

# Homework 1

Deadline: 2024/10/6 (Tue.) 23:59

~~Problem 1: Basic Numpy & Pandas~~

[hw1\\_basic\\_numpy\\_pandas.ipynb](#)

**Problem 2: Mixture of Normals**

[hw1\\_mixture\\_of\\_normals.ipynb](#)

~~Problem 3: Intersection Plot~~

[hw1\\_interaction\\_plot.ipynb](#)

~~Problem 4: Game of Life Simulation~~

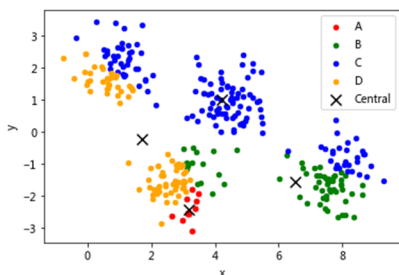
[hw1\\_mixture\\_of\\_normals.ipynb](#)

**Problem 5: K-means Clustering**

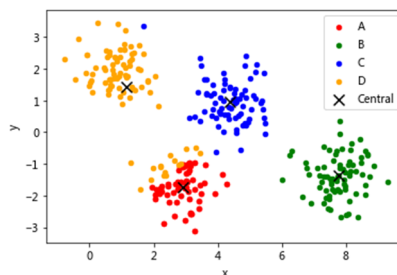
[hw1\\_kmeans.ipynb](#) (created by yourself)

Your task is to use Python (you can use only packages **numpy** and/or **Pandas**, together with plotting package like **matplotlib**. That says, you are NOT allowed to use any packages that can directly apply k-means.) to implement the well-known clustering algorithm, K-means, based on a synthetic dataset **cdata.csv**. This dataset contains two data columns, "X" and "Y", and one "cluster" column (1, 2, 3, and 4). In implementing K-means, you need to use "X" and "Y" as **features** for clustering while the "cluster" column is for your validation. Note that it is not necessary to perfectly clustering all of the data points into clusters. Also note that the "cluster" column cannot be used in clustering.

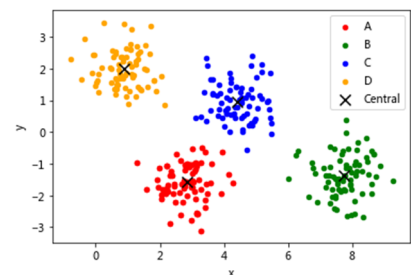
- (1) Randomly select data points as the initialized centroids. By default, please set  $K=4$ . Report and plot the process until convergence. The centroids also need to be plotted. An example is shown below. Note that it may not have 3 rounds (it can be 4 or 5 rounds, depend on initialized centroids).



Round 1

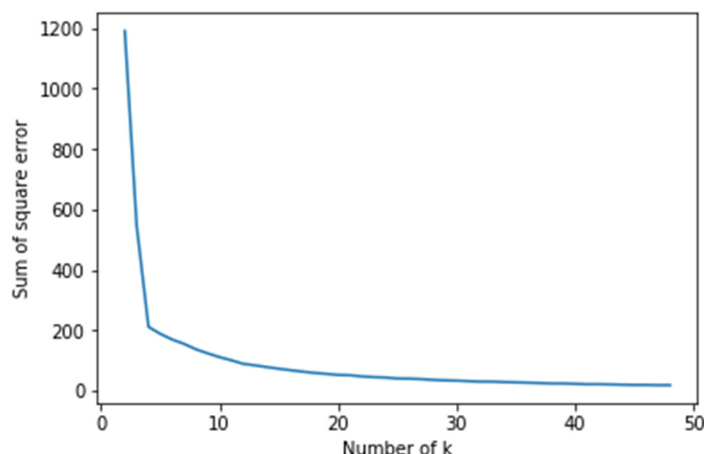


Round 2

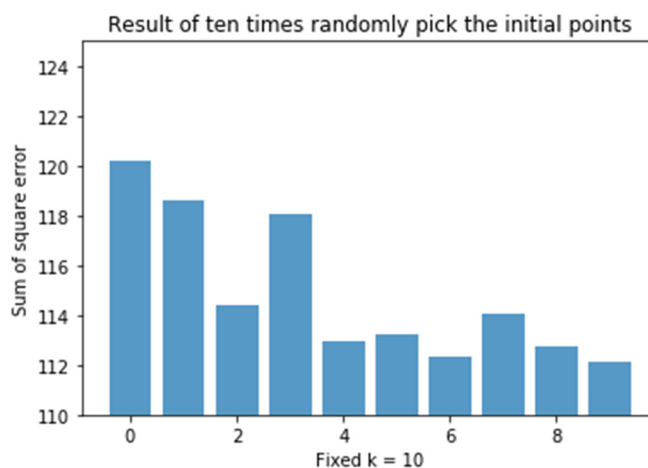


Round 3

- (2) Re-execute your K-means clustering algorithm by changing K from 2 to 50 (from 2 to 10 is also okay). Plot the K value (x-axis) vs. the value of Sum of Squared Error (SSE) (y-axis) as below. Note that it is reasonable and acceptable if the curve is 凹凸不平。☺



(3) Try 10 times of randomly initialized centroids, and plot their SSE values (y-axis) such as below.



## Important Notes

This is a homework for each **individual**. You are asked to **write comments** to describe the meaning of each part of your codes in either code block or markdown.

## How to Submit Your Homework?

Before submitting your homework, please zip all files, and name the file as "StudentID\_hw1.zip". For example, if your StudentID is H12345678, then your file name is: "**H12345678\_hw1.zip**". Then submit your file using NCKU Moodle platform <http://moodle.ncku.edu.tw>.

## Have Questions about This Homework?

Please feel free to visit TAs, and ask/discuss any questions in their office hours. We will be more than happy to help you.