

CS454&554 Homework 2: k -Means Clustering for Unsupervised Learning Spring 2024/2025

In this homework, your task is to implement k -means clustering to calculate cluster centroids on a 2D dataset.

You are provided with a data file called **data.csv**. For $k=[1, 2, 3, 4, 5, 6]$:

- Repeat the training process 10 times for each k value.
- Plot the mean reconstruction loss for each k in a single plot.
- Plot the clustering visualization for the trial with the least reconstruction loss for each k (use different colors/symbols for instances assigned to different clusters).

You can use the pandas, numpy, and matplotlib libraries for your implementation; you are not allowed to use any other external libraries. You can, for example, use numpy to calculate the distance between two vectors, but you are not allowed to call any library that implements k -means (when in doubt, send email to Emir).

This homework is due **April 2nd (Wednesday), 23:00**.

Your submission should include a short report of your findings, the plots, and your source code.

Upload your report **as a pdf file** to LMS alongside your .py code file. Do not compress your submission.