

* KNN

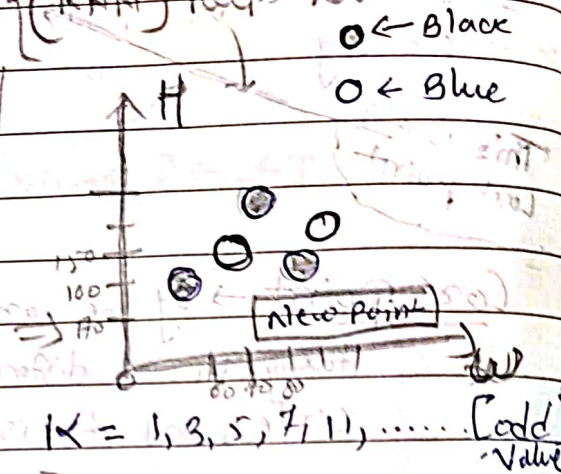
No of data point → Near neighbour

if I am asking New point
so, in which color it will
go (Black, or) Blue
we can't say.
(KNN) helps here

Why KNN?

→ Considering Dataset

X_1 (weight)	X_2 (Height)	Y
60	170	OB
70	160	OB
80	150	NO
90	140	OB
100	130	NO



* Note:- 12 N N

representing data over a graph

It is also a Hyperparameter $\Rightarrow K = (1, 3, 5, 7, \dots)$
(we can take K \uparrow)

- Steps:-
- I have to calculate distance to Each Data points
 - we have to choose 3 Nearest Distance
 - Then we have to find out Probability
 - Decide to which class it will belong

selecting
 $K=3$

given

(weight) X_1	(Height) X_2	0/1/0
60	175	OB
70	160	NO
75	155	OB
85	150	NO
95	180	OB

Distance

5

15.81

22.36

26.4

30.61

(175, 65)

(175, 60)

(160, 70)

(155, 75)

(150, 85)

(180, 95)

Euclidean D

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

find dist with

Each and Every

Data point

→ New Data point (belonging to these class)

OB
OB
NO

these new data point belongs to OB class

Probability

OB	NO
$\frac{2}{3}$	$\frac{1}{3}$

$\frac{2}{3}$ prob is more

Total 3 we are getting

K is hyperparameter
How to choose K value

$KNN \Rightarrow$ Classification = majority of class
Regression = mean value
In Regression we can select Even Value also

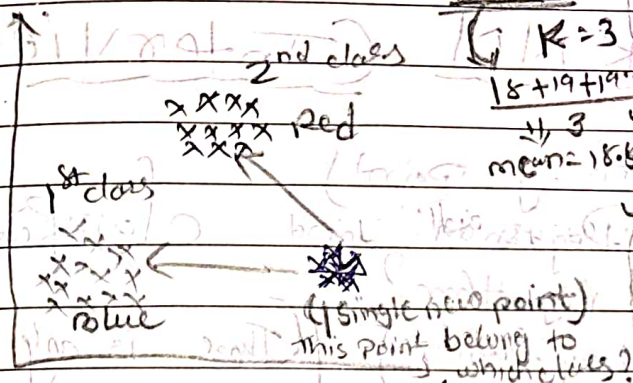
KNN :- calculation (calculating Distance)
Loss \rightarrow NO Loss

Optimization \rightarrow NO optimization

Hence, it is called as Lazy learner

K-NN Example

Regression Training in my Algorithm



Calculation
 $K=3$
 $\frac{18+19+19}{3}$
mean = 18.5

Regression Eg (mean value)

(x_1)	$x_2(H)$	$y(EMI)$
70	160	18
80	70	19
90	180	19.5
100	150	20
130	140	?

① \rightarrow first nearest will all points

② $\rightarrow K = \text{value} \rightarrow (1, 3, 5, 7, 9, \dots, n)$
(2, 4, 6, 8, \dots, n)
(we cannot even value)

(we can select Even value)

Why Not

Even Value

My New Data point belongs to which class?

\rightarrow we can't say,

So, we cannot take

Even value

Suppose $K=4$
Blue Blue
Red Red

Suppose $K=5$
B B
R R R

Here my new point belongs to R class

* Hamming Distance

$x_1 = x_2 \Rightarrow 0$
 $x_1 \neq x_2 \Rightarrow 1$

(x_1', x_2') (140, 60)
 (x_1'', x_2'') (160, 65)

Categorical Categorical Value
Here Dist

1	0	1
0	1	0
0	1	0
1	0	0

\Rightarrow 3

\Rightarrow If x_1 and x_2 are my categorical Data what we are going to use?

Distance =

K=3

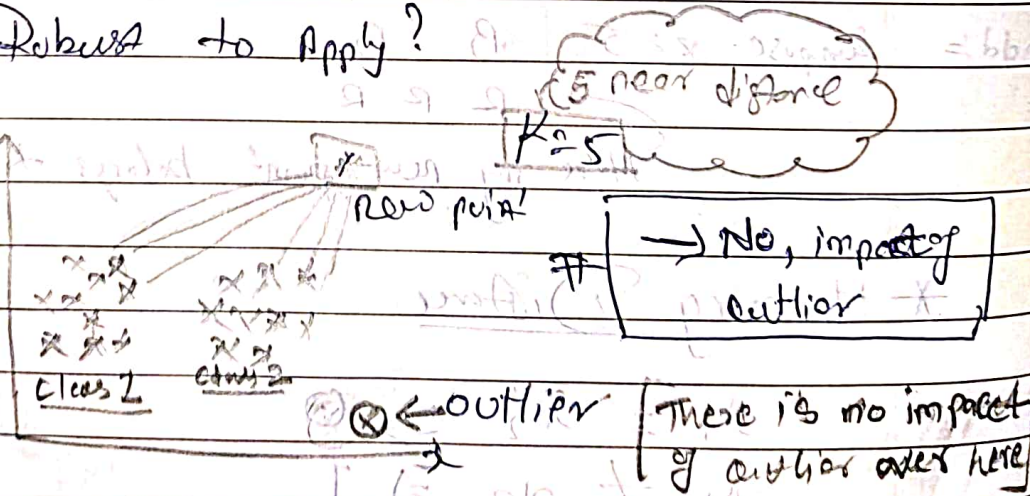
	X	Y	Hamming Dist			
1	AABC	0	1	\Rightarrow	<u>AABC</u>	AABC
2	ABCD	No	2		<u>AABC</u>	ABCD
3	ABBC	0	0		0(1)00 $\rightarrow 1$	00(1)1 $\rightarrow 2$
4	ACCD	No	3	(this point is obase)	ABBC	ABBC
5	AACD	0	3		ACCD	AACD
					0(1)11 $\rightarrow 3$	0(1)11 $\rightarrow 3$

All About KNN (Interview)

- ① \rightarrow KNN (K \rightarrow no. of Point)
- ② \rightarrow Used for both \rightarrow Regression and Classification
- ③ \rightarrow Is KNN is ML Algorithm?
 - \hookrightarrow No (It is lazy learner) (There is any Distance)
 - Lazy Learner because \rightarrow NO training
 - NO loss
 - NO optimization

- ④ \rightarrow Can we train Large Dataset?
 - \rightarrow NO \rightarrow (lots of memory)
 - \hookrightarrow lazy learner

- ⑤ Is it Robust to Apply?
 - \rightarrow Yes



- ⑥ In KNN do we required Scaling?
 - \Rightarrow Yes, because \rightarrow whenever distance is included
 - \rightarrow we required Scaling

Q. Can we use this for missing value? → "Yes"

→ Predict → weight height BMI

180 70 18

170 80 18.5

160 50 15.5

150 70 15.7

185 NA 20

NA 90 22.5

170 NA 21

NA 55 21.5

189 100 21.7

How we can impute missing value using KNN?

How we can predict the Null value

Treat Null values as Target but how?

→ Segregate Null and No Null values

Null

No Null values

This data will be my training data

180 70 18
 170 80 18.5
 160 50 15.5
 150 70 15.7

Training data means
 ↳ we are just fitting the data
 ↳ neighbor (x, y)
 ↳ Keeping the data
 we are not doing any loss or optimization

185 NA 20

Independent ↑ dependent

NA 90 22.5

dependent ↑ Independent

Q. If KNN is lazy learner how we can optimize it?

→ with the help of K-D Tree

we can optimize the distance

~~Brute force~~

* Tree → Pre order, Post order concept comes here

Page: _____

⑧ Can we reduce Time complexity of KNN
→ which method :- Kd-tree, Ball, Brute force
^cong Explain

Q) why KNN Algorithm Consider a "Lazy Learner"?

⇒ The KNN method simply saves the training data when it receives it; it does not learn and create a model. Instead of using the training data to discover any discriminative functions, instance-based learning is used, and the training data is also used for making predictions on dataset that have not yet been viewed.

Flow chart Explanation

$x_1, x_2, x_3, \dots, x_n$ | y

↓ It is just training Data

Data

Imp

over here Based on Distance we are going to decide this "New Point" belong to which class.

Whenever new point is coming it is just calculating Distance

New Point is coming (i.e. if test data is coming)

That's why is called Lazy Learner

* No loss / No Optimization

Here ⇒ "No Training" Data Over Here

→ we are just taking Data and calculating Distance