K-NEAREST NEIGHBOURS

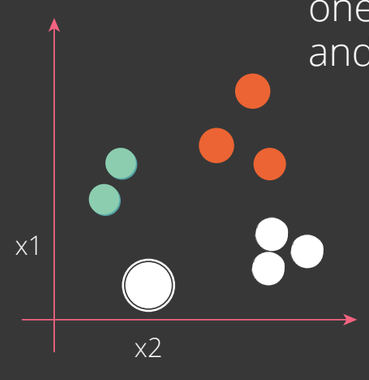
By: Janhavi Thakur

What is KNN?

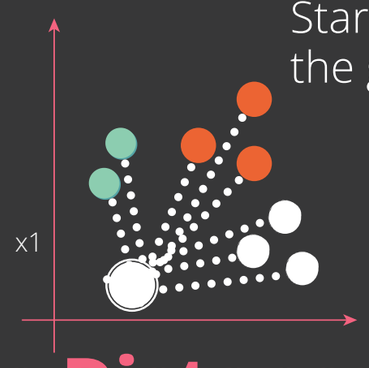
K-Nearest Neighbour algorithm is a simple but most used classification algorithm. It can also be used for regression.

KNN is non-paramateric(means that it does not makeany assumptions on the underlying data distribution) , instance-based(means that the algorithm doesn't explicitly learn a model.Instead it chooses to memorize a training instance)and used in a supervised learning setting.

We want to clarify the grey point into one of the three classes blue, orange and white.



Start by calculating a distance between the grey point and the k-nearest point



The Distance

Euclidean distance is calculated as the square root of the sum of square differences between a new point and an existing point across all input attributes.

Other popular distance measure include:

- Hamming Distance
- Manhattan Distance
- Minkowski Distance

ADVANTAGES

- An algorithm is easy and simple to implement.
- There is need to build a model,tune several parameters,or make additional assumptions.
- The algorithm is versatile.It can be used for classification, regression, ans search.

DISADVANTAGES

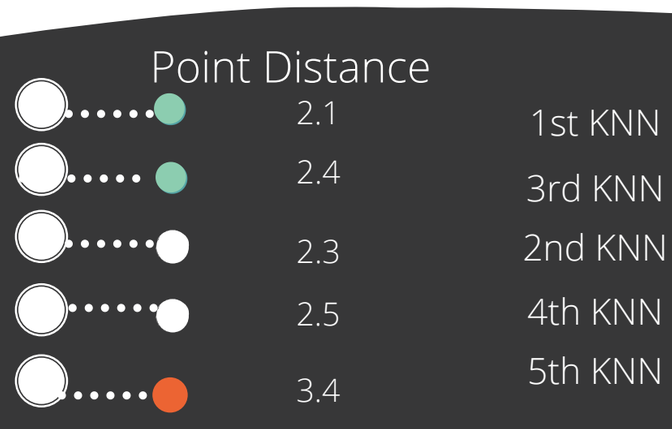
- The algorithm gets significantly slower as the number of the examples or independent variables increase.

How does K-NN algorithm works?

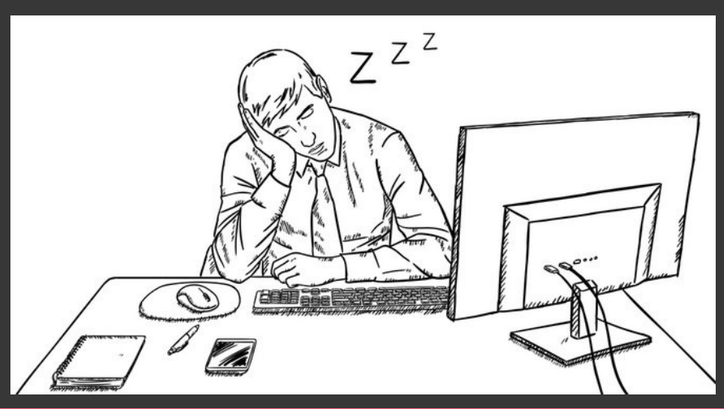
K_NN when used for classification-the output is a class membership(predicates a class-a discrete value).

There are three key elements of this approach: a set of labeled object,e.g., a set of stored records, a distance between objects,and the value of K,the number of the nearest neighbours.

Euclidean Distance= $d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$



Class ● wins the vote!!
Point ○ is predicted to be of class ●



KNN is also a lazy algorithm beacuse it is instance based.