

CSC 546/746
Assignment 7
(25 points)

1. (1 point) Create a new Jupyter Notebook project and name it as “hw07.ipynb”.
2. (2 points) Read data from the “hw07_data.csv”. This is a multiple feature dataset.
3. (3 points) Run the Scikit-Learn standard KMeans algorithm on the dataset. Set n_clusters to 40 and output the inertia value.
4. Finding the optimal number of clusters:
 - a. (3 points) Write a loop and run standard KMeans algorithm with different cluster numbers on the data. The cluster number should range from 5 to 150 with step of 5.
 - b. (2 points) Collect the inertia values and visualize the results with K values. (inertia values vs K values)
 - c. (2 points) Could you find the elbow point? What do you think the best K value is? (Answer this question in the hw07.ipynb either in a markdown cell or in the comments)
 - d. (2 points) Collect the silhouette scores of the above models and visualize the results with K values. (silhouette_scores vs K values)
 - e. (2 points) Could you find the best K value using this method? What do you think the best K value is? (Answer this question in the hw07.ipynb either in a markdown cell or in the comments)
5. Image segmentation:
 - a. (2 points) Load the picture “sea.jpg” to the project and display it.
 - b. (2 point) Process the data, so you can apply KMeans algorithm to cluster the data.
 - c. (2 points) Run Mini-Batch KMeans algorithm on the data with cluster number $K = 16$ and $K = 8$.
 - d. (2 points) Display the original picture and processed pictures. The three pictures should be displayed in the same row with proper labels.
6. Submit “hw07.ipynb” to the Blackboard.