1. What does k-means clustering do?

K-means clustering is a method to group a dataset into clusters based on similar attributes. Based on the value of k, it forms k amount of groups. We can try out different values of k to determine the best fit.

1. What does partitioning around medoid do?

Partitioning around medoids simply chooses a point from the dataset and forms clusters around those points. This different than k-means clustering because k-means clustering will form clusters around the means.

1. What are three methods of determining the number of clusters and how do they differ?

The three methods are empirical, elbow, and cross validation. The empirical method simply takes the square root of the number of observations divided by two. The elbow method plots the sum of squared error against the number of clusters. Wherever the plot bends, or “the elbow”, is the number of clusters to use. Cross validation takes a candidate k and use m-validation with k. It divides the data into m random parts, performs clustering on m-1 parts, and evaluates the clustering on the final part.

1. Please write any questions you may have.

None right now.