K-means_algortihm_Strategy1 (K-mean): Overview:

In this part, you are required to implement the K-means algorithm and apply your implementation on the given dataset (AllSamples.npy), which contains a set of 2-D points. You are required to implement the following strategy for choosing the initial cluster centers.

The Strategy1 is randomly pick the initial centers from the given samples.

You need to test your implementation on the given data, with the number k of clusters ranging from 2-10, output the finial coordinate of the centeriods and compute the loss based on the objective function.

(Referring to the course notes: When clustering the samples into k clusters/sets D_i, with respective center/mean vectors $\mu_1, \mu_2, \ldots \mu_k$, the objective function is defined as $\sum_{i=1}^k \sum_{x \in D_i} ||x - \mu_i||^2$))

You are highly suggested to use the build-in Jupiter Notebook to implement your algorithm. Other python environment is okay but you must be responsible for any numerical error caused by different programming environment.

What to submit:

Once launch the lab, you will be able to find two sets of initial k and points, which are associated with your ID. Please test your algorithm with this initialization.

The 1) finial coordinate of the centeriods and 2) the loss computed by the objective function should be submitted to the quiz titled as K-means-Strategy1(submission).

Note:

1. You should implement your own K-means algorithm.

2. You will not get any points for the project by simply programming here. Please remember you need to submit the results in the Quiz.

K-means_algortihm_Strategy2 (K-mean ++) : Overview:

In this part, you are required to implement the K-means algorithm and apply your implementation on the given dataset (AllSamples.npy), which contains a set of 2-D points. You are required to implement the following strategy for choosing the initial cluster centers.

The Strategy2 is pick the first center randomly; for the ii-th center (ii>1), choose a sample (among all possible samples) such that the average distance of this chosen one to all previous (ii-1) centers is maximal.

You need to test your implementation on the given data, with the number k of clusters ranging from 2-10, output the finial coordinate of the centeriods and compute the loss based on the objective function.

(Referring to the course notes: When clustering the samples into k clusters/sets D_iD_i, with respective center/mean vectors \mu_1μ_1, \mu_2μ_2, … \mu_kμ_k, the objective function is defined as \sum_{i=1}^k \sum_{x \in D_i} ||x - \mu_i||^2 $\sum_{i=1}^{k}\sum_{x\in D_i}||x-\mu_i||^2$))

You are highly suggested to use the build-in Jupiter Notebook to implement your algorithm. Other python environment is okay but you must be responsible for any numerical error caused by different programming environment.

What to submit:

Once launch the lab, you will be able to find two sets of initial kk and points, which are associated with your ID. Please test your algorithm with this initialization.

The 1) finial coordinate of the centeriods and 2) the loss computed by the objective function should be submitted to the quiz titled as K-means-Strategy2(submission).

Note:

1. You should implement your own K-means algorithm.

2. You will not get any points for the project by simply programming here. Please remember you need to submit the results in the Quiz.