LUMS
A Not-for-Profit University

School of Science and Engineering

AI 501 Mathematics for Artificial Intelligence

ASSIGNMENT 2 – SOLUTIONS

Due Date: 5 pm, Thursday, October 24, 2024.
Format: 07 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 (10 marks)

A transformation matrix $\mathbf{T}$ is shown below:

$$\mathbf{T} = \begin{bmatrix} 2 & -3 \\ 0 & 5 \end{bmatrix}.$$

(a) [**2 marks**] Given a position vector $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, explain what happens to $\mathbf{A}$ when it undergoes a transformation under $\mathbf{T}$?

(b) [**2 marks**] Using eigenvalue decomposition, represent $\mathbf{T}$ in terms of eigenvectors matrix, eigenvalues matrix and inverse eigenvectors matrix.

(c) [**1 marks**] State the eigenvalues of $\mathbf{T}$.

(d) [**1 marks**] State the associated eigenvectors of $\mathbf{T}$.

(e) [**4 marks**] By making a plot, show what happens to the eigenvectors when they undergo a transformation under $\mathbf{T}$. Include eigenvalues in your interpretation and contrast the result of this eigenvector transformation with that of position vector $\mathbf{A}$.

Solution:

(a) transforms to (-1,5)

(b) $\begin{bmatrix} 2 & -3 \\ 0 & 5 \end{bmatrix} = \mathrm{T} = \begin{bmatrix} 1 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \sqrt{2} \end{bmatrix}$

(c) eigenvalues: 2, 5

(d) eigenvectors: $\begin{bmatrix} 1 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$

(e) eigenvectors unchanged, scaling by eigenvalues shown

Problem 2 (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 either using the concept of projection or gradient of the cost function. 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. You can build on the polynomial regression formulation to answer this question.

Solution:

(a) Cost Function and Closed-Form Solution:

Rewrite the cost function as:

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

Expand:

$$L(w) = \frac{b_m}{2}\left(w^\top A^\top A w - 2y^\top A w + y^\top y\right) + \frac{\lambda b_r}{2} w^\top w.$$

Combine terms:

$$L(w) = \frac{b_m}{2} w^\top A^\top A w + \frac{\lambda b_r}{2} w^\top w - b_m y^\top A w + \text{constant terms}.$$

Projection Approach:

The ridge regression problem can be viewed as finding w such that y is projected onto the column space of A with regularization. This is equivalent to solving the normal equations:

$$(b_m A^\top A + \lambda b_r I)w = b_m A^\top y,$$

where λ is the regularization parameter and b_r is the weight applied to the regularization term.

Here: - A is $N \times d$ - $A^\top A$ is $d \times d$ - I is $d \times d$ (identity matrix) - $b_m A^\top y$ is $d \times 1$ - w is $d \times 1$

Solving for w:

$$w = (b_m A^\top A + \lambda b_r I)^{-1} b_m A^\top y.$$

Shape of Matrices: - A: $N \times d$ - $A^\top A$: $d \times d$ - I: $d \times d$ - $b_m A^\top y$: $d \times 1$ - w: $d \times 1$

Final Solution:

The weight vector w is:

$$w = (b_m A^\top A + \lambda b_r I)^{-1} b_m A^\top y.$$

(b) Bias-Variance Trade-Off:

The parameters b_m and b_r in the modified ridge regression cost function serve to weight the mean squared error (MSE) and the regularization term, respectively. Their values can significantly affect the model's performance:

- **Varying b_m:** Increasing b_m enhances the model's sensitivity to fitting the training data by emphasizing the reduction of MSE. This can lead to a model that fits the training data more closely, potentially decreasing bias. However, it can also increase the model's variance as it may start capturing noise as if it were a true signal, leading to overfitting.
- **Varying b_r:** Increasing b_r emphasizes the importance of the regularization term in the cost function. This can help prevent the model from becoming overly complex and fitting the training data too closely, which is especially useful when there is a risk of overfitting. A higher b_r typically increases bias as the model becomes less flexible and does not fit the training data as closely, but can reduce variance, making the model more robust to variations in the training data.

Problem 3 (15 marks)

As a data analyst for a major film studio, you've been tasked with developing a model to predict the box office success of upcoming movies. You believe that the movie's budget and the number of weeks in production are key factors influencing its performance.

Using a dataset of past movies, their budgets, production times, and box office earnings, create a predictive model that will help your studio make more informed decisions about which projects to green-light.

(a) Express the relationship between the budget, weeks in production, and box office earnings as a least-squares problem. Define the appropriate matrix notation and write down the least-squares objective function.

(b) Using the provided data, compute the coefficients of the linear regression model analytically.

Movie Title	Budget (\$ in Million)	Production Time (Weeks)	Box Office Revenue (\$ in Million)
A	10	25	30
B	5	20	15
C	10	20	25
D	5	15	10
E	15	35	35

Solution: We want to model the relationship between the budget and weeks in production of a movie, and its box office earnings using a multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where:

- y is the box office earnings,
- β_0 is the intercept,
- β_1 is the coefficient for budget,
- x_1 is the budget,
- β_2 is the coefficient for weeks in production,
- x_2 is the weeks in production,

We can express this problem in matrix notation as:

$$Y = \beta_0 + \beta_{1,2}^T X$$
$$Y = \beta^T X$$

where:

- Y is the vector of box office earnings,
- X is the matrix of independent variables (including a column of 1's for the intercept),
- β is the vector of coefficients $[\beta_0 \ \beta_1 \ \beta_2]$

The least-squares problem seeks to minimize the sum of squared residuals:

$$\min_{\beta} \mathcal{L}(\beta) = \frac{1}{2} \sum_{i=1}^{n} (y_i - \beta^T x_i)^2; \quad n = 1 \ \ till \ \ 5$$

The analytical solution for the coefficients β is given by:

$$\beta = (X^T X)^{-1} X^T Y;$$

$$Y = \begin{bmatrix} 30 \\ 15 \\ 25 \\ 10 \\ 35 \end{bmatrix}; \ X = \begin{bmatrix} 1 & 10 & 25 \\ 1 & 5 & 20 \\ 1 & 10 & 20 \\ 1 & 5 & 15 \\ 1 & 15 & 35 \end{bmatrix}; \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$X^TX : \begin{bmatrix} 5 & 45 & 115 \\ 45 & 475 & 1150 \\ 115 & 1150 & 2875 \end{bmatrix}$$

$$X^TY : \begin{bmatrix} 115 \\ 1200 \\ 2925 \end{bmatrix}$$

$$(X^TX)^{-1} : \frac{1}{det(X^TX)} adj(X^TX) = \begin{bmatrix} 3 & 1/5 & -1/5 \\ 1/5 & 2/25 & -1/25 \\ -1/5 & -1/25 & 14/575 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 0 \\ 2 \\ 5/23 \end{bmatrix}$$

Principal Component Analysis Overview

PCA is a powerful technique used for dimensionality reduction, enabling the projection of high-dimensional data onto a lower-dimensional subspace while preserving as much variance as possible.

- The process begins with mean subtraction, where we compute the mean of the dataset and subtract it from each data point, resulting in a dataset centered around zero.
- The next step is standardization, where each data point is divided by the standard deviation of the entire dataset for that dimension, transforming the data into a unit-free format with a variance of 1 along each axis. This ensures that subsequent analysis is not skewed by differences in scale among the variables.
- This is followed by the construction of the covariance matrix of the standardized data. We find its eigenvalues and corresponding eigenvectors. The eigenvalues indicate the amount of variance captured by their corresponding eigenvectors, and the eigenvectors are scaled according to the magnitude of their eigenvalues, creating a set of orthogonal basis vectors that represent the principal components. The principal subspace corresponding to the largest eigenvector captures the most variance in the data.
- We can project any new data point onto the principal subspace by first standardizing it using the mean and standard deviation of the training data. The projection yields the coordinates in the context of the standardized dataset.
- Finally, to transform these projections back in the original data space, we must undo the standardization by multiplying the standardized projections by the standard deviation and then adding the mean back. This allows us to visualize and interpret the projected data points in relation to the original dataset.

These steps are also illustrated in the figure below:

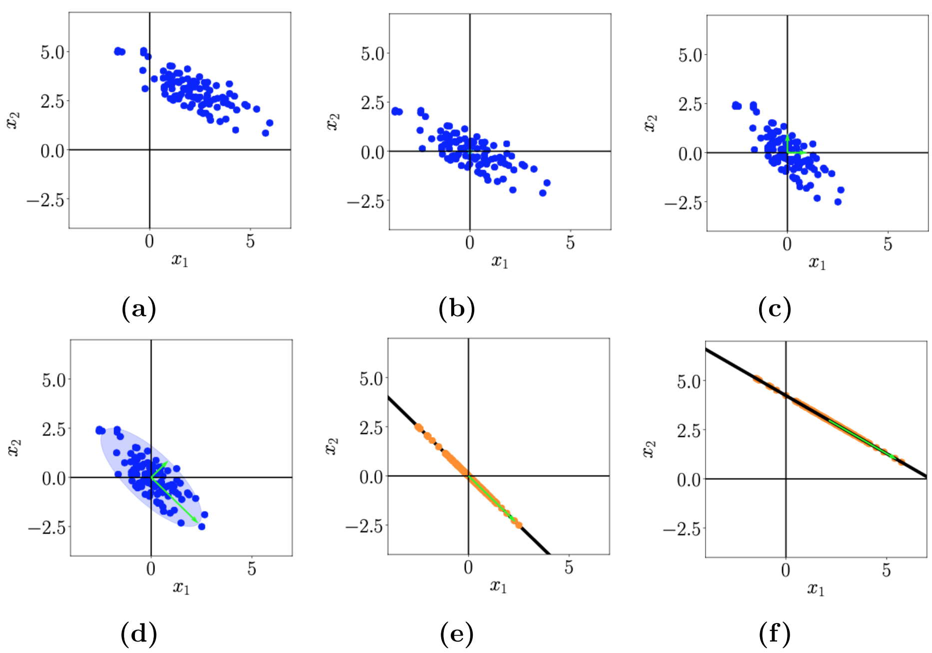

Figure 1: (a) Original dataset. (b) Centering by subtracting mean. (c) Dividing by standard deviation. (d) Compute eigenvalues and eigenvectors. (e) Project data onto principal subspace. (f) Undo standardization, move projected data in original space.
Figure source: Mathematics for Machine Learning by Deisenroth, pg. 337.

Problem 4 (10 marks)

Given the following matrix $\mathbf{X}$,

$$\mathbf{X} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \\ 7 & 8 \\ 6 & 5 \\ 9 & 3 \\ 11 & 10 \\ 12 & 9 \end{bmatrix}$$

for $n = 7$ and $d = 2$.

We will make use of Principal Component Analysis (PCA) to reduce the dimensions of the matrix $\mathbf{X}$ from $d = 2$ to $d = 1$ by carrying out the following steps:

(a) [**2 marks**] Plot the data points on a 2-dimensional plane.

(b) [**4 marks**] Compute the principal components using the procedure taught in class (refer to the slides) and plot them as well.

(c) [**4 marks**] Now, project the original data matrix $\mathbf{X}$ onto its first principal component and plot on a 1-dimensional number line.

Problem 5 (20 marks)

Welcome to the grand AI Adventure, where intellect meets intrigue! As part of this thrilling scavenger hunt, you'll traverse the campus, unraveling clues that lead you closer to a hidden treasure. Each solved riddle brings you a step nearer to the final prize.

Your journey culminates in the ultimate challenge — a puzzle that involves the enigmatic matrix Z. Gather your wits, and let's dive into the clues that will lead you to victory!

You are given an m x n matrix Z, described by five captivating facts:

1. **Hidden Dimensions:** The matrix's reflection in a mirror, $Z^T Z$, holds secrets of two hidden dimensions, 9 and 4.
2. **The Guiding Vector:** One of these dimensions has a cosmic companion — the normalized eigenvector $e_1 = \frac{1}{\sqrt{5}} [1\ 2]^T$ — a beacon pointing towards the truth.
3. **The Cosmic Canvas:** Z is painted with a cosmic brush and within its strokes lies the left singular matrix U. The third stroke, $u_3 = [\frac{2}{3}\ -\frac{1}{3}\ -\frac{2}{3}]^T$, holds a vital piece of the puzzle.
4. **The Cosmic Dance:** Two celestial bodies, v_1 and v_2, orbit around Z, leaving trails of stardust: $Zv_1 = \frac{1}{\sqrt{5}} [5\ 2\ 4]^T$ and $Zv_2 = \frac{2}{\sqrt{5}} [0\ 2\ -1]^T$ (v_1 and v_2 being the first and second column of the matrix V, respectively)
5. **The Celestial Key:** The hidden treasure lies on a floor determined by the cosmic energy of Σ_{22}.

The fate of the universe rests in your hands. Can you crack the code and unveil the mysteries of the matrix?

Your Mission:

(a) [**2 marks**] **Unveil the Matrix:** Determine the dimensions of this cosmic enigma, Z.

(b) [**6 marks**] **Paint the Cosmic Canvas:** Calculate the left singular matrix, U, to reveal the hidden patterns.

(c) [**6 marks**] **Trace the Celestial Dance:** Determine the right singular matrix, V, to map the orbits of v_1 and v_2.

(d) [**4 marks**] **Unleash the Cosmic Energy:** Unveil the singular values Σ to understand their significance.

(e) [**2 marks**] **Find the Celestial Treasure:** Decipher the cosmic energy of Σ_{22} to determine the floor number where the prize is hidden.

Solution:

(a) Z is a $m \times n$ matrix.
Dimensions of u_3 must be $m \times 1$, hence $m = 3$.
For Zv_1 to have dimensions 3×1, the dimensions should be $Z_{3\times n}(v_1)_{n\times 1}$, hence $n = 2$.
Size of Z is a 3×2

(b)

$$U = [u_1\ u_2\ u_3]$$

$$u_1 = \frac{1}{\sqrt{\lambda_1}} Zv_1; \lambda_1 = 9 \text{ and } Zv_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 5 & 2 & 4 \end{bmatrix}^T$$

$$u_2 = \frac{1}{\sqrt{\lambda_2}} Zv_2; \lambda_2 = 4 \text{ and } Zv_2 = \frac{2}{\sqrt{5}} \begin{bmatrix} 0 & 2 & -1 \end{bmatrix}^T$$

$$u_3 = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \end{bmatrix}$$

Hence:

$$U = \begin{bmatrix} \frac{5}{3\sqrt{5}} & 0 & \frac{2}{3} \\ \frac{2}{3\sqrt{5}} & \frac{2}{\sqrt{5}} & -\frac{1}{3} \\ \frac{4}{3\sqrt{5}} & -\frac{1}{\sqrt{5}} & -\frac{2}{3} \end{bmatrix}$$

(c)

$$V = [v_1 \; v_2]$$

$$v_1 = e_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

We know that $v_1^T v_2 = 0$ as they are both orthogonal. So

$$\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} v_{2a} \\ v_{2b} \end{bmatrix} = 0$$

$$v_{2a} = -2v_{2b}$$

Both vectors should be normalized so

$$v_{2a}^2 + v_{2b}^2 = 1$$

Solve these two equations simultaneously to get:

$$V = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

(d)

$$\Sigma = \begin{bmatrix} \sqrt{9} & 0 \\ 0 & \sqrt{4} \\ 0 & 0 \end{bmatrix}$$

(e) 2^{nd} floor

Problem 6 (10 marks)

Let matrix $M \in \mathbb{R}^{m \times m}$ and $M^H = M$.

(a) **[3 marks]** Prove that all the eigenvalues of M are real valued. [Hint: Consider the inner product $\langle M\mathbf{x}, \mathbf{x}\rangle$ for an eigenvector x of M.]

(b) **[2 marks]** Prove that u and v are orthogonal, given that u and v are eigenvectors to distinct eigenvalues of M.

(c) **[3 marks]** A matrix M is positive semi-definite if $\langle M\mathbf{x}, \mathbf{x}\rangle = \left(M\mathbf{x}\right)^T \mathbf{x} \geq 0$ for all non-zero vectors $\mathbf{x}$. Suppose M is positive semi-definite.

Prove that all its eigenvalues are non-negative.

(d) **[2 marks]** Prove that if M is a projection matrix (i.e., $M^2 = M$), all its eigenvalues are either 0 or 1.

Solution:

(a)

$$\langle Mx, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle$$

and

$$\langle Mx, x\rangle = \langle x, M^H x\rangle$$
$$\langle x, Mx\rangle = \langle x, \lambda x\rangle = \bar{\lambda}\langle x, x\rangle$$

But $\langle Mx, x\rangle = \langle x, Mx\rangle$ since $M = M^H$, thus we must have

$$\lambda\langle x, x\rangle = \bar{\lambda}\langle x, x\rangle$$

hence $\lambda = \bar{\lambda}$, which is only possible if λ is real.

(b) Let λ be the eigenvalue to u and μ be the eigenvalue to v. We know from (a) that the eigenvalues are real. Then

$$\langle Mu, v\rangle = \langle \lambda u, v\rangle = \lambda\langle u, v\rangle$$

and

$$\langle u, My\rangle = \langle u, \mu v\rangle = \mu\langle u, v\rangle$$

Since $M = M^H$, we have $\langle Mu, v\rangle = \langle u, Mv\rangle$ and therefore $\lambda\langle u, v\rangle = \mu\langle u, v\rangle$ or $(\lambda - \mu)\langle u, v\rangle = 0$. Since $\lambda \neq \mu$ by assumption it follows that we must have $\langle u, v\rangle = 0$, that means u and v are orthogonal.

(c) For an eigenvector $\mathbf{x}$ corresponding to eigenvalue λ:

$$\langle M\mathbf{x}, \mathbf{x}\rangle = \langle \lambda\mathbf{x}, \mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle.$$

Since $\langle M\mathbf{x}, \mathbf{x}\rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x}\rangle > 0$, it follows that $\lambda \geq 0$.

(d) Let λ be an eigenvalue of M with corresponding eigenvector $\mathbf{x}$, so $M\mathbf{x} = \lambda\mathbf{x}$. Applying the projection property $M^2 = M$:

$$M^2\mathbf{x} = M(\lambda\mathbf{x}) = \lambda M\mathbf{x} = \lambda^2\mathbf{x}.$$

But $M^2\mathbf{x} = M\mathbf{x} = \lambda\mathbf{x}$, so:

$$\lambda^2\mathbf{x} = \lambda\mathbf{x}.$$

Since $\mathbf{x} \neq 0$, it follows that $\lambda^2 = \lambda$, giving $\lambda(\lambda - 1) = 0$. Thus, $\lambda = 0$ or $\lambda = 1$.

Problem 7 (25 marks)

Use MATLAB or Python to answer this question. Share m-file or Colab notebook as submission. An RGB image consists of 3 channels (Red, Green, Blue).

(a) [**1 mark**] What is the vector size of a horizontally flattened 32×32 RGB image?

(b) [**1 mark**] What are the dimensions of a matrix that consists of 50 such horizontally flattened images that are vertically stacked on top of each other?

(c) [**15 marks**] The dataset attached `(click to download)` has 50 horizontally flattened images, each of size 32×32, vertically stacked on each other. By treating each pixel as a feature, compute Euclidean distance of the test image `(click to download)` with all 50 images. Plot image-number on x-axis and distance from test image on y-axis.

(d) [**8 marks**] By sorting the distance in ascending order and reading the corresponding labels in the data file, (and also by your own judgement of the celebrity in the test image and its 05 nearest neighbours), explain whether the test image is correctly classified if 5 nearest neighbors are considered. Specify labels of the 05 nearest neighbours.

Solution: (a) 1×3072
(b) 50×3072
(c)

(d) 5 nearest neighbours: image numbers 43, 6, 29, 41, 1
corresponding labels: ['fawad khan'], ['ahsan khan'], ['bilawal bhutto'], ['fawad khan'], ['morgan freeman']

— End of Assignment —