

CMPE 255-03, Spring 2024

Assignment #2

Release on March 7th, 2023

Due 11:59pm on Sunday, March 17th, 2023

Notes

This assignment should be submitted in Canvas as a format of ipython notebook (assignment2_yourFirstName_LastName.ipynb).

No late assignments will be accepted.

You may collaborate on homework but must write independent code/solutions. Copying and other forms of cheating will not be tolerated and will result in a zero score for the homework (minimal penalty) or a failing grade for the course. Your work will be graded in terms of correctness, completeness, and clarity, not just the answer. **Thus, correct answers with no or poorly written supporting steps may receive very little credit.**

NOTE: Please do not use any package/library including scikit-learn library except NumPy, Pandas, and Matplotlib.

1. (6 pts) Assigning clusters using K-means/K-mean++ clustering algorithm

Please download cluster_data1.csv.

X1	X2	X3	X4
6.7	3	5	1.7
6.3	2.9	5.6	1.8
5.6	3	4.5	1.5
7.6	3	6.6	2.1
6	3.4	4.5	1.6
6.4	3.2	5.3	2.3
7.7	2.8	6.7	2
4.8	3	1.4	0.3
5	3	1.6	0.2
5	3.4	1.6	0.4

K-means algorithm is a method to automatically cluster similar data examples together. K-means is an iterative procedure that starts by guessing the initial centroids, and then refines this guess by repeatedly assigning examples to their closest centroids and then recomputing the centroids based on the assignments until converge.

Note that the converged solution may not always be ideal and depends on the initial setting of the centroids.

To address this issue, K-means++ was introduced.

Please implement K-means clustering and K-means++ algorithm from scratch.
Then, please find the optimum K (the number of clusters).
Then, assign each data point to the corresponding clusters.

2. (4 pts) Assigning class using K-NN classification algorithm

Please download cluster_data2.csv.

X1	X2	X3	X4
5.21	3.65	1.42	0.25
5.07	3.41	1.43	0.19
5.85	2.65	4.14	1.27
5.64	2.73	4.03	1.23
6.55	2.9	5.54	2.05

Let's assume the resultant clusters from problem #1 are representing each class.
Please assign class for each data point from cluster_data2.csv file using K-NN method.
Let's assume K equals to 5 in K-NN method.

NOTE: You need to implement K-NN classification algorithm from scratch.