kuala lumpur

school of ai

Machine Learning 101
K-Means Theory

What is Unsupervised machine learning

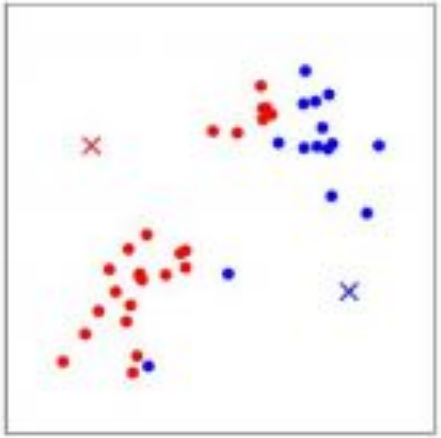
- Unsupervised learning is where you only have input data (X) and no corresponding output variables.
- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

What is K-Means

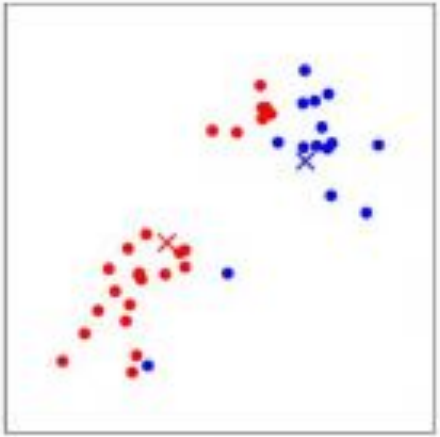
- K-Means is one of the most popular "clustering" algorithms. K-means stores k centroids that it uses to define clusters. A point is considered to be in a particular cluster if it is closer to that cluster centroid than any other centroid.

(a)

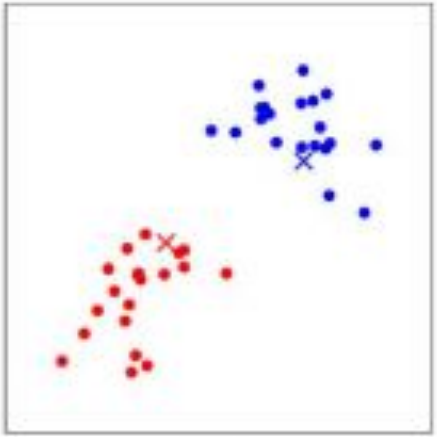
(b)



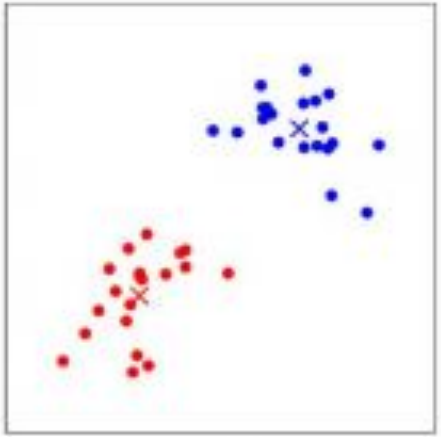
(c)



(d)

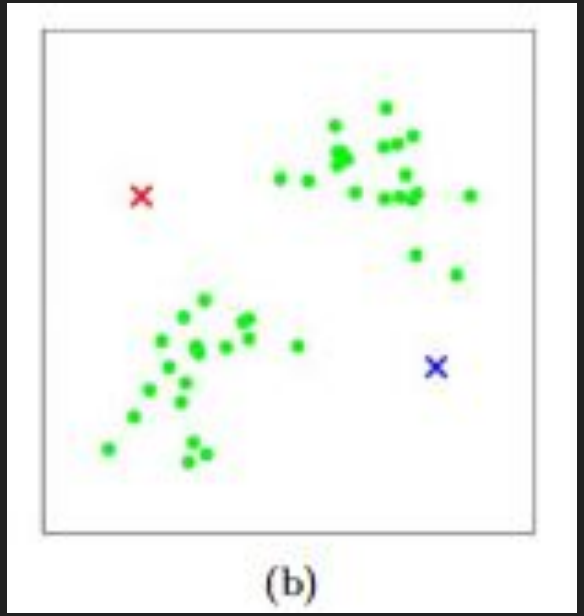


(e)

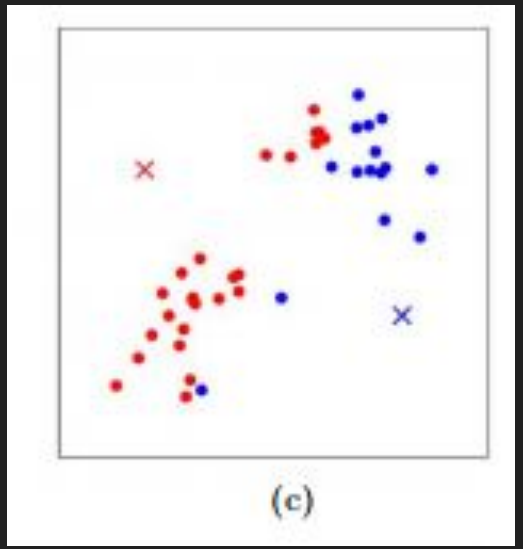


(f)

Choose K-Centroids Randomly



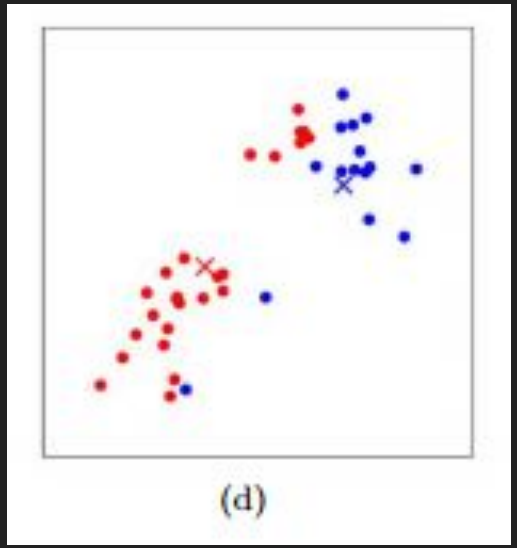
Calculate minimum distance to centroid and assign point according to minimum distance



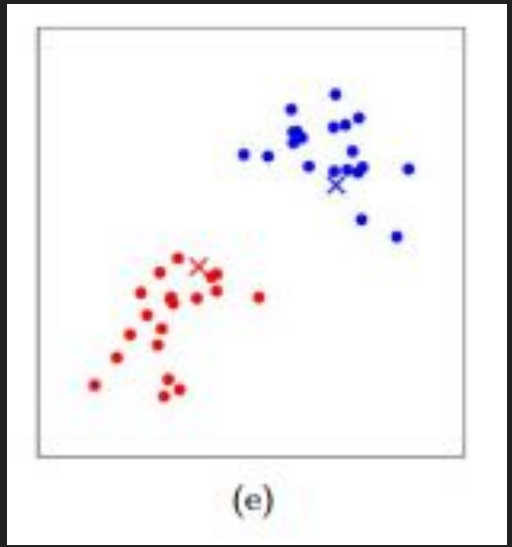
$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

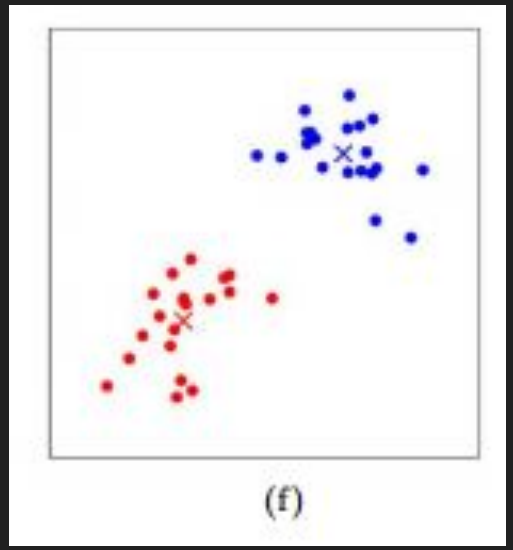
Calculate new centroid by calculating the average position according to all values assigned to the previous centroid



Recalculate minimum distance to the new centroids and reassigned accordingly



Repeat steps until convergence is reached

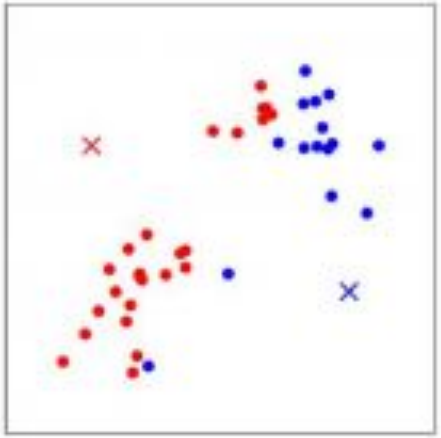


3 Basic Steps in K-Means Clustering

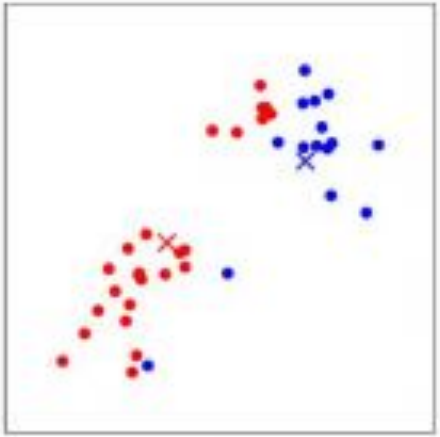
1. Start with choosing random K points as controids
2. Obtain the points with the minimum distance to the centroid
3. Calculate the new centroid by obtaining the average of the points
4. Repeat steps 2 and 3 until centroid converges (centroid does not change) or reach number of iteration required.

(a)

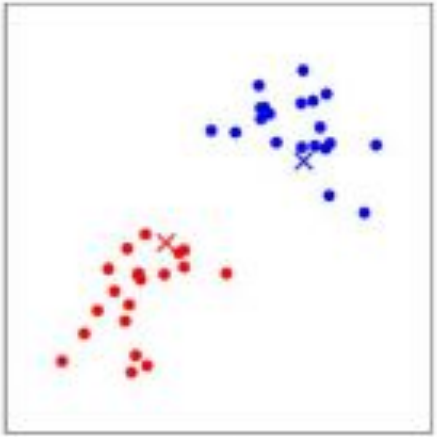
(b)



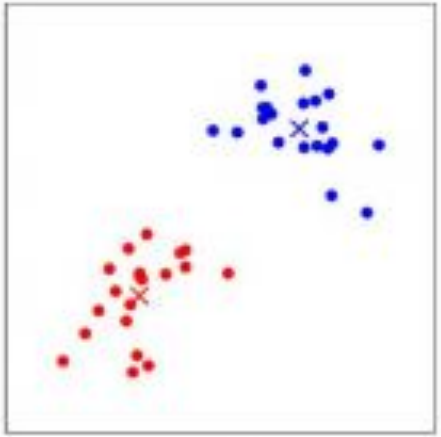
(c)



(d)



(e)



(f)

References:

https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

https://www.youtube.com/watch?v=hDmNF9JG3lo

https://youtu.be/8dqdDEyzkFA

http://stanford.edu/~cpiech/cs221/handouts/kmeans.html

https://github.com/huseinzol05/Machine-Learning-Numpy/blob/master/clustering/k-mean/mushroom-k5.ipynb

https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/