

Week 10 - Clustering

Exercises

Question 1.

This question asks you to manually perform K-means clustering manually, with $K = 2$, on a small sample, $n = 6$, and $p = 2$ features.

	x_1	x_2
Obs.		
0	1	4
1	1	3
2	0	4
3	5	1
4	6	2
5	4	0

a. Plot the observations

► [Click here for answer](#)

b. Randomly assign a cluster label to each observation. In Python you can use `np.random.randint`. Report the cluster labels for each observation.

► [Click here for answer](#)

c. Compute the centroid for each cluster.

► [Click here for answer](#)

d. Assign each observation to the centroid to which it is closest, in terms of Euclidean distance. Report the cluster labels for each observation.

► [Click here for answer](#)

e. Find where the k-means cluster centers (e.g. where questions (c) and (d) stops changing), then color your plot according to these cluster labels.

► [Click here for answer](#)

Question 2.

Describe two techniques to help select the number of clusters when using K-Means.

► [Click here for answer](#)

Question 3.

Suppose we have a dissimilarity matrix as follows:

$$\begin{bmatrix} & 0.3 & 0.4 & 0.7 \\ 0.3 & & 0.5 & 0.8 \\ 0.4 & 0.5 & & 0.45 \\ 0.7 & 0.8 & 0.45 & \end{bmatrix}$$

This means the dissimilarity between the first and second observation is 0.3, second and fourth is 0.8 ect.

a. Sketch or code a diagram that results from hierarchically clustering these four observations using **complete** linkage.

► [Click here for answer](#)

b. Suppose we cut the dendogram from question (a) such that two there are two clusters, which observations are in which cluster?

► [Click here for answer](#)

c. Sketch or code a diagram that results from hierarchically clustering these four observations using **single** linkage.

► [Click here for answer](#)

d. Suppose we cut the dendogram from question (c) such that two there are two clusters, which observations are in which cluster?

► [Click here for answer](#)

[NbConvertApp] Converting notebook Clustering_Exercises.ipynb to html

```
[NbConvertApp] Writing 287714 bytes to Clustering_Exercises.html
```