

Geometric Data Analysis HW 1

Gilad Turok, gt2453
gt2453@columbia.edu

February 19, 2023

1 k -means vs Single-Linkage Clustering

I generated data with three 2-dimensional Gaussians with identity covariance matrices. I cluster this data with k -means and single-linkage clustering with means that vary. I use 4 different sets of means, beginning from very close to each other to further and further apart:

$$\begin{aligned}\mu_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & \mu_2 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & \mu_3 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \mu_1 &= \begin{bmatrix} -2 \\ 2 \end{bmatrix}, & \mu_2 &= \begin{bmatrix} 0 \\ -2 \end{bmatrix}, & \mu_3 &= \begin{bmatrix} 2 \\ 2 \end{bmatrix} \\ \mu_1 &= \begin{bmatrix} -5 \\ 5 \end{bmatrix}, & \mu_2 &= \begin{bmatrix} 0 \\ -5 \end{bmatrix}, & \mu_3 &= \begin{bmatrix} 5 \\ 5 \end{bmatrix} \\ \mu_1 &= \begin{bmatrix} -10 \\ 10 \end{bmatrix}, & \mu_2 &= \begin{bmatrix} 10 \\ -10 \end{bmatrix}, & \mu_3 &= \begin{bmatrix} 10 \\ 10 \end{bmatrix}\end{aligned}$$

I noticed that the k -means algorithm is very sensitive to the initial means. If the initial means are too far apart, the algorithm will not converge. If the initial means are too close together, the algorithm will converge to a local minimum. The single-linkage clustering algorithm is not sensitive to the initial means. The single-linkage clustering algorithm will always converge to the same result, regardless of the initial means. The k -means algorithm will converge to different results depending on the initial means. The k -means algorithm will converge to the same result as the single-linkage clustering algorithm if the initial means are close enough to the true means. The k -means algorithm will converge to a different result than the single-linkage clustering algorithm if the initial means are too far apart from the true means.

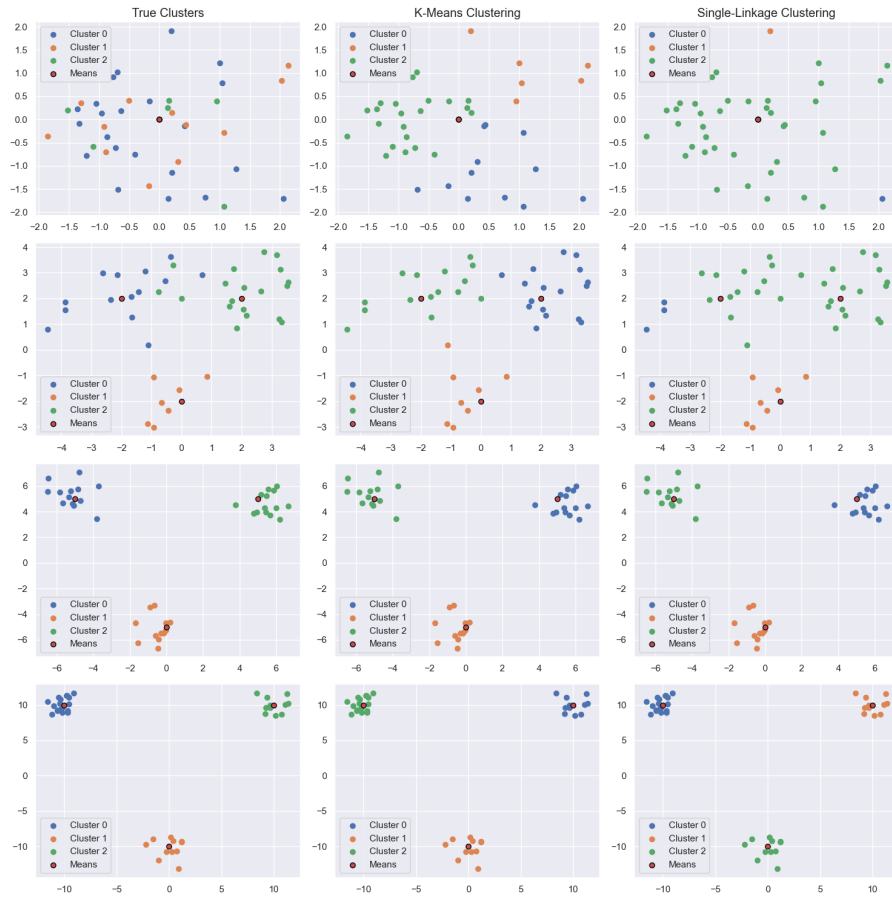


Figure 1: k -means vs single-linkage clustering on 3 Gaussians with means at different distances