KNN (K-nearest neighbors)

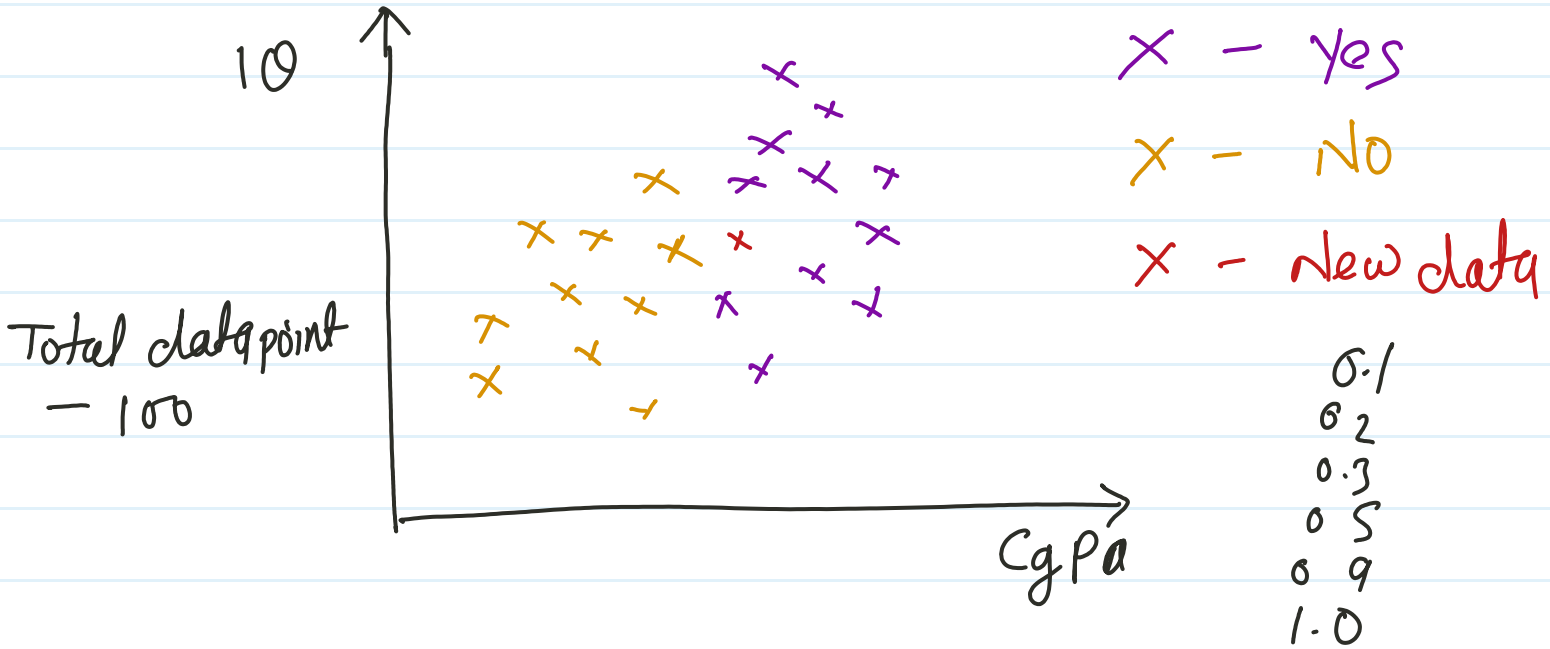
It is used to build both Reg. and class. problem.

KNN

- classification
- Regression

Eg:-

cgpa	IQ	placement
7	75	Y
8	71	Y
9	80	Y
6	45	N
5	60	N



We have a new datapoint and we want to predict it as student get placement Y/N.

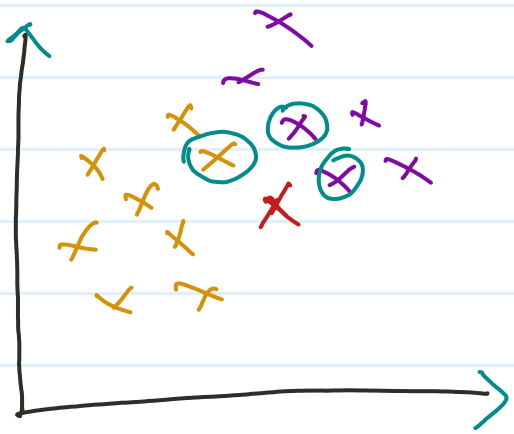
We have 100 data point, We will calculate distance of every data point from New data then we will sort the distance in ascending order.

KNN - K - nearest neighbors

K is a hyperparameter, We can assume it in odd numbers like for this problem

we are taking K = 3

As we took K = 3 so we need to check top 3 distance in ascending



These three are the nearest of new data

X - Y
X - Y
X - N

From 3, 2 has Y and 1 has N

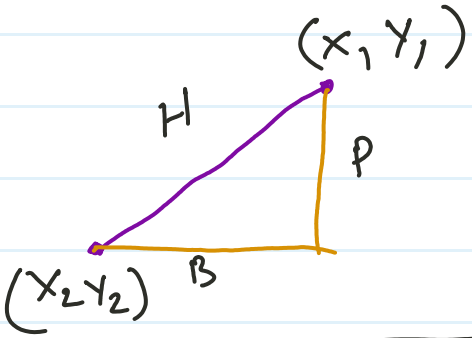
It works on majority so it will be fall into yes class.

So how kNN basic works on this distance based approach.

⇒ For classification -

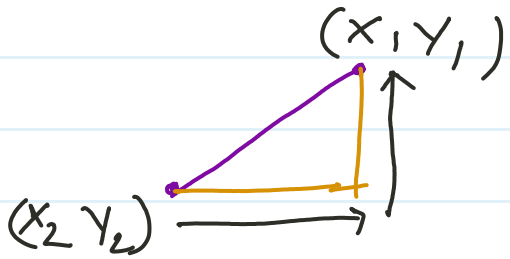
Two types of distance methods are given -

① Euclidean Distance



$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

② manhattan Distance

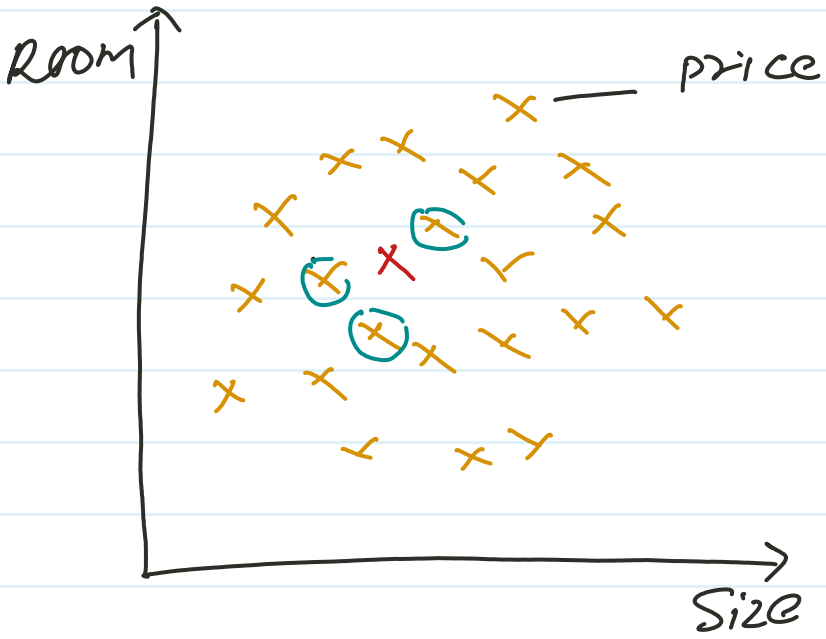


⇒ For Regression problem

Directly understand by example

size	Room	price
—	—	20L
—	—	40L
—	—	35.5L

⇒ since y is continous variable it is regression problem



$k = 3$

Same approach we calculate distance of

all the data point for new point sort them ascending,

we select 3 sorted distance and calculate the avg. of all of three distance and it will be our final output.

⇒ Limitation –

① we cann't use with large dataset.

② hug impact by outlier

③ Sensitive to missing value