Agenda

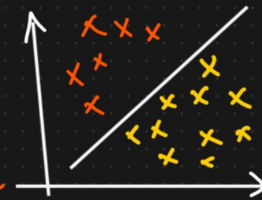
① Naive Baye's Algorithm (Classification Algorithm) ✓

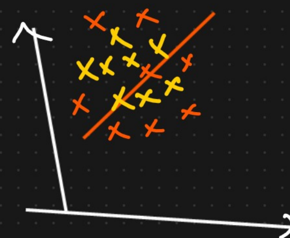
② Practicals {Linear Regression, Ridge, Lasso, $\boxed{\text{ElasticNet,}}$ Logistic} ✓

Assignment

& KNN

③ KNN ✓





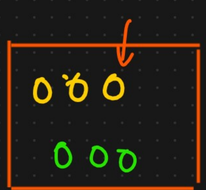
① Naive Baye's Intuition

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

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

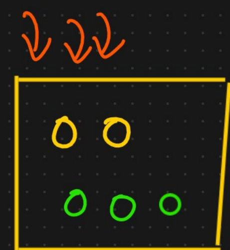
Independent Events

Dependent Events



First Event

$P(y) = 3/6 = \frac{1}{2}$ ✓



$\to P(G/y) = \frac{3}{5} =$

Dependent

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

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

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

Conditional Probability → Naive Baye's

Dependent events

A and B $\qquad$ B and A

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

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

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

→ Bay's Theorem

$P(w)$

$P(o|w)$

$P(o)$

$P(w|o)$

Dataset

$$\underbrace{X_1 \quad X_2 \quad X_3 \quad X_4 \quad - \quad - - \quad X_n}_{\text{I/P features}} \longleftrightarrow y \quad \text{O/P}$$

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

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

Binary dataset

x_1	x_2	x_3	O/P
			Yes
			No

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

$$P(y = \text{Yes}/x_i) = \frac{P(\text{yes}) * P(x_1/\text{yes}) * P(x_2/\text{yes}) * P(x_3/\text{yes})}{\cancel{P(x_1) * P(x_2) * P(x_3)}} \leftarrow \text{\# fixed}$$

Constant

Ignore

$$P(y=No/x_i) = \frac{P(No) * P(x_i/No) * P(x_2/No) * P(x_3/No)}{\cancel{P(x_1) * P(x_2) * P(x_3)}} \leftarrow \# \text{ fixed}$$

Constant → Ignore

Let say $\boxed{P(yes/x_i) = 0.13}$ $\boxed{P(No/x_i) = 0.05}$

$$\overset{1}{P(yes/x_i)} = \frac{0.13}{0.13+0.05} = \frac{0.13}{0.18} = 72\%$$

$$\overset{0}{P(No/x_i)} = \frac{0.05}{0.13+0.05} = 0.28 = 28\%$$

≥50%

<50%

⇒ 100%

DATASET

Naive Bayes

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
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

Outlook

	Yes	No	P(Y)	P(N)
Sunny	2	3	$\boxed{2/9}$	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5
	9	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		

PLAY

		P(Yes)	P(No)
Yes	9	$\boxed{9/14}$	$\boxed{5/14}$
No	5		
	14		

→ (Sunny, Hot) → O/p Naive Bayes ⇒ O/p No

$$P(Yes/(Sunny, Hot)) = \frac{P(Yes) * P(Sunny/Yes) * P(Hot/Yes)}{\cancel{P(Sunny) * P(Hot)}} = \frac{\cancel{9}}{14_7} * \frac{\cancel{2}^1}{\cancel{9}} * \frac{2}{\cancel{9}}$$

$$= \frac{2}{63} = 0.031$$

$$P(No/(Sunny, Hot)) = \frac{P(No) * P(Sunny/No) * P(Hot/No)}{\cancel{P(Sunny) * P(Hot)}}$$

$$= \frac{\cancel{5}}{\cancel{14}_7} * \frac{3}{\cancel{5}} * \frac{\cancel{2}^1}{\cancel{8}}$$

$$= \frac{3}{35} = 0.08571$$

Real Probability

$P(Yes/(Sunny, Hot)) = 0.031$

$P(No/Sunny, Hot) = 0.0857$

$$P(Yes/(Sunny, Hot)) = \frac{0.031}{0.031 + 0.0857}$$

$$= 27\%$$

$$P(No/Sunny, Hot) = 1 - 0.27 = 73\%$$

DATASET

1000

X				Y
x_1	x_2	x_3	x_4	O/P

ML Model

⇓

Test Data { Train Test Split }

Dataset { Train ← 70%, Test ← 30% }

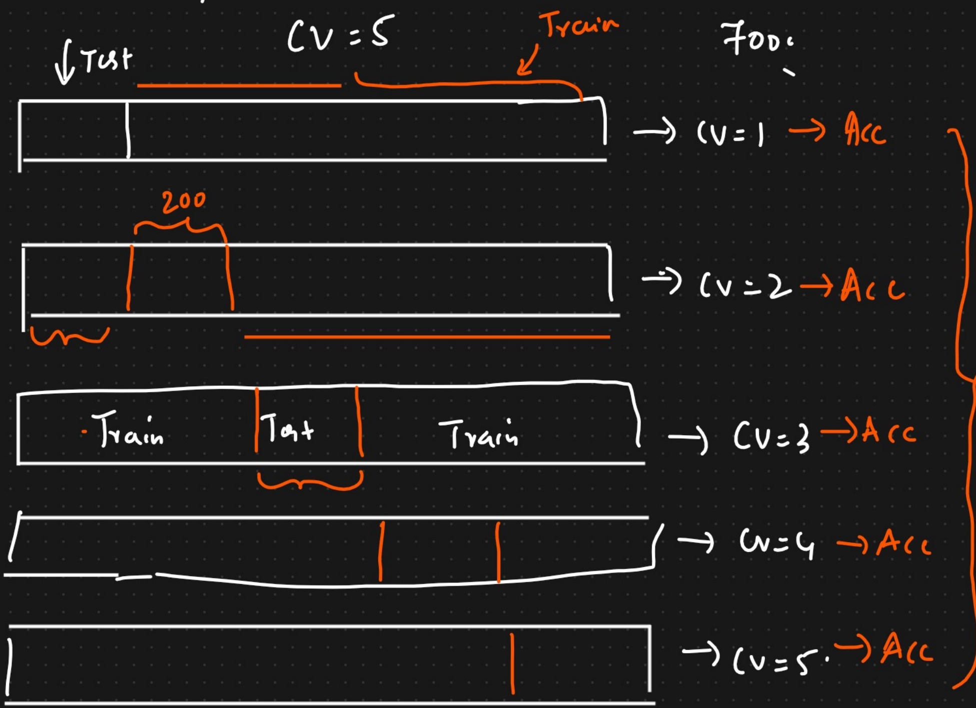
CROSS VALIDATION

1000 Datapoint

Train = 70%   Test = 30%

700   300

CV = 5



$\frac{1000}{5} = \boxed{200}$ → Test Size

{ Training Data } → CROSS VALIDATION

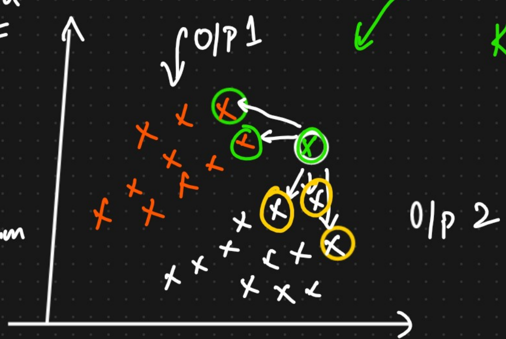
Very good generalized model.

① K Nearest Neighbour (KNN) → Classification and Regression Problem

K → Nearest Neighbour Distance Calculated

K=5 → odd

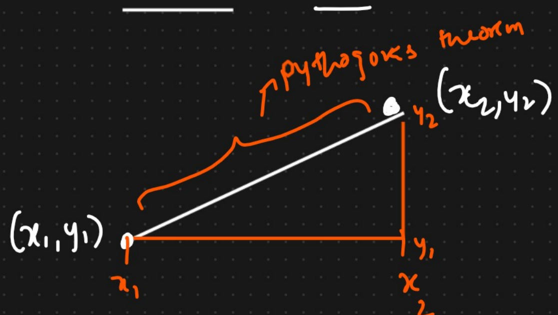
Hyperparameter Tuning



① Eucledian Distance

② Manhattan Distance.

① Eucledian Distance



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

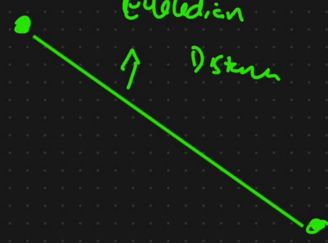
(x_2, y_2, z_2)

(x_1, y_1, z_1)

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

② Manhattan

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



IRON MAN

GPS

↓

USA

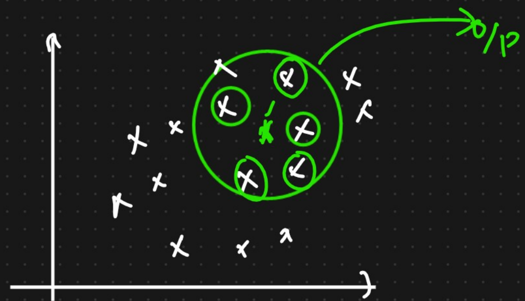
Regression

$\boxed{K=5}$ →

Metrics

↑

$\boxed{MSE}$



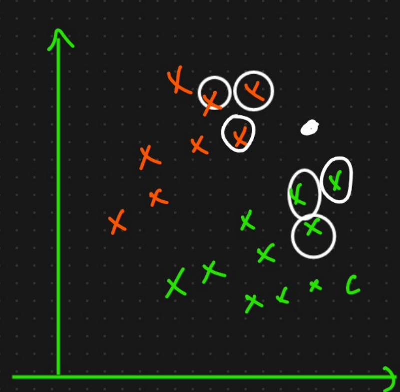
K=1 → Accuracy ↙

K=2 → "

K=3 → "

⋮

K=6



K=odd