CSCI 4380/6380 – Data Mining Fall 2023

Assignment 1 Due date: September 17th, 11:59pm, EST

---

1. (35 points) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
(a) What is the mean of the data? What is the median?
(b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).
(c) What is the midrange of the data?
(d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
(e) Give the five-number summary of the data.
(f) Show a boxplot of the data.
(g) How is a quantile–quantile plot different from a quantile plot?

2. (15 points) Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

(a) Calculate the mean, median, and standard deviation of age and %fat.
(b) Draw the boxplots for age and %fat.
(c) Draw a scatter plot and a q-q plot based on these two variables.

3. (10 points) Suppose we measure the heights of 10 people, $person_1, person_2, \cdots, person_{10}$. If we create a matrix $\boldsymbol{S}$ where

$$\boldsymbol{S_{ij}} = height(person_i) - height(person_j)$$

is the matrix $\boldsymbol{S}$ symmetric? What is the trace($\boldsymbol{S}$)?

4. (5 points) Determine whether or not the vectors

$$x_1 = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

are linearly independent.

5. (10 points) Show that λ is an eigenvalue of A and list two eigenvectors corresponding to this eigenvalue:

(a) $A = \begin{pmatrix} 0 & 4 \\ -1 & 5 \end{pmatrix} \quad \lambda = 4$

(b) $A = \begin{pmatrix} 0 & 4 \\ -1 & 5 \end{pmatrix} \quad \lambda = 1$

6. (25 points) Install PyCharm/VS Code + Anaconda in your laptop.