an engineering team working on stabilizing the foundation of a futuristic building. The structural integrity of the building depends on three support columns represented by vectors in $\mathbb{R}^3$. These columns are initially described by the vectors:

$$\mathbf{v_1} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \quad \mathbf{v_2} = \begin{pmatrix} 4 \\ -2 \\ 2 \end{pmatrix}, \quad \mathbf{v_3} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

However, upon inspection, you realize that some of the columns may be providing redundant support, leading to inefficiencies and potential instability.

(a) **(3 points)** Prove that the vectors $\mathbf{v_1}$ and $\mathbf{v_2}$ are linearly dependent. Explain the significance of this linear dependence in the context of the structural stability of the building.

(b) **(5 points)** Apply the Gram-Schmidt orthogonalization process to the set $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$. Discuss how the process helps to break the linear dependence and provide new, orthogonal directions that improve the structural support for the building.

Problem 3 (10 marks)

The Mystery of the Hidden Dimensions

In the ancient kingdom of AI501, the wise King had three legendary artifacts: the Scepter of Independence, the Crown of Basis, and the Shield of Rank. These powerful artifacts were said to protect the kingdom from chaos by maintaining the balance of the royal Vector Space, V. One day, a sinister sorcerer, known as Ibrahim, stole the artifacts and scattered them across different subspaces, threatening the very structure of the kingdom.

Desperate to restore order, the King called upon his most trusted advisors, the students of AI501, to retrieve the artifacts by solving a series of puzzles. The students began by studying the kingdom's primary defense: a mysterious matrix A, which governs the stability of the entire Vector Space. The matrix is of size 4×4, and its columns represent different vectors that describe the forces protecting the kingdom. The matrix A is given by

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 1 & 0 & 1 & 0 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

The puzzles were as follows:

- The **Scepter of Independence** can only be retrieved if the students identify whether the columns of matrix A are linearly independent. Can the students prove the linear independence of the columns of A? If not, what is the dimension of the subspace they span? [5 marks]
- The **Crown of Basis** was hidden in a secret subspace of the Vector Space V. The dimension of this subspace was equal to the rank of matrix A. The students must determine the rank of matrix A to find the exact location of the subspace. What is the rank of matrix A? [2.5 marks]
- Finally, to retrieve the **Shield of Rank**, the students need to discover a basis for the column space of matrix A. How can the students determine the basis for this subspace using the columns of matrix A? [2.5 marks]

The King depends on the students to solve these puzzles, for, without the artifacts, the kingdom of Mathematics of AI will crumble into chaos. Will the students be able to determine the linear independence, rank, and basis of matrix A in time to save the kingdom?

Problem 4 (8 marks)

Story Similarity and Distance Measures

Suppose you are analyzing stories, each represented by a vector with 3 elements describing the following characteristics: drama, comedy, and science fiction. You want to compare pairs of stories using these dimensions, and are given the following examples:

$$v_{\text{Story A}} = \begin{bmatrix} 0.9 \\ 0.05 \\ -0.1 \end{bmatrix}, \quad v_{\text{Story B}} = \begin{bmatrix} 0.8 \\ 0.02 \\ 0.78 \end{bmatrix}, \quad v_{\text{Story C}} = \begin{bmatrix} 0.6 \\ 0.85 \\ 0.89 \end{bmatrix}$$

(a) **(2 points)** Compute the Chebyshev distance between Story A and Story B. What shortcoming might arise from using this distance for this comparison?

(b) **(2 points)** Compute the Manhattan distance between Story A and Story B. Discuss why Manhattan distance may not be a good metric for this comparison.

(c) **(2 points)** Compute the Euclidean distance between Story A and Story B. Why might this not be the most appropriate metric in this case?

(d) **(2 points)** Using Cosine Similarity, determine which of Story B or Story C is closer to Story A. What feature/characteristic would you attribute this closeness to? Be sure to show your working.

Problem 5 (10 marks)

Matrix Inverses: A Rigorous Investigation

Let A be an $m \times n$ matrix, and let A have both a left inverse L and a right inverse R. Prove the following statements:

(a) **(3 points)** Show that A must be a square matrix ($m = n$). Specifically, prove that if A has a left inverse L and a right inverse R, then m must equal n.

(b) **(3 points)** Prove that the left inverse L and the right inverse R of A are equal. In other words, show that $L = R$ if both L and R exist.

(c) **(4 points)** Consider a matrix A where $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$.

(i) Compute the left and right inverses of A if they exist.

(ii) Use your results to demonstrate that A does not satisfy the conditions needed for having both a left and right inverse.

(iii) Explain why the specific matrix A you worked with in part (i) does or does not fit the general conditions for having left and right inverses.

Problem 6 (10 marks)

The Adventure of the Regression Model

In a small town, there is a local bakery that wants to predict its daily sales based on the number of customers visiting the store. The bakery owner, Ms. Baker, has collected data from the past month, recording the number of customers and the corresponding daily sales in dollars. Ms. Baker is keen to understand the relationship between the number of customers and the sales using a univariate linear regression model. She needs your help to apply this model to her data and evaluate its performance.

(a) **(5 points) Conceptual Understanding:**
Ms. Baker wants to understand the concept of residuals and how they reflect the quality of the model. Describe what residuals are, how they are calculated, and why minimizing the residual sum of squares is important in fitting the model. Provide an intuitive explanation of how the residuals impact the fit of the regression line.

(b) **(5 points) Practical Application:**
Ms. Baker provides you with the following dataset:

Number of Customers(x)	Daily Sales(y)
10	120
15	180
20	240
25	300
30	360

Using the data, calculate the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for the linear regression model $y = \beta_0 + \beta_1 x$. The equations you need are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Show all steps involved in your calculations.

Compute the residual sum of squares (RSS) for the model you obtained. Evaluate the goodness of fit by calculating the total sum of squares (TSS) and the coefficient of determination (R^2). Show all your calculations and explain what the R^2 value indicates about the model's performance.

Problem 7 (10 marks)

Modified Ridge Regression Using Projection

Given a dataset with N observations and d features, with A being the design matrix and y the response vector, the modified ridge regression cost function is:

$$L(w) = \frac{b_m}{2}\|Aw - y\|^2 + \frac{\lambda b_r}{2}\|w\|^2,$$

where b_m is the weight for the mean squared error term, b_r is the weight for the regularization term, and λ is the regularization parameter.

(a) **(6 points)** Derive the closed-form solution for the weight vector w in this modified ridge regression problem using the concept of projection. Also, specify the shape of each matrix involved in the derivation and the final shape of the weight vector w.

(b) **(4 points)** Discuss how varying b_m and b_r independently affects the bias-variance trade-off in the model.

Problem 8 (12 marks)

The Dilemma of the Automated Drone

You are developing an automated drone system to monitor a secure area for potential intrusions. The drone uses a machine learning model to classify detected movements as either "Intruder" or "Non-Intruder." The performance of your model is evaluated using the following confusion matrix:

	Actual Intruder (Positive)	Actual Non-Intruder (Negative)
Predicted Intruder	a_{11}	a_{12}
Predicted Non-Intruder	a_{21}	a_{22}

(a) **(3 points)** Suppose you are concerned about missing actual intruders (i.e., failing to detect real threats). Which evaluation metric would you prioritize to minimize this issue and why? Explain the importance of this metric in the context of your application.

(b) **(3 points)** If you had to choose between maximizing precision or maximizing specificity, which would you choose for this drone system and why? Provide a detailed explanation of how each metric impacts the performance of the system in this scenario.

(c) **(6 points)** Given the confusion matrix representation above, derive expressions for the following metrics:

- Accuracy
- Precision
- Recall
- Specificity

Additionally, answer the following:

(i) Given that the model never predicts the negative class, and the Accuracy is equal to the F1 score, what should be the value of a_{12}?

Problem 9 (10 marks)

The Tale of the Expanding Universe

In a distant mathematical universe, a group of interdimensional explorers discover a peculiar phenomenon: as they traverse from one dimension to another, their perception of distance and volume within a *unit ball* dramatically shifts.

Hint: The volume V_n of a ball of radius R in n-dimensional space is given by:

$$V_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} R^n,$$

where $\Gamma(x)$ is the Gamma function, and for positive integers x, $\Gamma(x) = (x-1)!$.

(a) **(4 points)** Upon entering the universe, they encounter a spherical planet which has a unit radius in their familiar 3D world. They observe that the volume of the planet (which corresponds to the volume of a unit ball in 3 dimensions) is approximately 4.19 cubic units. The explorers, intrigued by how this might change in higher dimensions, decide to explore unit balls in various dimensions n. Use the given formula for V_n to calculate the volume of the unit ball in $n = 10$ and $n = 100$ dimensions.

(b) **(3 points)** The explorers then notice something strange: as they increase the dimensionality beyond 10, the majority of the volume seems to *disappear* from the center of the unit ball and shifts to its surface. Explain this phenomenon mathematically, and show how the proportion of the volume near the center diminishes as n increases.
Hint: Analyse the fraction of the volume of unit radius ball that lies inside a shell between unit radius ball and a ball of radius $0 < r_0 \leq 1$, and show that for most of the volume lies inside the shell even when $r_0 \to 1$.

(c) **(3 points)** Realizing that this loss of volume at the center could have serious implications for navigation, the explorers decide to quantify their findings. They calculate the distance from the origin at which half the volume of the unit ball resides. Find and explain how this distance changes as the number of dimensions increases, particularly comparing $n = 3$, $n = 10$, and $n = 100$.

Problem 10 (6 marks)

Linear, Affine and Convex Combination We require you to plot the linear, affine, and convex combinations of vectors in $\mathbf{R}^3$. Write a Python script to generate three random vectors v_1, v_2, v_3, where each coordinate of these vectors is drawn from a standard normal distribution. We require you to submit both the code and the resulting plots.

Problem 11 (10 marks)

Properties of the p-Norm

Let $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and define the p-norm for $1 \le p < \infty$ as:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

(a) **(4 points)** Prove that for any vector $\mathbf{x} \in \mathbb{R}^n$, the p-norm satisfies the following properties:

- $\|\mathbf{x}\|_p = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
- $\|\alpha\mathbf{x}\|_p = |\alpha|\|\mathbf{x}\|_p$ for any scalar $\alpha \in \mathbb{R}$.
- $\|\mathbf{x} + \mathbf{y}\|_p \le \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (triangle inequality).

Use Minkowski's inequality to prove the triangle inequality:

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{\frac{1}{p}} \le \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}, \forall p \ge 1.$$

(b) **(3 points)** Show that the p-norm is monotonically decreasing with respect to p, i.e., for $1 \le p \le q < \infty$, prove that:

$$\|\mathbf{x}\|_p \ge \|\mathbf{x}\|_q.$$

(c) **(3 points)** We want to investigate the behavior of the p-norm as $p \to \infty$. Prove the following:

$$\lim_{p \to \infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \max_i |x_i|.$$

— End of Assignment —