DSA5103 Assignment 1

Instructions While all languages are acceptable, it is recommended that you code using Python or MATLAB. You must write your own code. Your submission should be in softcopy including the code. Due on Feb 9, 11:59pm.

1. **Convexity**

   Show by definition that f is convex

   (1) $f(x) = x^2, x \in \mathbb{R}$,

   (2) $f(x) = x_1^2 + x_2^2 + 2x_1 + 4, x = (x1; x2) \in \mathbb{R}^2$.

2. **Matrix inverse**

   (1) Generate an $n \times n$ random matrix X from whatever distribution (e.g., every entry X_{ij} is uniformly distributed on $[0, 1]$).

   (2) Prove by definition that $A = I + X^T X$ is symmetric and positive definite (Here I denotes the identity matrix).

   (3) For different values of $n = 1000, 2000, 3000, 4000, 5000, 6000, \ldots$, measure the time of inverting the matrix A: $B = A^{-1}$ (e.g., measure by "tic" "toc" function). Plot a figure of $\log_{10}$(time) against n.

   (4) What is the largest n that your device can handle? How long it takes to invert this large $n \times n$ matrix A?

3. **Principal component analysis** on the iris data set (see `http://archive.ics.uci.edu/ml/datasets/Iris` for data description.) It contains 3 species (Setosa/Versicolor/Virginica) of 50 instances each ($n = 150$), and $p = 4$ features (sepal length, sepal width, petal length, petal width)

   (1) Download the iris data set from the above link, or load it directly in Python

   ```
   from sklearn import datasets
   iris = datasets.load_iris()
   ```

   or in MATLAB

   ```
   iris = iris_dataset;
   ```

sepal length	sepal width	pedal length	petal width	species
				setosa
				setosa
				setosa

Table 1

(2) Transform each feature to the same scale (subtracting the mean and dividing by the standard deviation). Report in Table 1 the first three samples after feature scaling.

(3) Compute the $p \times p$ covariance matrix and report the its numerical values

$$\Sigma = \begin{bmatrix} \quad & \quad \\ \\ \end{bmatrix}.$$

(4) Compute eigenvalue decomposition of $\Sigma = QDQ^T$. Report the numerical values of Q and D

$$Q = \begin{bmatrix} \quad & \quad \\ \\ \end{bmatrix}, \quad D = \begin{bmatrix} \quad & \quad \\ \\ \end{bmatrix}.$$

(5) Give the first principal component. Recast the data onto the first principal component and plot the results.

(6) Give the first two principal components. Recast the data onto the first two principal components and plot the results.