K-means Algotithm

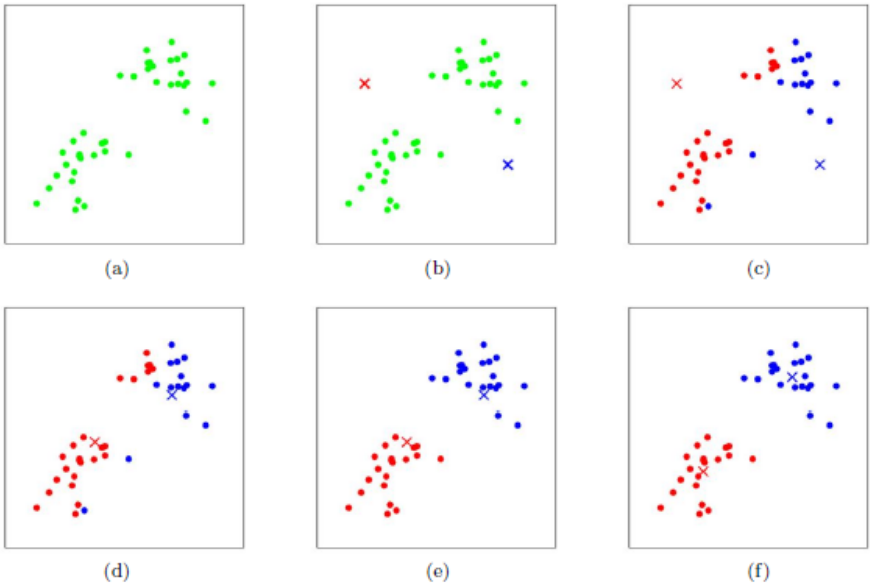
First part

Concept:

K-means algorithm is an unsupervised clustering algorithm, which adopts an iterative method and tries to divide the data set into K non-overlapping clustering, in which each data point only belongs to one group. It tries to make the data points within the cluster as similar as possible while making the cluster as different as possible.

Principle(central idea):

pre-determined constant K, constant K means that the final number of cluster categories, first randomly selected initial point as the center of mass, and by calculating the similarity between each sample and center of mass (here as the Euclidean distance), similar to the sample points to the classes, and then to recalculate the center of mass of each class (that is, for the class center), repeat this process, until the center of mass is no longer change, eventually will determine the category of each sample and the center of mass of each class.



Second Part

Algorithm Flow:

1. Select the number of clustering K (when K-means algorithm transfers hyperparameters, it only needs to set the maximum k value)
2. Randomly generate K clustering, and then determine the clustering center, or directly generate K centers.
3. Determine the clustering center of each point.
4. Then calculate the new clustering center.
5. Repeat the above steps until the convergence requirements are met.(This is usually the fixed center point that no longer changes.)