CSC4007 – Advanced Machine Learning
Vien Ngo, EEECS, QUB

Lab 7: K-Means Clustering

In this lab, we will use k-means clustering to cluster unlabeled data. The implementation is with Jupyter Notebook. We will use only Numpy and Matplotlib libraries. This lab will run on week 7. If you could not finish them during this lab, you can finish at home. Through this lab, you will practice to write code to understand and re-implement the k-means clustering algorithm learned in Lecture 06.

STEP 1: Download these files: *clustering.ipynb* (main code) and *clustering.data* (data file).

STEP 2: Open the notebook file: "*clustering.ipynb*" and learn how to write code in Python (using Numpy only) to implement the k-means clustering algorithm from scratch. (The code also contains an example of using scikit-learn as demonstration: optional).

STEP 3: The code in "*clustering.ipynb*" use data generated in program. Modify the code to load data in file *clustering.data (Note that the data consists of both input and output as ground-truth),* and re-work Step 2.