Introduction to K-NN (K-Nearest Neighbors) algorithm

Supervised Learning -

↳ Regression Task

K- Nearest Neighbors Algorithm :-

↳ K: No of Neighbors

To get prediction

Aggregate the output of the nearest nighbors

K-Nearest Neighbors:- K Nearest Neighbour is a simple algorithm that stores all the available cases and classifies the new data or case based on a similarity measure. It is mostly used to classifies a data point based on how its neighbours are classified.

Example :-

Training Dataset - To train model and to figure out your prediction

Univariate Task $f(X_1)$ - Using only one feature for prediction

	Age (in year) X_1	Weight (in kgs) Y
x_1	10.8	30
x_2	19	36
x_3	24.8	45.5
x_4	35	90

K=1 : one nearest neighbors

K-NN ($X_1 = 10$) → Predict the weight i.e $\hat{y}$

↓

Age of person is 10

The closest age from the training dataset from $X_1 = 10$ is x_1 [10.8] so predicted weight $\hat{y}$ is 30

K=2: Two nearest Neighbor

K-NN ($X_1 = 10$) $\rightarrow \hat{Y} \rightarrow \{30, 36\}$

x_1: Closest from training dataset (X_1)

x_2: second closest From training dataset (X_2)

$$\text{Mean (Avg)} = \frac{1}{2}(30+36) = \frac{1}{2}(30) + \frac{1}{2}(36)$$

$w_{closest}$ $w_{second\ closest}$

$\frac{1}{2}$ weight to closest $\frac{1}{2}$ Weight to second closest

Distance weighted prediction :-

[closest]	[second closest]
$d_{closest}$	$d_{second\ closest}$
$\text{distance}_{closest}$	$\text{distance}_{second\ closest}$
$x_1 - X_1$	$x_2 - X_1$
$10.8 - 10$	$19 - 10$
$[0.8]$	$[9]$

In distance weighted prediction we are going to give more weight to closest value, instead of giving half-half weight to both second closest & closest value

$$w_{sample} \propto \frac{1}{distance_{sample}}$$

Note :- $\sum$ Weight of the value should be 1.

$$w_{closest} = \frac{\frac{1}{0.8}}{\frac{1}{0.8} + \frac{1}{9}} \qquad w_{second\ closest} = \frac{\frac{1}{9}}{\frac{1}{0.8} + \frac{1}{9}}$$

$$\frac{1.25}{1.25 + 0.11} + \frac{0.11}{1.25 + 0.11}$$

$$\approx 1$$

$$\therefore \hat{y} = w_{closest} \times 30 + w_{second\ closest} \times 36$$

Distance weighted Prediction