

How to Determine k

Praneet Nandan

September 2021

We take our favourite number, which in my case is 7.

.

On a serious note, we should have a k which is big enough so that the distances of the points to the centroids is minimized, and on the other hand putting a k tooo large will subdivide obvious clusters also into different parts.

.

So maybe a quantity like the total distance of the points from their closest centroid, multiplied by the total number of centroid points. Then we try to minimize this.

.

Another idea is to minimize the distance to closest centroid and maximixe the distance to the next closest.