Naive Baye's :-

→ used for classification

→ works on Baye theorm.

ex:- rolling a dice

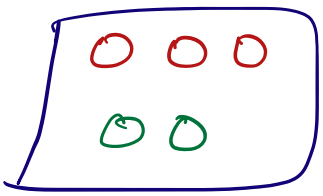
{1, 2, 3, 4, 5, 6}

1. $P(1) = \frac{1}{6}$

$P(2) = 1/6$

$P(3) = 1/6$

} Independent events.

2. 

$P(R) = 3/5$ → total

→ first event

$P(4) = 2/4 = 1/2$

↳ out of we took out 1 Red.

$P(R \text{ and } 4) = P(R) * P(4/R)$

↳ Conditional probability

$P(A \text{ and } B) = P(A) * P(B/A)$

$P(A \text{ and } B) = P(B \text{ and } A)$

$$P(A) * P(B/A) = P(B) * P(A/B)$$

$$P(B/A) = \frac{P(B) * P(A/B)}{P(A)}$$

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad \ldots \quad x_n$ → I/P

y → o/p

train data

we will use Bayes theorem.

$$P(y/x_1, x_2, x_3 \ldots x_n) = \frac{P(y) * P(x_1, x_2 \ldots x_n / y)}{P(x_1, x_2 \ldots x_n)}$$

$$= \frac{P(y) * P(x_1/y_1) * P(x_2/y_2) \ldots P(x_n/y_n)}{P(x_1) * P(x_2) * \ldots \quad \text{---} \quad P(x_n)}$$

ex:

x_1	x_2	x_3	x_4	y
				yes
				no

$$P(y=yes/x_i) = P(yes) * P(x_1/yes) * P(x_2/yes) * P(x_3/yes) * P(x_4/yes)$$

Constant (ignore) ← $P(x_1) * P(x_2) * P(x_3) * P(x_4)$

$$P(y=No/x_i) = P(No) * P(x_1/No) * P(x_2/No) * P(x_3/No) * P(x_4/No)$$

Constant (ignore) ← $P(x_1) * P(x_2) * P(x_3) * P(x_4)$

Let,

$$P(yes/x_i) = 0.13, \quad P(No/x_i) = 0.05$$

we know that, in classification

$\geq 0.5 \Rightarrow 1$

$< 0.5 \Rightarrow 0$

Normalization:

$$P(yes/x_i) = \frac{0.13}{0.13 + 0.05} = 0.72 = 72\%$$

$$P(No/x_i) = 1 - 0.72 = 0.28 = 28\%$$

note: we take the outlook, temp, play tennis dataset.

outlook

	yes	No	P(y)	P(N)
Sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5
Total =	9	5		

Temp

	yes	No	P(y)	P(N)
Hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cold	3	1	3/9	1/5
total =	9	5		

play

		P(y)	P(N)
yes	9		
No	5	(9/14)	(5/14)
Total	14		

Problem stat: we get a test data (Sunny, Hot)

predict the o/p ?

$$P(\text{yes} / \text{sunny}, \text{Hot}) = \frac{P(\text{yes}) * P(\text{Hot}/\text{yes}) * P(\text{Sunny}/\text{yes})}{P(\text{Sunny})\ P(\text{Hot})}$$

cut, this is const

$$= \frac{9}{14} * \frac{2}{9} * \frac{2}{9}$$

$$= 2/63 = 0.031$$

My $P(\text{No} / \text{Sunny}, \text{Hot}) = 0.085$

Now, performing normalization:

$P(\text{yes} / \text{sunny}, \text{Hot}) = 1 - 0.73 = 27\%$

$$P(\text{No} / \text{Sunny}, \text{Hot}) = \frac{0.085}{0.031 + 0.085} = 73\%$$

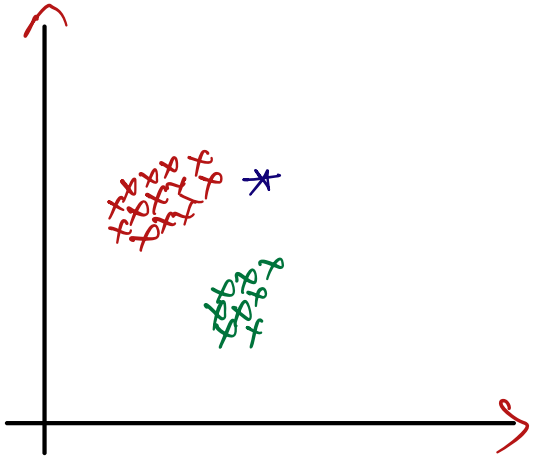
So, cut 'No' value is more, o/p is No

DONE!

KNN: (K- Nearest Neighbours)

→ Can be used to solve both classification and Regression problems

Classification problem:

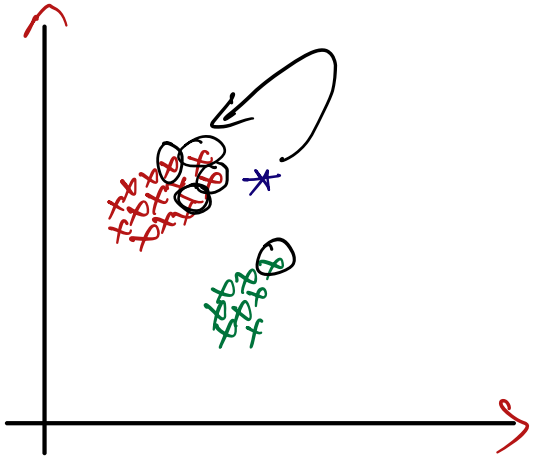


Steps:

→ we declare a value for K which is a hyper paramtre.....

let, k=5

⇒ this means, it will search for the nearest '5' points.



⇒ as you can see, it will go to the category which have more neighbours.

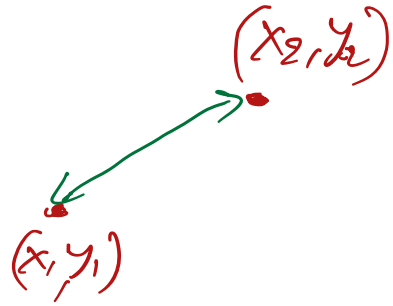
in this case, it is red

⇒ but how are we calculating that dist?

Eucledian Dist Manhatten Dist

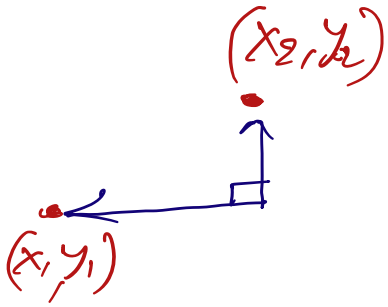
Euclidian Dist:

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

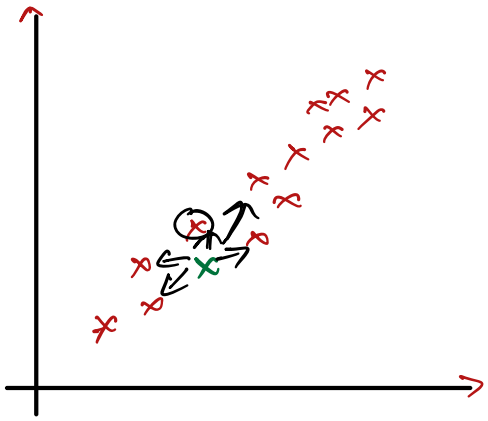


Manhatten Dist:

$$|x_1-x_2|+|y_1-y_2|$$



Regression:



it takes the avg of distances

We will check the o/p using Error Rate

note: ① KNN is not good when there are outliers

② " " " " " " is imbalanced data.