Joshua Potter, jop13

K-Means Algorithm

The k-means algorithm accepts a selection of data of the same type and attempts to predict k-amount “clusters” from x-amount of elements, where each element belongs in a certain cluster. This algorithm is used when to defined labels for data, where there is an unknown amount of labels. Within machine learning labels are used to categorize the data, and without labels it’s difficult the user to define rule sets to classify the data. The k-means algorithm predicts the amount of clusters based on the data provided, which in turn creates labels to classify all the information.

As the algorithm iterates through each element in the dataset, it determines whether it belongs within a cluster. Then, after adding the element to the cluster, it updates that cluster’s “mean” value to further represent that cluster for future comparisons against the dataset. This continues until all elements are placed within one of k-amount of clusters.