Machine Learning – COMP3032

Tutorial and Lab Practice 5 – Week 6

This lab pactice focuses on the concepts and techniques of unsupervised learning: clustering.

Tutorial

1. Review the terminology and concepts introduced and algorithems taught in Lecture 5.

2. How would you define clustering? Can you name two clustering algorithms?

3. What are some of the main applications of clustering algorithms?

4. Describe the technique of the elbow rule to select the right number of clusters when using K-Means.

5. What is a silhouette score? How to use silhouette scores to select the right number of clusters when using K-Means?

6. Can you name and describe some K-Means improvement algorithms?

Lab Practice

1. Download, open and run the program tut05.py. Read and understand the program.

   1) Understand each figure plotted. What is the purpose of the last figure?
   2) Revise the program by adding one more interation of running K-Means, and add the result to the last figure.
   3) Observe the movements of the centroids.
   4) What happens if changing the number of clusters to 4 or 8 when running K-Means?