**KMEANS, KNN and DIFFERENCES**

KMEANS

*K*-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the *K*-means clustering algorithm are:

1. The centroids of the *K* clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents.  The *Κ*-means clustering algorithm uses iterative refinement to produce a final result. The algorithm inputs are the number of clusters *Κ* and the data set. The data set is a collection of features for each data point. The algorithms start with initial estimates for the *Κ*centroids, which can either be randomly generated or randomly selected from the data set. The algorithm then iterates between two steps:

1. Data assignment step:

Each centroid defines one of the clusters. In this step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance

2. Centroid update step:

In this step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome.

KNN

KNN Algorithm is based on **feature similarity**: How closely out-of-sample features resemble our training set determines how we classify a given data point:

KNN can be used for **classification** — the output is a class membership (predicts a class — a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It can also be used for **regression** — output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

DIFFERENCES

KNN Algorithm is based on feature similarity and K-means refers to the division of objects into clusters (such that each object is in exactly one cluster, not several).

KNN is a classification technique and K-means is a clustering technique.

More differences between KNN and KMeans

K-Nearest Neighbors

• Supervised Technique

• Used for Classification or Regression

• Used for classification and regression of known data where usually the target attribute/variable is known before hand.

• KNN needs labelled points

K-Means Clustering

• Unsupervised Technique

• Used for Clustering

• Used for scenarios like understanding the population demographics, social media trends, anomaly detection, etc.

• K-Means doesn’t require labelled points