📝K-Nearest Neighbour (KNN) Algorithm

* Supervised Machine Learning | Classification type
* It stores all the available data and classifies the new data point based on the similarity.
* It is non-parametric algorithm which means it does not make any assumption on underlying data.
* It is lazy learner algorithm because it doesn’t learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. 👌

✍️How KNN Work

1. Select the number K of the neighbours.
2. Calculate the Euclidean distance of **K number of neighbour.**
3. Take the K nearest neighbours as per the calculated Euclidean distance.
4. Among the K neighbours, count the number of the data points in each category.
5. Assign the new data points to that category for which the number of the neighbour is maximum.
6. Our model is ready.

✍️HOW TO SELECT THE VALUE OF K

* The value k must be odd number
* It should not be too high or too low
* If the value of k is too low, it can be noisy and leads to the outlier in the method
* If the k value is too high, it may cause some difficulties

📝ADVANTAGE:

* Simplest and easiest algorithm.
* Robust to the noisy training data.
* It can be effective if the training dataset is large.

📝DISADVANTAGE:

* It always determines the values of k, which sometimes complex to determine
* The computation cost is high, becausse of calculating the Euclidian distance between new data point and all the training data.

📒EXAMPLE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Age | Gender | Sport | Distance |
| Ajay | 32 | M-0 | Football | 27.02 |
| Mark | 40 | M | Neither | 35.01 |
| Zara | 34 | F-1 | Cricket | 9.00 |
| Sachin | 55 | M | Neither | 50.01 |
| Rahul | 40 | M | Cricket | 35.01 |
| Snth | 15 | M | Cricket | 10.00 |
| Michel | 15 | M | Football | 10.05 |
| Angelina | 5 | F | ? |  |

Distance = √(x2-x1)2+(y2-y1)2 ------/> Euclidian distance

We have selected our k = 3, x = age, y = gender (male=0, female=1)

So, three nearest neighbours will be selected, the nearest will be those having small value in the distance column.

* Zara
* Snth
* Michel

From the above three, zara is the closest so we can categorize Angelina to zara class. So, Angelina sport class will be cricket.