

Homework 2

Jakubik, Gavin
gjakubik@nd.edu

November 8, 2021

1 Compare Initial Centroids

These two did not generate the same result. I prefer the second initialization, which is $(7, 7), (7, 14)$, because the grouping makes more sense logically. All of the teams who won more games in 2017 were together, while the teams who lost more games in 2017 were together.

2 Compare Features

I like this clustering more than the last, because the groups seem more obvious even visually, so it seems better to cluster. In this case, the teams ranked highly in 2015 are in one cluster while those ranked lower in that same year are in another.

3 Choose a good K

My initial centroids were: $(1, 1), (10, 10), (25, 25)$

These worked well, and I like it better than a $K = 2$ cluster because it breaks one of the previous clusters into two obvious clusters. In this case, the teams ranked very high in 2015 are in their own group, while the others are now all teams that were good in 2015. The remaining two groups are teams good in both years and then good in 2015 but worse in 2017. These are the most logical and useful clusters out of the ones we have seen because they break the data into meaningful groups.