Applied Linear Algebra in Data Analaysis
SVD & Dimensionality Reduction Assignment

Marks: 14

1. For a square $\mathbf{A} \in \mathbb{R}^{n\times n}$, the SVD tells us how a unit sphere in $\mathbb{R}^n$ is distorted by the linear transformation performed by $\mathbf{A}$. This degree of distortion can be quantified using the singular values of $\mathbf{A}$, which is the 2-norm *condition number*,

$$\kappa = \frac{\sigma_1}{\sigma_n}$$

   (a) Explain why $\kappa \geq 1$? [**Marks: 1**]

   (b) What is condition number of a singular matrix? [**Marks: 1**]

   (c) If $\mathbf{A}$ is non-singular, show that $\kappa = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2$ [**Marks: 1**]

   (d) Condition numbers can also be defined based on other p-norms. The general p-norm condition number is given by, $\kappa_p = \|\mathbf{A}\|_p \|\mathbf{A}^{-1}\|_p$. Evaluate the 1-norm, 2-norm and ∞-norm condition numbers for the following matrices. How do these number compare with each other? [**Marks: 3**] (i) $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; (ii) $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 10 & -9 \end{bmatrix}$; (iii) $\mathbf{A} = \begin{bmatrix} 1 & 5 \\ -1 & 1 \end{bmatrix}$.

   (e) Conditions numbers play an important role in practice. We had earlier an example of an ill-conditioned system $\mathbf{Ax} = \mathbf{b}$. Consider the following systems, where: [**Marks: 2**] (i) $\mathbf{A}_1 = \begin{bmatrix} 1 & -1 \\ 10 & -9 \end{bmatrix}$; and (ii) $\mathbf{A}_2 = \begin{bmatrix} 1 & -10 \\ 1 & 10 \end{bmatrix}$.

   For $\mathbf{b} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$, what are the solutions $\mathbf{x}_1 \left(= \mathbf{A}_1^{-1}\mathbf{b}\right)$ and $\mathbf{x}_2 \left(= \mathbf{A}_2^{-1}\mathbf{b}\right)$? [**Marks: 2**]

   Suppose there is an error in the measurement of $\mathbf{b}$, and we have $\tilde{\mathbf{b}} = \begin{bmatrix} 9 \\ 1 \end{bmatrix}$. The relative error in $\mathbf{b}$ is given by $\delta b = \frac{\|\mathbf{b}-\tilde{\mathbf{b}}\|_2}{\|\mathbf{b}\|_2}$. What are the new solutions $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$? [**Marks: 2**]

   Calculate δx_1 and δx_2, the relative errors in $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively? How do these compare to δb? [**Marks: 2**]

   Note: Through this problem, you should be able to see that an ill-conditioned system has a large condition number, which can amplify error and thus lead to large uncertainty in the solutions.

2. [Programming] **Feature standardization**. Feature standardization is a common preprocessing step in machine learning and data analysis. It involves transforming the features of a dataset so that they have a mean of zero and a standard deviation of one. This is done to ensure that the features are on the same scale, which can help improve the performance of certain machine learning algorithms.

   You are given a dataset (hw.csv) consisting of heights (in inches) and weights (in pounds) of individuals. Read this file and organize the data height and weight data into a matrix $\mathbf{X} \in \mathbb{R}^{N \times 2}$, where N is the number of individuals. Carry out the principal component analysis and find the principal components of the data. Let's define the dimensionality of the data as the number of principal components required to explain 95% of the variance in the data.

(a) What is the dimensionality of the original data? [**Marks: 2**]

(b) Instead of measuring the height inches, let's measure it in micrometers and weight in tonnes instead of pounds. Convert the colmun corresponding to the height from inches to micrometer. Let's call this new data,

$$\mathbf{X}_m = \mathbf{X} \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \begin{cases} d_1 = 2.54 \times 10^4 \\ d_2 = 4.53592 \times 10^{-4} \end{cases}$$

Redo the principal component analysis on the $\mathbf{X}_m$. What is the dimensionality this modified data? Why is it different from the previous case? [**Marks: 2**]

One way to address the differences in the PCA results with $\mathbf{X}$ and $\mathbf{X}_m$ is to standardize the data before performing PCA. Standardizing the data involves transforming each feature so that it has a mean of zero and a standard deviation of one.

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2 \end{bmatrix}, \; \hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mu_i \mathbf{1}}{s_i}, \; \mu_i = \frac{1}{N} \mathbf{1}^\top \mathbf{x}_i, \; s_i = \frac{1}{N-1} \|\mathbf{x}_i - \mu_i \mathbf{1}\|_2^2$$

Perform PCA on the standardized data $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}_m$, and verify if the analysis results are different. [**Marks: 4**]

3. [Programming] **Multilead Electrocardiogram**. A multi-lead electrocardiogram (ECG) refers to a cardiac monitoring technique that involves recording electrical activity from the heart using multiple electrodes placed on the body. Each lead provides a different perspective on the heart's electrical activity, allowing for a more comprehensive assessment of its function.

In a standard 12-lead ECG, there are 10 electrodes placed on specific locations on the limbs and chest. These electrodes create 12 different "views" of the heart, each representing the electrical activity in a particular direction. The 12-lead ECG is widely used in clinical settings to diagnose various cardiac conditions, such as arrhythmias, ischemia, and other abnormalities.

The source of the electrical potential measured as ECG through bipolar electrode is the 3D current dipole $\boldsymbol{\phi}(t)$ formed by the electrical activtity of the cardiac muscle cells.

$$\boldsymbol{\phi}(t) = \begin{bmatrix} \phi_x(t) & \phi_y(t) & \phi_z(t) \end{bmatrix}^\top \mathbb{R}^3$$

where, $\phi_x, \phi_y, \phi_z \in \mathbb{R}$ are the componets of the current dipole along the x, y, and z axes. The voltage recorded by any bipolar lead $v(t)$ is proportional to the component of the current dipole along a spatial direction $\mathbf{l} \in \mathbb{R}^3$.

$$v(t) = \mathbf{l}^\top \boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)^\top \mathbf{l}$$

When the voltage is sampled at N time instants separated by the sampling time Δt, the voltage recorded by the bipolar lead is given by the vector $\mathbf{v} \in \mathbb{R}^N$.

$$\mathbf{v} = \mathbf{\Phi} \mathbf{l}$$

where,

- $\mathbf{v} = \begin{bmatrix} v(0) & v(\Delta t) & \cdots & v((N-1)\Delta t) \end{bmatrix}^\top \in \mathbb{R}^N$
- $\mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(0) & \boldsymbol{\phi}(\Delta t) & \cdots & \boldsymbol{\phi}((N-1)\Delta t) \end{bmatrix}^\top \in \mathbb{R}^{N \times 3}$

The 12 lead ECG measurement can now be represented as the following,

$$\mathbf{V} = \mathbf{\Phi L}$$

where, $\mathbf{V} \in \mathbb{R}^{N \times 12}$ is the matrix of the 12 lead ECG measurements, and $\mathbf{L} \in \mathbb{R}^{3 \times 12}$ is the matrix of the spatial directions of the 12 bipolar leads.

Assuming the the measurements are noise free, if we compute the full SVD of the matrix $\mathbf{V}$, how many of the singular values will be zero? Explain why? [**Marks: 2**]

You are given two files, containing 12-lead ECG recording from two experiments (ecg1.csv and ecg2.csv), each containing the matrix $\mathbf{V} \in \mathbb{R}^{N \times 12}$. Carry out the SVD of the matrix $\mathbf{V}$ with the 12 -lead ECG data to find the principal components and the variance explained by each of these principal components. How many principal components have a non-zero variance associated with them? If this number is greater than 3, why is it greater than 3 if the 12 lead ECG data is generated by the process, $\mathbf{V} = \mathbf{\Phi L}$? [**Marks: 4**]

Instead of looking at principal components with non-zero variance, let's look at the number of principal components that are needed to explain 95% of the total variance in the data. Are the number of principal components explaining 95% of the variance different between the two experiments? Can you explain why there might be a difference? [**Marks: 4**]