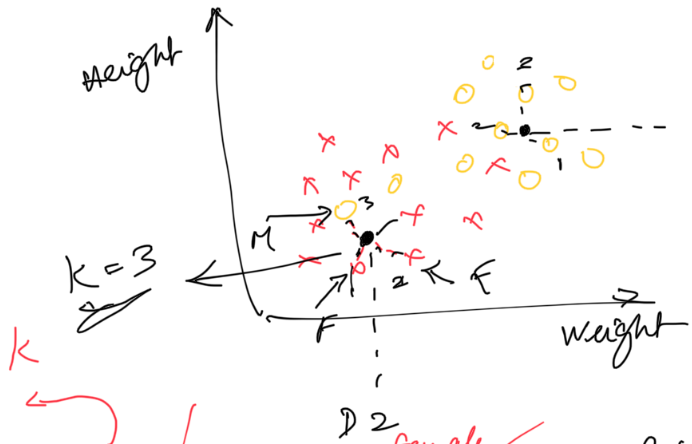
KNN

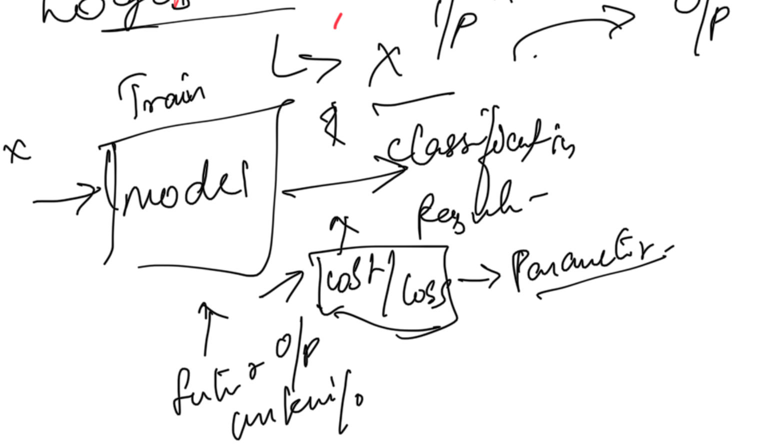


x – Female
o – Male

2F/1F => F
M/M/M => 3M

K-Nearest Neighbours!!

Logistic

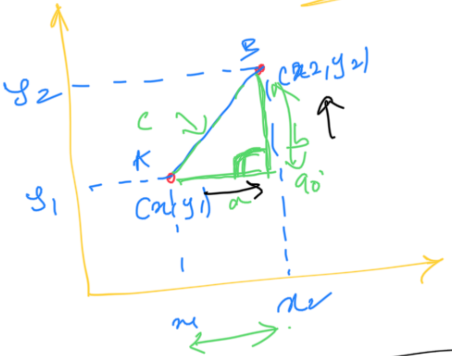


K.N.N

Distance – Estimate

Distance method

Distance



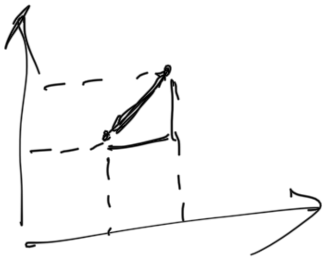
$$a^2 + b^2 = c^2$$

$$\rightarrow (x_2 - x_1)^2 + (y_2 - y_1)^2$$

$$\Rightarrow c^2$$

Euclidean distance $\Rightarrow$ $\boxed{c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$

Manhattan distance $\rightarrow$ $|(x_2 - x_1)| + |(y_2 - y_1)|$



Customer	Age	Income	# of Credits	Classifier	
② George	35	35K	3	N	15.16
① Rachael	22	50K	2	Y	15
Steve	63	200K	1	N	152
③ Tom	59	170K	1	N	122
John	37	50K	2	? N	

⇒ John

$$\Rightarrow \sqrt{(37-59)^2 + (50{,}000 - 170{,}000)^2 + (2-1)^2}$$

$$\Rightarrow \sqrt{(22)^2 + (12014)^2 + (1)^2}$$

$$\Rightarrow \sqrt{484 + 144{,}000 + 1}$$

① [4, N, N] → N

$\boxed{K = 3}$ 5

~~Case~~ 1

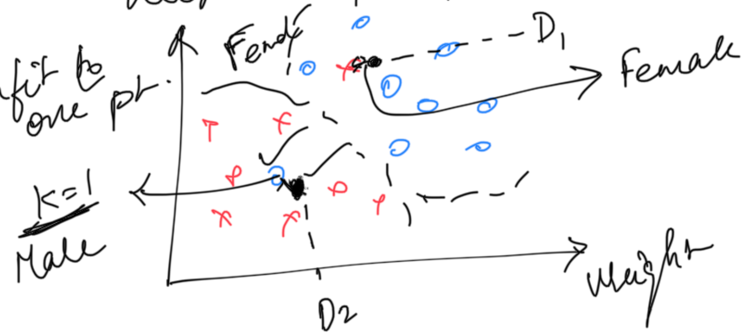
① $K = 1$ → classification is closer Rachael.

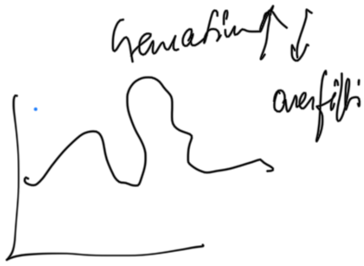
$K = 1$ = Male.

$K - N - N$ Male

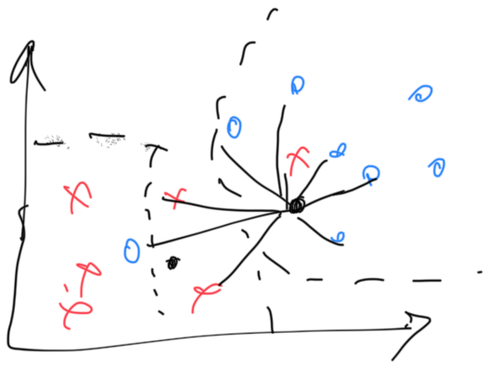
O/p → Y real

Overfit to one pt.





Case 2:



$K = 11$

More Generalized

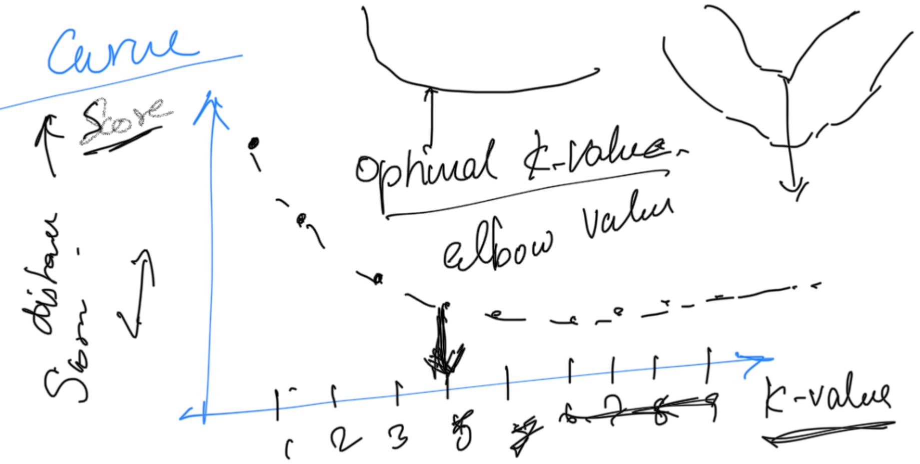
Underfitting

Case ⇒ k = 2 There is possibility of ambiguity

Model → Clear decision

→ Random class.

Elbow Curve



Case ⇒ IV

K = 3

Age	Loan	Default	Distance	
25	40,000	N	102K	N
35	60,000	N	82K	N
45	80,000	N	62K (3)	N
20	20,000	N	122K	N
35	120,000	N	22K (1)	N
52	18,000	Y	124K	Y
23	95,000	Y	47K (2)	Y
48	142,000	?		

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

N

142

142,000

1,20,000

$(48-23)^2$

$(48 - 23)^2$ ↑ 9 ↑ 142,000

① → 120,000
95,000
② → 80,000
③

→ We should → Scaling & Normalization

→ ① Min-Max Scaler
→ ② Standard — Standard-Scaler.

$$X_i = \frac{X - X_{min}}{X_{max} - X_{min}} \rightarrow 0 \rightarrow 1$$

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

Min - Max Scaler.

→ Max → 142,000
→ Min → 18,000

$$\frac{80,000 - 18,000}{142,000 - 18,000} = 0.5$$ ✓

↳ can be outliers also

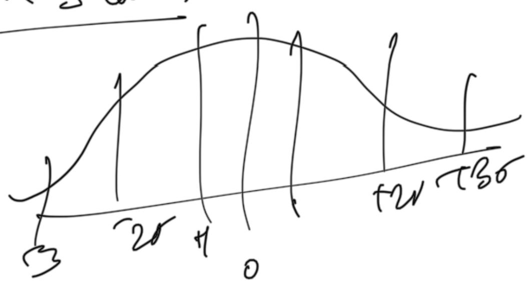
Min-Max → Influenced by outliers. !!!

Standard Scaler

$Z = X - \mu$

Standard Scaler

$$z = \frac{x - \mu}{\sigma}$$



Age	Loan	Distance	
~~0.151~~	0.178	0.71	N → ③
0.468	0.338	0.35	N
0.261	0.5	0.09	N → ①
0	0.016	0.87	N
0.468	0.872	0.40	N
1	0	0.12	Y → ②
0.093	0.000	0.78	Y
0.873	1		

→ effect of higher magnitude will be reduced

→ [N, N, Y] → N

→ ① faster
② Equal Ground → calculate

↳ negates value advantage (units, tens, thousands).

⇒ Robust Scaler:

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$

⇒ Robust Scaler:-

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$

⇒ Minkowski:-

p 3, 4, 5

$$\Rightarrow \left[\sum_{i=1}^{n} |x_i - y_i|^{p}\right]^{1/p} \qquad \sqrt{} \; 1/2$$

$p = 2$

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

$p = 1$ → Manhattan Distance

$k = 5, 7, 3$

Lazy algorithm/model.