CI6227 Data Mining – Assignment 3

In this assignment you will get some practical experience working with a clustering algorithm. The gist of the assignment is to use an implementation of a K-means algorithm and run it to demonstrate the results of good and bad choices of initial centroids.

Assignment

- Find (or write yourself) a tool that can visualize iterations of K-means algorithm.
- Run it several times
    - You can use *any* 2-dimensional dataset with *clearly visible* clusters
    - You can do *any* number of clusters $K \geq 2$
    - Take screenshots of the first 5 iterations of every run
    - Screenshot should show the data points, the centroids, and the current point-centroid assignment
- Submit the 5 screenshots from a successful run (centroids reach the true clusters) and from an unsuccessful run (centroids do not converge into true clusters)

Reporting

Your submission for this assignment is 10 image files (JPG or PNG). Screenshots for the successful run should be named **A1, …, A5** (e.g., A1`.jpg`, A2`.jpg`, *etc.*); screenshots for the unsuccessful run should be named **B1, …, B5**.

Submission

Submission should be done in NTULearn. Access the assignment submission page through the left navigation bar by selecting "Assignments". Submit ten image files named A1, … A4, A5, B1, B2, …, B5. Submissions are accepted up to Thursday, December 4th 23:59:59.