Homework 12 Solutions

Due: Thursday, April 25, 2024 at 12:00pm (Noon)

Project Report

Consider the following two-dimensional dataset:

a. Display your output of \texttt{plot.kmeans()}. Does your plot match your expectations?

\textbf{Solution:} Plot should look like the image below. Some possible answers could include: all numbers 0-9 are present, numbers in general are fuzzier or less fuzzy than expected, certain numbers like 7 are fuzzier due to variation in how people write it, etc.

\begin{center}
\includegraphics[width=0.4\textwidth]{plot_kmeans.png}
\end{center}

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

\textbf{Solution:} Many possible answers. Ex) Manhattan distance, cosine distance

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

\textbf{Solution:}
If $K < 10$, clusters would be more difficult to read/interpret since some would be a blend of two or more "similar" digits. (Ex: 1 and 7 might be the same cluster).

If $K > 10$, our clusters will begin distinguishing between different ways of writing the same digit. (Ex: 4 with the leftmost edge slanted vs vertical).