3rd Day → Machine Learning Algorithms

Agenda

① Practicals → Simple Examples
② Naive Bayes Intuition
③ KNN algorithms

Previous Session

① Linear Regression
② Ridge & Lasso
③ Logistic Regression

⇒ Complex

① Naive Bayes Intuition {Classification}

⇓

{BAYES THEOREM}

Rolling a Dice

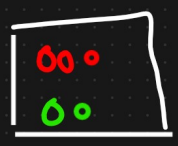
{1, 2, 3, 4, 5, 6}

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

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

{Independent Events}

Dependent Event



First Event ✓

$P(R) = \frac{3}{5}$ → R ✓

Given Marble

$P(G) = \frac{2}{4} = \frac{1}{2}$ → G

Dependents

$P(G) = \frac{2}{5}$

$P(R) = \frac{3}{4}$

$P(R \text{ and } G) = P(R) * P(G/R)$ → Conditional probability

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

$\Rightarrow$ P(A and B) = P(B and A) {Yes}

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

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

Bayes Theorem

(RUX Naive Bayes

I/P

X_1 X_2 X_3 X_4 X_5 ... X_n → y o/p

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

Yes or No

$$= \frac{P(y) * P(x_1/y_1) * P(x_2/y_2) * P(x_3/y_3) \cdots P(x_n/y_n)}{P(x_1) * P(x_2) * P(x_3) * \cdots P(x_n)}$$

DATASET

x_1	x_2	x_3	x_4	y
				Yes ✓
				No ✓

$$P(y=Yes/x_i) = \frac{P(Yes) * P(x_1/Yes) * P(x_2/Yes) * P(x_3/Yes) * P(x_4/Yes)}{P(x_1) * P(x_2) * P(x_3) * P(x_4)}$$

Constant → Ignore # fixed

$$P(y=No/x_i) = \frac{P(No) * P(x_1/No) * P(x_2/No) * P(x_3/No) * P(x_4/No)}{P(x_1) * P(x_2) * P(x_3) * P(x_4)}$$

Constant → # fixed

x_i → Yes / No

$P(Yes|x_i) = 0.13$ $\quad P(No|x_i) = 0.05$

$\Downarrow$ $\quad \geq 0.5 \Rightarrow 1$

$< 0.5 \Rightarrow 0$

$$P(Yes|x_i) = \frac{0.13}{0.13+0.05} = 0.72 = 72\%$$

$P(No|x_i) = 1 - 0.72 = 0.28 = 28\%$

DATASET

Binary class

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

x_1 Outlook — $P(Sunny/Yes)$

	Yes	No	P(y)	P(No)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rain	3	2	3/9	2/5
Total	9	5		

Temperature

	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
Total	9	5		

PLAY

		P(Yes)	P(N)
Yes	9	9/14	5/14
No	5		
Total	14		

→ Test (Sunny, Hot) → O/P

$$P(Yes|(Sunny, Hot)) = P(Yes) * P(Sunny/Yes) * P(Hot/Yes)$$

$P(\text{Sunny}) * P(\text{Hot})$

$= \frac{1}{14} * \frac{2}{1} * \frac{2}{9}$

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

$$P(\text{No} \mid \text{Sunny, Hot}) = \frac{P(\text{No}) * P(\text{Sunny} \mid \text{No}) * P(\text{Hot} \mid \text{No})}{P(\text{Sunny}) * P(\text{Hot})} \rightarrow \text{Constant}$$

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

$= \frac{3}{35} = 0.085$

$P(\text{Yes} \mid \text{Sunny, Hot}) = 0.031 \quad = 1 - 0.73 = 0.27 = 27\%$

$P(\text{No} \mid \text{Sunny, Hot}) = 0.085 = \frac{0.085}{0.031 + 0.085} = 0.73 = 73\%$

$\rightarrow$ (Sunny, hot) $\longrightarrow$ Yes or No

Answer $\rightarrow$ No ✓

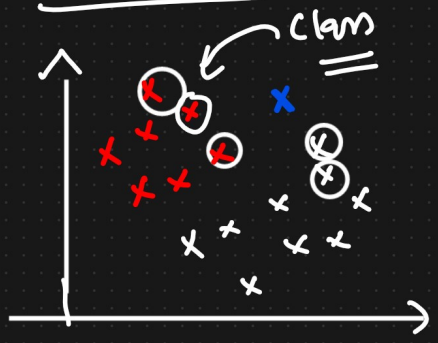
Assignment

('Overcast', 'Mild') $\longrightarrow$ Naive Bayes ?

② KNN Algorithm {K Nearest Neighbour}

- Classification
- Regression

① Classification



K Nearest Neighbour

K = 5

Maximum No

Red = 3

White = 2

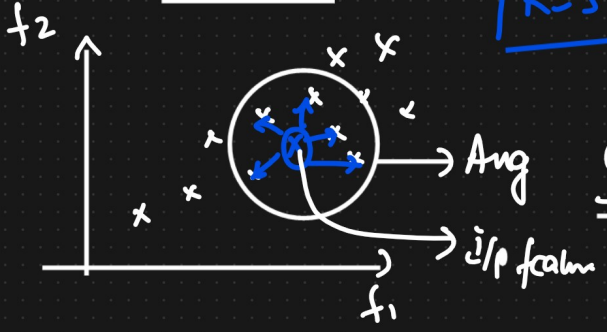
K = 1 to 20

↑

K = 5 → Hyperparameter

Error Rate ↑↑.

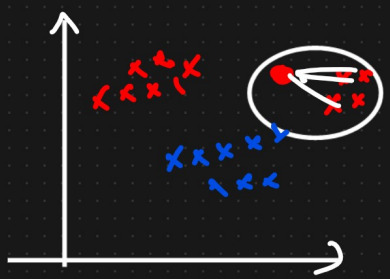
Regression



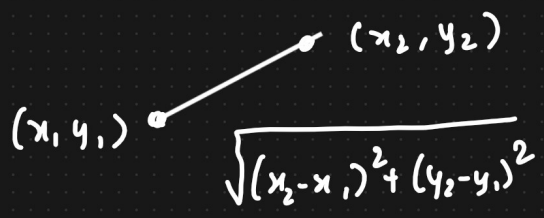
K Nearest

① Outliers

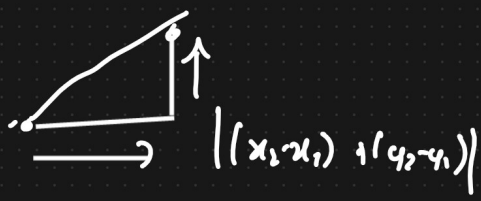
② Imbalanced



Euclidean Distance ✓



Manhattan Distance ✓



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