Intro2ML Homework for clustering

Lecturer's Name: Wen-Chieh Fang

Spring 2024

- Deadline: 05/31(Fri) 11:00 PM
- Format: Submit your homework to the e-learning platform.

1 Programming assignment

1. k-means clustering assignment

   - Mission: Write Python3 code to do clustering using k-means. Use the "elbow" method to help you select the optimal number of clusters by fitting the model with a range of values for k. Plot the clusters in 3-d figures.
   - Data set: seeds Data Set
     - Read the data set description.
   - Approaches:
     - Clustering algorithm (required): k-means.
       - Using scikit learn library: sklearn.cluster.KMeans.
       - Using Yellowbrick to perform the "elbow" method: Elbow Method
     - Plot the clusters (required): the two tools are required for 3-d projections to work.
       - Using Matplotlib: Matplotlib: Visualization with Python
       - Using Matplotlib toolkit: mplot3d
     - Other data pre-processing or feature engineering methods (optional): You can apply any technique you prefer.
   - Submission: Please submit two files. The first file should be a zip archive containing the Python 3 file (.ipynb) named with your student ID, along with any other necessary files to run your code. The second file should be a report (.pdf) named with your student ID, including a screenshot of your code's output and a description of the results."
   - Reference
     - Clustering
     - K-means Clustering
     - Three-Dimensional Plotting in Matplotlib