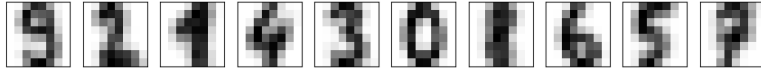


Homework 11 Report

May 1, 2019

1. Display your output of plot Kmeans(). Does your plot match your expectations?



Yes, the plot matches my expectations. I was impressed by how well some numbers (2, 3, 0, etc.) were easily distinguishable, but also surprised how almost unrecognizable some numbers (7,8) looked.

2. In this assignment, you implemented k-means through a Euclidean distance metric. Describe other distance metrics that can be used and how they cluster inputs.

You could use Manhattan Distance (the absolute difference between two points) instead of euclidean distance. Depending on the dataset, the Manhattan distance could potentially have less distortion in kmeans then when using Euclidean distance. Source: "K-means with Three different Distance Metrics" <https://pdfs.semanticscholar.org/a630/316f9c98839098747007753a9bb6d05f752e.pdf>

3. What would you expect the clusters centers (centroids) to look like if use $K < 10$? $K > 10$?

If $K < 10$, then I would predict some similar groups that are unclear in the plot (7,8,9) would be grouped together. If K was larger than 10 then a group with high variety would be split into multiple groups.