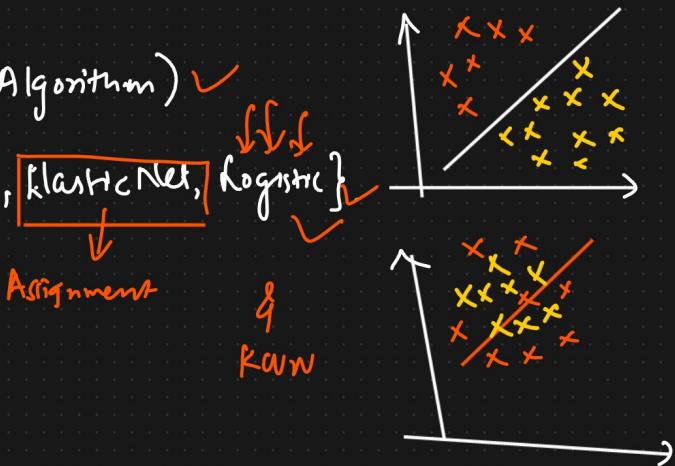


Agenda

- ① Naive Bayes Algorithm (Classification Algorithm) ✓
- ② Practicals {Linear Regression, Ridge, Lasso, **ElasticNet**, Logistic} ✓
- ③ KNN ✓



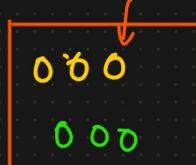
① Naive Bayes Intuition

Rolling a Dice {1, 2, 3, 4, 5, 6}

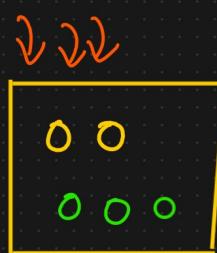
$$P(1) = \frac{1}{6} \quad P(2) = \frac{1}{6} \quad P(3) = \frac{1}{6}$$

Independent Events

Dependent Events



First Event
 $P(y) = 3/6 = \frac{1}{2}$



$$\rightarrow P(a/y) = \frac{3}{5} = \frac{3}{5}$$

Dependent ↓ ↓

$$P(y \text{ and } a) = P(y) * P(a/y)$$

$$= \frac{1}{2} * \frac{3}{5} = \frac{3}{10}$$

Naive Bayes

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

↑↑

Conditional Probability

A and B

B and A

Dependent Events

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

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

000
00

$$P(\omega)$$

$$P(0|\omega)$$

000
00

$$P(0)$$

$$P(\omega|0)$$

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

Bayes' Theorem

Dataset

I/P features

$$x_1 \ x_2 \ x_3 \ x_4 \ - \ - \ x_n$$

$$y \nearrow 0/p$$

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

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

$$P(x_1) * P(x_2) * P(x_3) * \dots * P(x_n)$$

Binary dataset

$$x_1 \ x_2 \ x_3 \ 0/p$$

Yes

No

$$P(x_1, x_2, x_3 / \text{Yes})$$

$$P(y=\text{Yes} / x_i) = \frac{P(y_{ci}) * P(x_1/y_{ci}) * P(x_2/y_{ci}) * P(x_3/y_{ci})}{P(x_1) * P(x_2) * P(x_3)}$$

Constant $\rightarrow P(x_1) * P(x_2) * P(x_3) \leftarrow \# \text{ fixed}$
 Ignore

$$P(y = N_0 | x_i) = P(N_0) * P(x_i | N_0) * P(x_2 | N_0) * P(x_3 | N_0)$$

Constant $\rightarrow P(x_1) * P(x_2) * P(x_3) \leftarrow \# \text{ fixed}$
 Ignore

let say

$$P(y_{as} | x_i) = 0.13$$

$$P(N_0 | x_i) = 0.05$$

$$P(y_{as} | x_i) = \frac{0.13}{0.13 + 0.05} = \frac{0.13}{0.18} = 72\% \quad < 50\%$$

$$P(N_0 | x_i) = \frac{0.05}{0.13 + 0.05} = 0.28 = 28\% \quad \Rightarrow 100\%$$

DATASET

Naive Bayes

Outlook

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

	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

Temperature (Overcast, Cool)

	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
	9	5		

	Yes	No	$P(y_{as})$	$P(N_{as})$
	9	5		
	9/14	5/14		
	9/14	5/14		

$\rightarrow (\text{Sunny}, \text{Hot}) \rightarrow \text{O/p Naive Bayes} \Rightarrow \text{O/p } \underline{\text{No}}$

$$P(\text{Yes} | (\text{Sunny}, \text{Hot})) = \frac{P(\text{Yes}) * P(\text{Sunny} | \text{Yes}) * P(\text{Hot} | \text{Yes})}{P(\text{Sunny}) * P(\text{Hot})} = \frac{\frac{9}{14} * \frac{2}{5} * \frac{1}{2}}{\frac{4}{7} * \frac{9}{14}} = \frac{2}{63} = 0.031$$

$$P(\text{No} | (\text{Sunny}, \text{Hot})) = \frac{P(\text{No}) * P(\text{Sunny} | \text{No}) * P(\text{Hot} | \text{No})}{P(\text{Sunny}) * P(\text{Hot})}$$

$$= \frac{7}{14} * \frac{3}{5} * \frac{2}{8}$$

$$= \frac{3}{35} = 0.0857$$

$P(\text{Yes} | (\text{Sunny}, \text{hot})) = 0.031$

$P(\text{No} | (\text{Sunny}, \text{hot})) = 0.0857$

Real Probability

$$P(\text{Yes} | \text{Sunny, hot}) = \frac{0.031}{0.031 + 0.0857}$$

$$= 27\%$$

$$P(\text{No} | \text{Sunny, hot}) = 1 - 0.27 = 73\%$$



X	Y
$x_1 \ x_2 \ x_3 \ x_4$	O/P

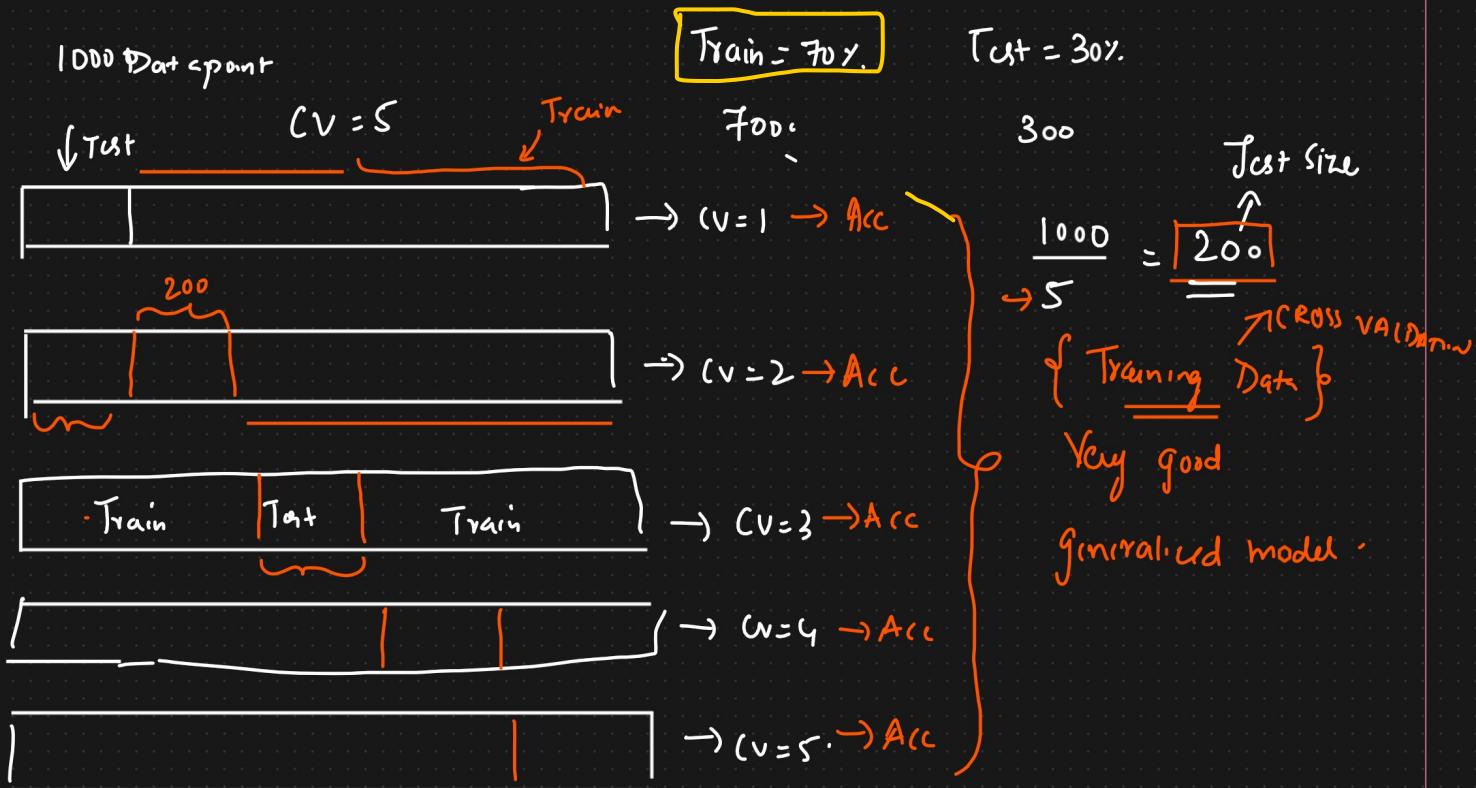
NL Model

\Downarrow

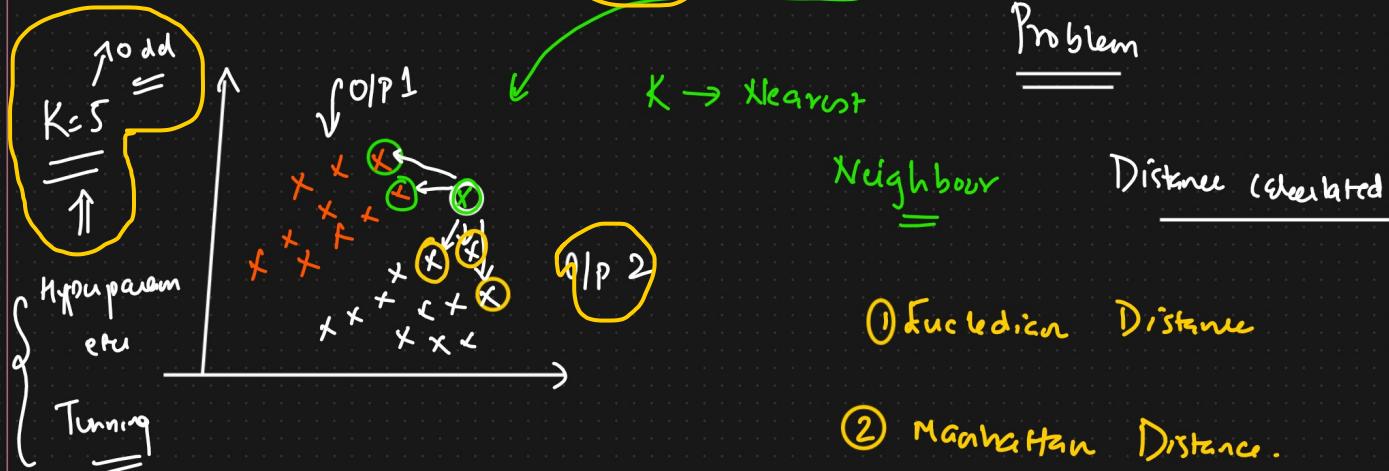
Just Data { Train Test Split }

Dataset { Train \leftarrow 70%,
Test \leftarrow 30% }

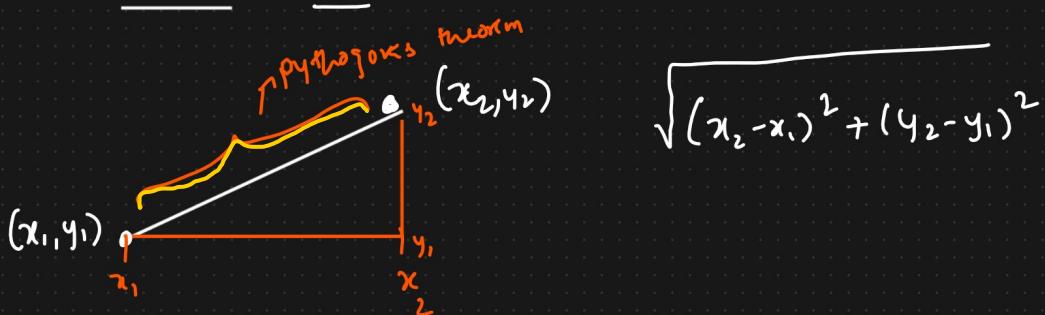
CROSS VALIDATION



④ K Nearest Neighbour (KNN) → Classification and Regression



① Euclidean Distance



(x_1, y_1, z_1)

(x_2, y_2, z_2)

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

②

Manhattan



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



IRON MAN

GPS

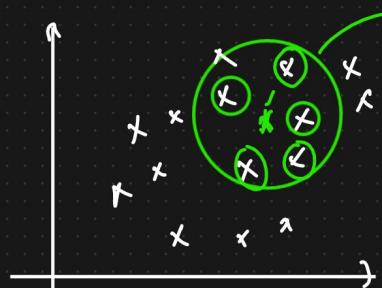
↓

USA



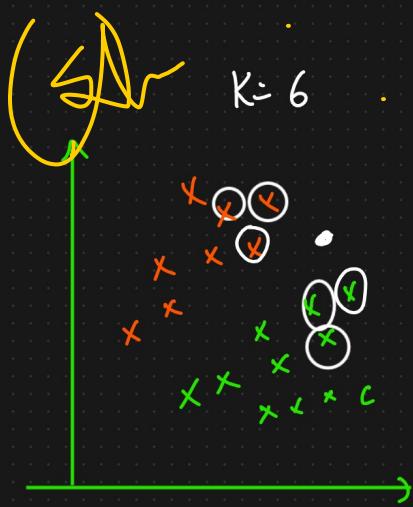
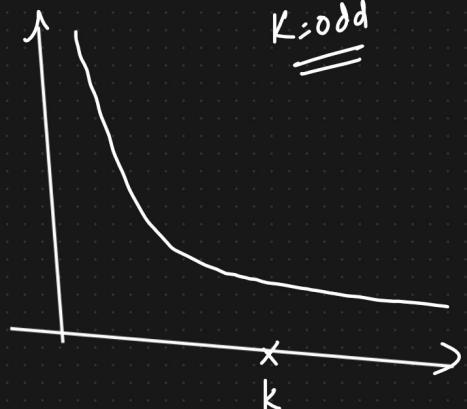
Regression

$$K=5 \rightarrow$$



$$\begin{matrix} \text{Metrics} \\ \hline \text{MSE} \end{matrix}$$

$K=odd$



$K=1 \rightarrow \text{Accuracy } \downarrow$

$K=2 \rightarrow \text{ " }$

$K=3 \rightarrow \text{ " }$

!

}