Quiz 3

11:30am-12:15pm, Friday, October 13, 2017

Use your own python k-means and hierarchical functions to answer every question.

1. Do k-means clustering of the iris dataset. Use the first, second, and third setosa flower as the initializing centroid. Consider all of the four flower attributes. Specify $k = 3$. Answering the following questions:

(1) (15 points) What is the new centroid after the first iteration?

(2) (15 points) What is the new centroid after the second iteration?

(3) (15 points) How many iterations does it take your function to reach convergence?

(4) (15 points) After convergence, how many setosa, versicolor, virginica flowers were clustered correctly?

2. Do hierarchical clustering of the dataset `SCLC_study_output_filtered_2.csv` using single linkage and euclidean distance.

(1) (20 points) What is the first pair of samples to be merged? What is the second pair of samples to be merged?

(2) (20 points) Compare clustering results from k-means with results from hierarchical in terms of how many samples are corrected clustered. For k-means, specify $k = 2$ and randomly choose two initial centroids.