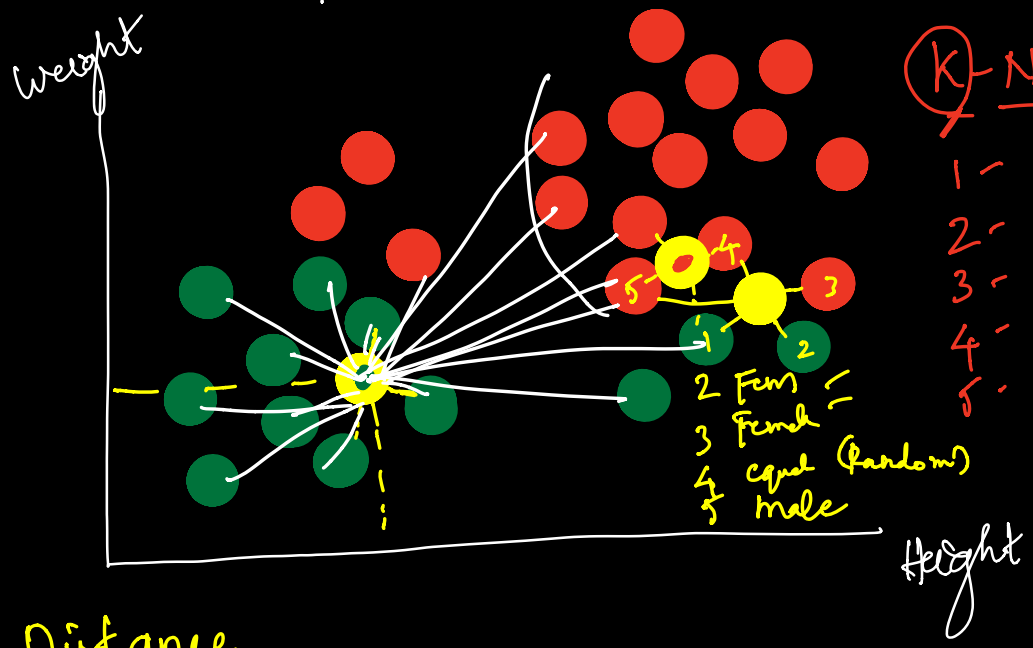
1. KNN Algorithm
2. Naive Bayes Algorithm

---

1. KNN

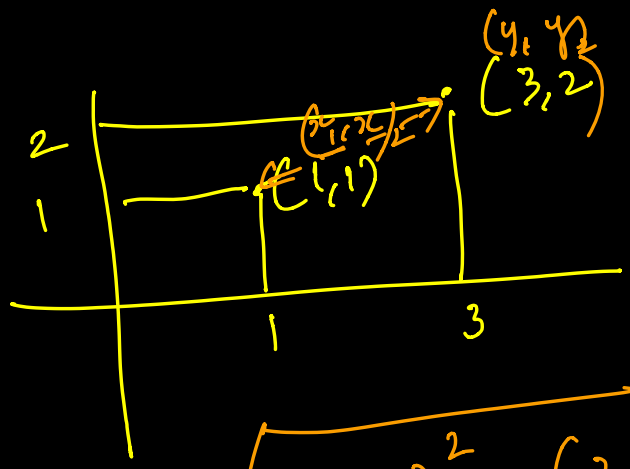
Supervised learning
Regression , classification
Simple , distance based Algorithm



Distance

Euclidean Distance ✓

$$\sqrt{\sum (x_i - y_i)^2}$$ ✓



$$\sqrt{(3-1)^2 + (2-1)^2}$$

$$\sqrt{2^2 + 1^2} = \sqrt{5}$$

$$= 2.414$$

Manhattan

$$\sum |x_i - y_i|$$

Minskowski

$$\sum_{i=1}^{k} \left(|x_i - y_i|^q \right)^{1/q}$$

$q = 2$

$$\sum \left((x_i - y_i)^2 \right) \quad (y_2)$$

Euclidean

$$\sum \sqrt{(x_i - y_i)^2}$$

K=2

K=3 Odd

Regression

	Customer	Age	Income	No of Credit	Class
②	George	35	35K	3	No (15.16)
①	Rachel	22	50K	2	Yes (15)
④	Stere	63	200K	1	No (152.23)
⑤	Tom	59	170K	1	No (122)
	John	37	50K	2	No?

$$\sqrt{(37-59)^2 + (50K - 170K)^2 + (2-1)^2}$$

122

K is too small → sensitive to the (overfit) noise points

K is too large → this might be more generic (Underfit)

$K < \text{sqrt}(n)$ → n => number of sample

Age	Loan	Default	Distance
25	40,000	N.	102000
35	60,000	N.	82,000
45	80,000	N. (3)	62,000
20	20,000	N.	1,22,000
35	120,000	N. (1)	22,000
52	18,000	Y	1,24,000
23	95,000	Y (2)	47,000
48	142000	? (N)	

K=3

0.62

25²

47000²

(6-i) Min-Max Scaling (Normalization)

Max → 1
Min → 0

Max → 142000
Min → 18000

$$X = \frac{X - min}{max - min}$$

$$\frac{142000 - 18000}{124000}$$

$$\frac{124000}{124000} = 1$$

$$\Rightarrow \frac{95000 - 18000}{124000}$$

$$= \frac{77000}{124000} = 0.620$$

Max = 52 Min = 20

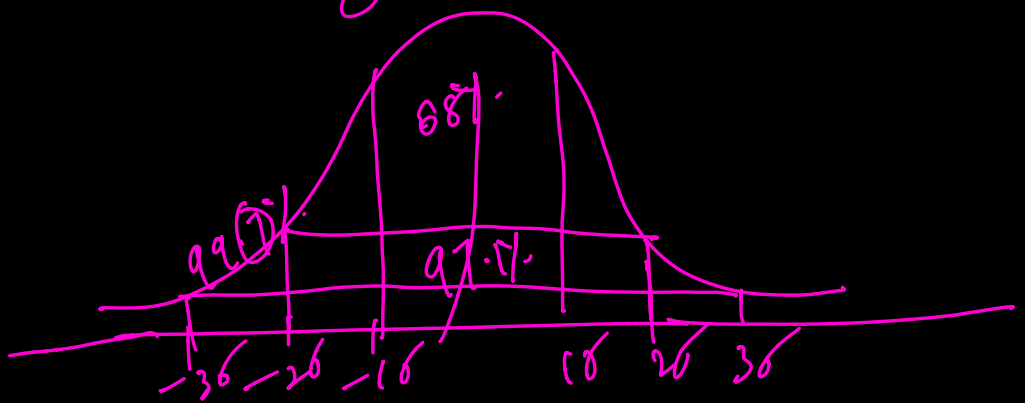
$$\frac{48-20}{52-20} = \frac{28}{32} = 0.875$$

Age	Loan	Distance	Default
0.1562	0.177	0.71	N
0.46	0.33	③ 0.35	N
0.78	0.5	① 0.09	N
0	0.01	0.87	N
0.46	0.82	0.40	N
1	0	② 0.12	Y
0.093	0.62	0.78	Y
0.87	1	~~2~~	N

Standardization

$$Z = \frac{X - \mu}{\sigma}$$

Z-Score



99.7% ⇒ −3 to +3

−5 +5

Box plot

Z-Score

−5 +5

Min-Max

✓

1

Standardization ✓

| +5

−3 +3

Linear Regression

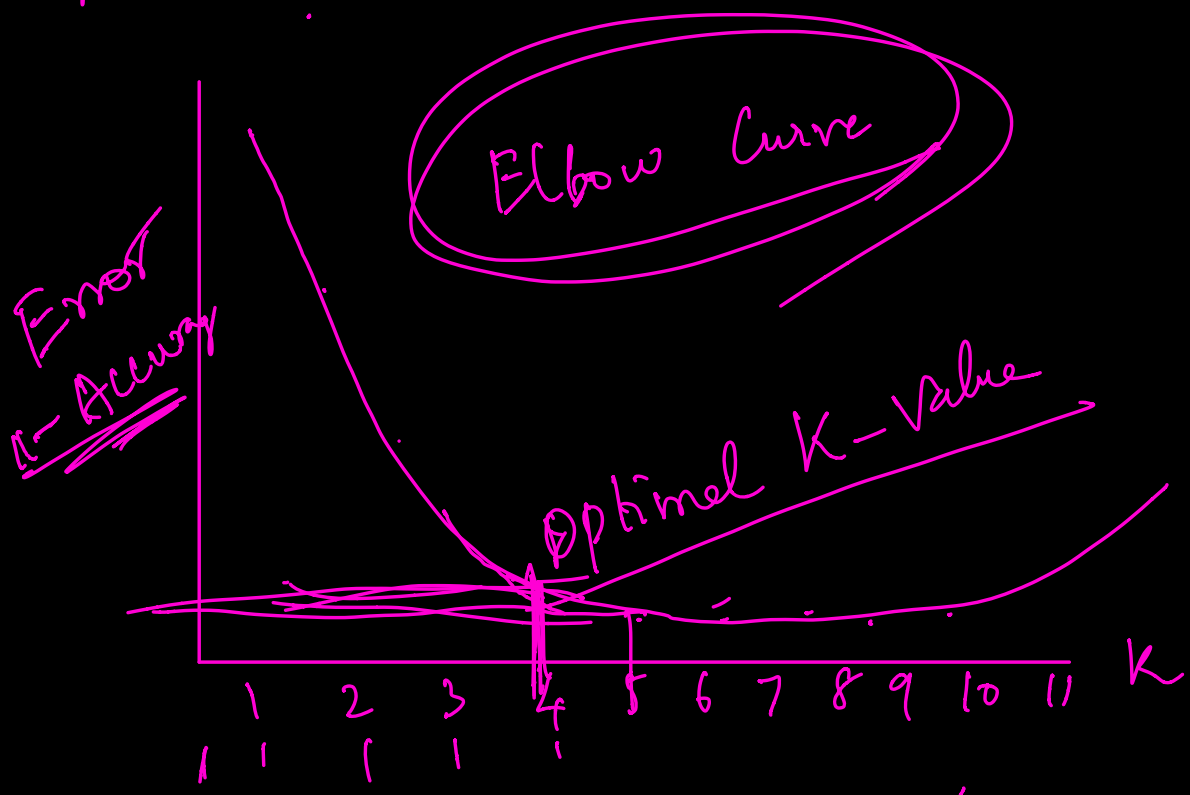
model => $mx + b$ → Predict

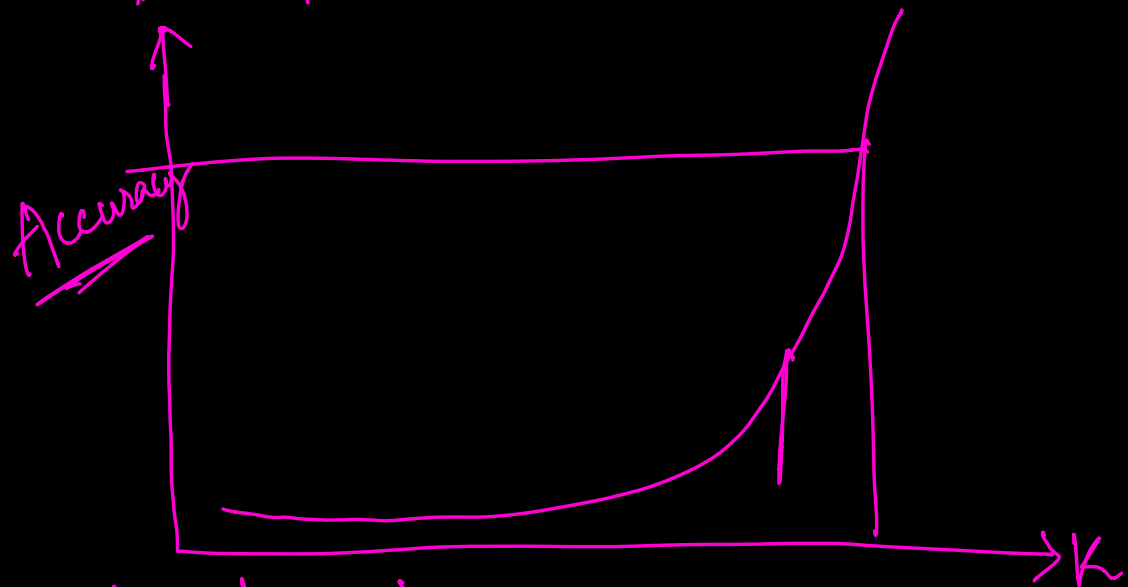
KNN

model => No model

Lazy model

K = ?





Disadvantages:
more time to classify

Need to calculate lot of distance

Choosing right k is tricky

Need large Number of samples

KD Tree can be used to reduce the complexity of time

Naive Bayes Algorithm

Supervised Algorithm

classification

Naive Bayes classifier works on the principle of conditional probability as given by Bayes theorem

Toss two coins :

$SS = \{HH, HT, TT, TH\}$

$P(\text{Getting two Heads}) = \frac{1}{4}$

P(Second coin being head given first coin is tail)

$= \{TT, TH\}$

$= \frac{1}{2}$

P(getting two heads given first coin is head)

$= \{HH, HT\}$

$= \frac{1}{2}$

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

P(Second coin being head given first coin is tail) =

P(A/B) {HH, HT, TH, TT}

A = Second coin being head

B = first coin is tail

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

$$P(A) = \frac{2}{4}$$

$$P(B) = \frac{2}{4}$$ {HH, HT, TH, TT}

P(B/A) = first coin is tail given second coin being head

$$\{\bar{H}H, TH\}$$

$$= 1/2$$

$$\frac{\frac{2}{4} * \frac{1}{2}}{\frac{2}{4}} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2}}$$

$$\frac{1}{4} \times \frac{2}{1}$$

$$= \frac{1}{2}$$

Class Prior Probablity

Likelihood

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

Posterior Probablity

Predictor Prior Probablity.

A = Target

B = Independant Data

1. Day = Holiday
   Discount = yes
   Free Delivery = Yes

P(Buy = Yes) P(Buy = No)

P(Buy = Yes | Day = Holiday, Discount = Yes, Free Delivery = Yes)

P(Buy = Yes) * P(Day = Holiday | Buy = Yes) * P(Discount = Yes | Buy = Yes) * P(Free Deliv = Yes | Buy = Yes)

P(Day = Holiday) * P(Discount = Yes) * P(Free Deliv = Yes)

Likelihood Tables

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy		
		Yes	No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy		
		Yes	No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

Calculating Conditional Probability of purchase on the following combination of da
discount and free delivery:

Where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let A = Buy

$$P(A|B) = P(\text{Yes Buy} \mid \text{Discount = Yes, Free Delivery = Yes, Day = Holiday})$$

$$= \frac{P(\text{Discount = Yes} \mid \text{Yes}) * P(\text{Free Delivery = Yes} \mid \text{Yes}) * P(\text{Day = Holiday} \mid \text{Yes}) * P}{P(\text{Discount=Yes}) * P(\text{Free Delivery=Yes}) * P(\text{Day=Holiday})}$$

$$= \frac{(19/24) * (21/24) * (8/24) * (24/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.986$$

$$\frac{\frac{24}{30} * 8/24 * 19/24 * 21/24}{11/30 * 20/30 * 23/30}$$

$$\Rightarrow 0.986 =$$

$P(\text{NoBuy} / - - - -)$

$= 0.178$ ..

$0.$

0.629

Play = yes

0.3107 ~~Play =~~

0.628

0.62156

$P(\text{No}) = 0.212$