

Data Science with Python :

K-Nearest Neighbour

Algorithm #731

PRESENTED BY:

DEEPTHI M

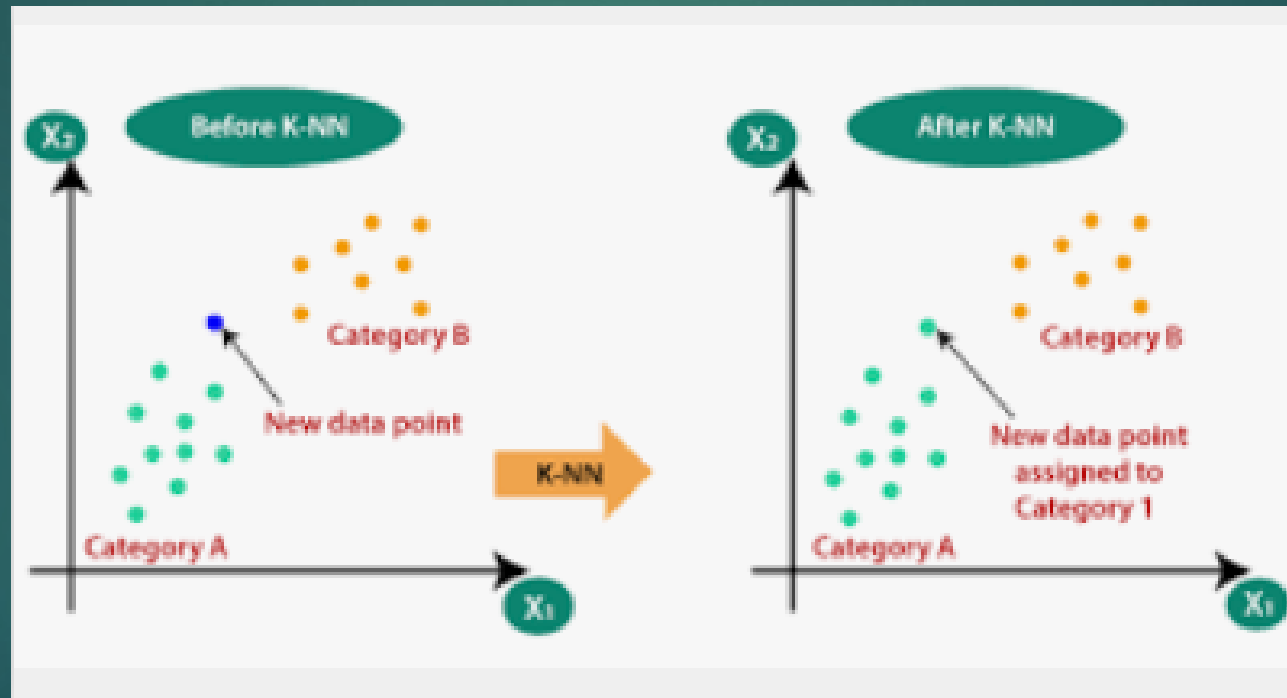
GITHUB LINK: [HTTPS://GITHUB.COM/DEEPTHI1107](https://github.com/DEEPTHI1107)

BATCH NUMBER: 05

SERIAL NUMBER: 172

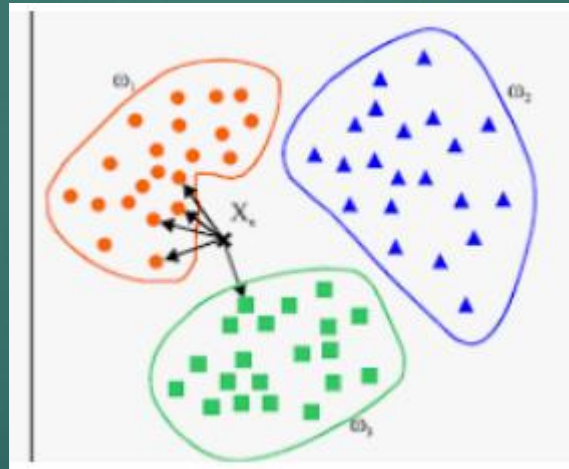
K-Nearest Neighbor(KNN) Algorithm for Machine Learning:

KNN is a supervised machine learning algorithm.
It is used when output is discrete or continues in nature.



What is KNN?

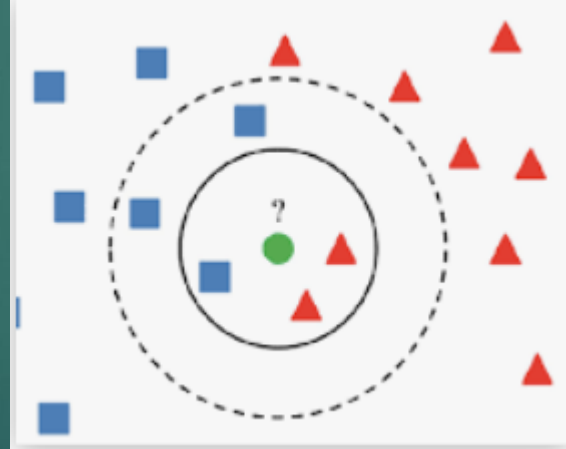
KNN algorithm assumes that similar data points lie near each other. So when new data point is introduced it characterizes the data point on based on distance generally through Euclidean distance, and then it assigns the category on the most frequency category or average based on k- value.



Choosing of k-value in KNN:

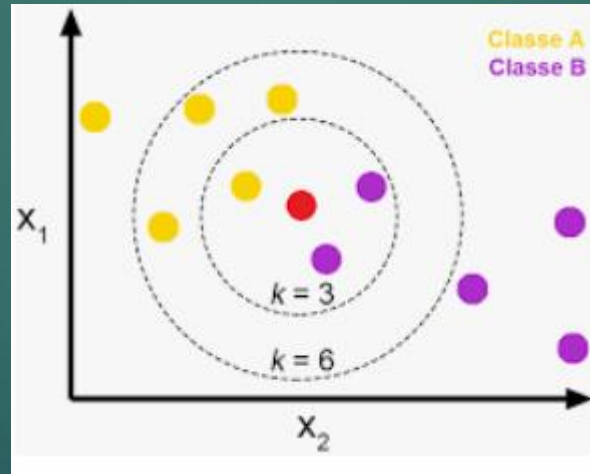
Choosing of k-value is critical because it is used to solve the problem of bias-variance trade-off.

- Low k-value, it is influenced by outliers.
- High k-value, might introduce data points from other category data points.



How to choose k-value?

- Generally, we assign k-value randomly and check the accuracy.
 - Low error-rate and high accuracy is good.
 - Derive a plot between error rate and k-value denoting values in a defined range
- And then choose the k-value which have a min error rate.
- $\text{Sqrt}(n)$ for optimal k-value.






Advantages:

- No training period (lazy learner).
- New data can be added seamlessly.
- Easy to implement.

Disadvantages:

- Does not work well on large datasets.
- Does not work well with high- dimensions.
- Sensitive to noisy data, missing values and outliers.



I hope you all got an over on the topic:
KNN

Hope, everyone enjoyed the video.