Homework Assignment 2

1. (10 points) Suppose that you have a data set of $m(>1)$ data points $\{\mathbf{x^1} ... \mathbf{x^m}\}$ and each data point has $n(>1)$ attributes. You may store them in an $m \times n$ object-feature matrix X.

   (a) (6 points) What's the meaning of the $m \times m$ matrix XX^T and Why? What's the meaning of the $n \times n$ matrix X^TX and Why?

   (b) (2 points) What's the meaning of each eigenvalue of the matrix XX^T ? What's the meaning of each eigenvalue of the matrix X^TX ?

   (c) (2 points) What's the meaning of the rank of the matrix XX^T ? What's the meaning of the rank of the matrix X^TX ?

2. (4 points $\times 5 = 20$ points) Derive the gradient information of the following functions at a point $\mathbf{x} \in \mathbb{R}^n$:

   You may consult any textbooks or published papers, but you have to show the details of your derivation **Step by Step** .

   (a) $\nabla(\mathbf{c}^T\mathbf{x})$, $\forall \mathbf{c}, \mathbf{x} \in \mathbb{R}^n$

   (b) $\nabla(\|\mathbf{x}\|_2)$, $\forall \mathbf{x} \in \mathbb{R}^n$

   (c) $\nabla(\|\mathbf{x}\|_2^2)$, $\forall \mathbf{x} \in \mathbb{R}^n$

   (d) $\nabla(\mathbf{x}^T M \mathbf{x})$, $\forall M \in \mathbf{M}_{n\times n}$ and $\mathbf{x} \in \mathbb{R}^n$

   (e) $\nabla(\|A\mathbf{x} - b\|_2^2)$, $\forall A \in \mathbf{M}_{m\times n}, b \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$

3. (30 points) You are given a dataset named "data.csv".

   Consider the following optimization model for the centroid-based clustering:

$$\text{Minimize} \quad \sum_{i=1}^{N}\sum_{j=1}^{k} w_{ij}||\mathbf{x}^i - \mu^j||_2^2$$

subject to

$$\sum_{j=1}^{k} w_{ij} = 1 \quad \forall i = 1, \cdots, N$$

$$\sum_{i=1}^{N} w_{ij} \geq 1 \quad \forall j = 1, \ldots, k$$

where $\quad w_{ij} \in \{0, 1\}, \quad \mu^j \in \mathbb{R}^n$

   (a) (15 points) For $k = 1, 2, 3, 4, 5$, use the k-means algorithm to find the respective center point and members of each cluster.

   (b) (5 points) Plot your results of (a). Make sure to use different colors for different clusters and clearly mark the position of the center point of each cluster.

(c) (10 points) Use the Elbow rule method and Silhouette method to determine the "best" number of clusters. What will be your final choice? Why?

4. (40 points) Consider the following dataset $\{\mathbf{x}^i : i = 1, ..., 6\}$ of 6 data points :

$$\mathbf{x}^1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{x}^2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \mathbf{x}^3 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \quad \mathbf{x}^4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}^5 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{x}^6 = \begin{pmatrix} 4 \\ 1 \end{pmatrix},$$

(a) (6 points) Use the l_2-norm to calculate the pairwise-distance between the 6 data points and plot your results on a graph.

(b) (10 points) For $k = 2, 3$, use the same optimization model in Problem 2 to find the center point and members of each cluster and plot them as requested before.

(c) (6 points) Use the l_0-norm to calculate the pairwise-distance between the 6 data points and plot your results on a graph.

(d) (10 points) For $k = 2, 3$, replace the l_2-norm by l_0-norm in the objective function of the optimization model used in (b) to find the center point and members of each cluster and plot them as requested before.

(e) (8 points) Compare your results of (b) and (d). How different are the results using different norm functions? Tell me what you have learned from this exercise.