This assignment was locked Jun 28 at 2:59am.

Project Overview:

In this part, you are required to implement the k-means algorithm and apply your implementation on the given dataset, which contains a set of 2-D points. You are required to implement two different strategies for choosing the initial cluster centers.

Strategy 1: randomly pick the initial centers from the given samples.

Strategy 2: pick the first center randomly; for the i-th center (i>1), choose a sample (among all possible samples) such that the average distance of this chosen one to all previous (i-1) centers is maximal.

You need to test your implementation on the given data, with the number k of clusters ranging from 2-10. Plot the objective function value vs. the number of clusters k. Under each strategy, plot the objective function twice, each start from a different initialization.

(Referring to the course notes: When clustering the samples into k clusters/sets Di, with respective center/mean vectors $\mu_1, \mu_2, \ldots \mu_k$, the objective function is defined as

)

Algorithms:

k-Means Clustering

Resources:

A 2-D dataset to be downloaded from this link: Dataset.

Workspace:

Any Python programming environment.

Software:

Python environment.

Language(s):

Python. (MATLAB is equally fine, if you have access to it.)

Required Tasks:

1. Write code to implement the k-means algorithm with Strategy 1.

2. Use your code to do clustering on the given data; compute the objective function as a function of k (k = 2, 3, …, 10).

3. Repeat the above step with another initialization.

4. Write code to implement the k-means algorithm with Strategy 2.

5. Use your code to do clustering on the given data; compute the objective function as a function of k (k = 2, 3, …, 10).

6. Repeat the above step with another initialization.

7. Submit a short report summarizing the results, including the plots for the objective function values under different settings described above.

Deliverables and due date(s):

The code and reports are **due by June 27 .**

What to Submit:

1. Code file with comments explaining what you do for each part as directed

2. A report that summarizes the results and includes the plots for each of the objective function values.