Lab 7

Goal: For this lab, we will look at clustering approaches using scikitlearn

Part 1: Setup

- Familiarize yourself with the information available at https://scikit-learn.org/stable/index.html
- Install scikitlearn with pip
- Watch the video on scikitlearn posted on the course website

Part 2: Generation of data

- Write a python program that uses a userdefined function that takes a user's input to create a dataset using the make_blob function. The program should ask the user for the number of samples, clusters and the standard deviation to be used. Scale the data to 0 mean and standard deviation of 1 using scikitlearn and save the data in a csv file.

Submit your documented Python code

Part 3: Determining the number of clusters

Determine the number of clusters for the Boston House Prices dataset, using the k-means algorithm (look for the knee in the curve). You can either delete or transform any categorical attributes. Determine the number of clusters with and without scaling of the resulting numerical attributes.

Hand in the screen shot of your curve to determine the cluster number

Part 4: Application of k-means

Apply the k-means algorithm and one more clustering algorithm of your choice available in scikitlearn to the transformed House Prices Dataset and plot the resulting cluster assignments using different colors for the various clusters using subplots, side by side. Label your plots appropriately and submit a screenshot of the plots.