LUMS

A Not-for-Profit University

School of Science and Engineering

AI 501 Mathematics for Artificial Intelligence

ASSIGNMENT 2

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.

Problem 1 (10 marks)

A transformation matrix **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 **A** when it undergoes a transformation under **T**?

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

(c) [**1 marks**] State the eigenvalues of **T**.

(d) [**1 marks**] State the associated eigenvectors of **T**.

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

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.

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

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.

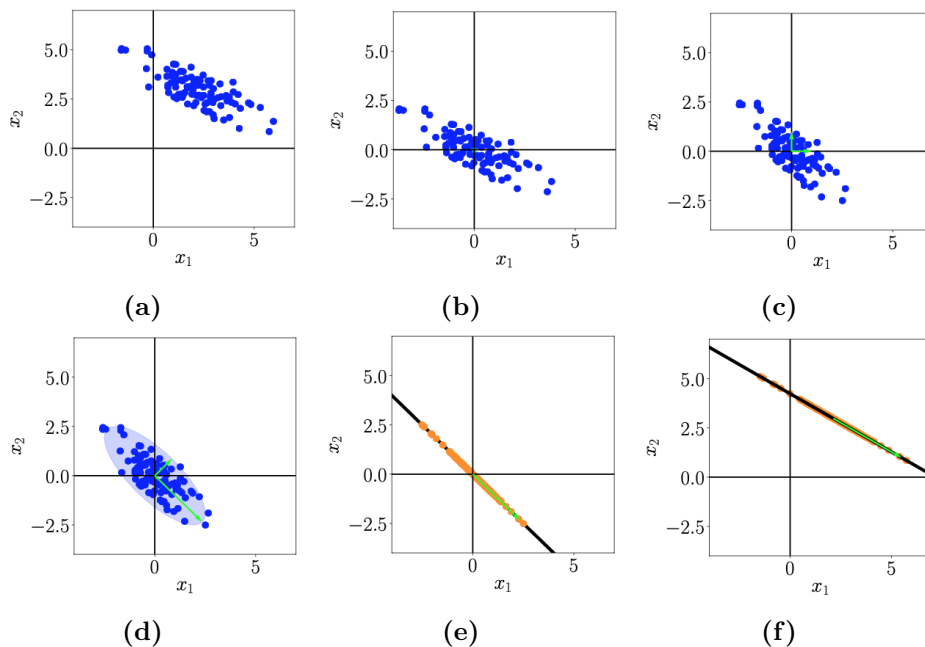These steps are also illustrated in the figure below:

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.

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.

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.

— End of Assignment —