Naive Baye's Algorithm (Classification)

① Probability [Independent And Dependent Events]

② Baye's Theorem

③ Naive Baye's Mathematical Intuition.

④ Variants of Naive Baye's

① Probability [Independent And Dependents Events]

Independent Events

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

$Pr(1) = \frac{1}{6}$ $Pr(2) = \frac{1}{6}$ $Pr(3) = \frac{1}{6}$.

Tossing a Coin. {H, T}

$Pr(H) = \frac{1}{2}$ $Pr(T) = \frac{1}{2}$.

Dependent Event

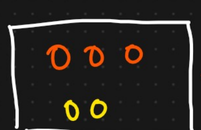
① What is the probability of first removing a orange marble and then removing a yellow marble

1st Event

↓

$Pr(O) = 3/5$

$\boxed{Pr(Y/O) = \frac{2}{4}}$ ⇒ Conditional Probability



$Pr(O \text{ and } Y) = Pr(O) * Pr(Y/O)$ ⇒ Conditional probability.

$= 3/5 * 2/4 = \frac{\cancel{6}\,3}{\cancel{20}\,10} = \boxed{3/10}$

$$\boxed{Pr(A \text{ and } B) = Pr(A) * Pr(B/A)}$$
⇓
Dependent

$$\boxed{Pr(A \text{ or } B) = Pr(A) + Pr(B)}$$
⇓
InDependent

② Baye's Theorm

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

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

$$\boxed{P(A/B) = \frac{Pr(A) * Pr(B/A)}{Pr(B)}}$$ ⇒ Baye's Theorm

$Pr(A/B)$ = Probability of Event A given B has occured

$Pr(A)$ = Probability of Event A

$Pr(B)$ = " " " B

$Pr(B/A)$ = Probability of Event B given A has occured.

Dataset

x_1	x_2	x_3	O/P ⇒ y
–	–	–	Yes
–	–	–	No
–	–	–	Yes
–	–	–	No
–	–	–	Yes

$$Pr(\overset{A}{y}/\overset{B}{(x_1, x_2, x_3)}) = \frac{Pr(y) * Pr(x_1, x_2 x_3)/y)}{Pr(x_1, x_2, x_3)}$$

$$= \frac{Pr(y) * Pr(x_1/y) * Pr(x_2/y) * Pr(x_3/y)}{Pr(x_1) * Pr(x_2) * Pr(x_3)}$$

$$Pr(y = yes / (x_1, x_2\, x_3)) = \frac{Pr(yes) * Pr(x_1/yes) * Pr(x_2/yes) * Pr(x_3/yes)}{\cancel{Pr(x_1) * Pr(x_2) * Pr(x_3)}}$$ Remove

$$Pr(y = No \mid (x_1, x_2, x_3)) = \frac{Pr(No) * Pr(x_1/No) * Pr(x_2/No) * Pr(x_3/No)}{\cancel{Pr(x_1) * Pr(x_2) * Pr(x_3)}}$$ Remove.

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

Temperature

	Yes	No	P(E/Yes)	P(E/No)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5

$$Pr(Yes \mid (Sunny, Hot)) = \frac{Pr(Yes) * Pr(Sunny/Yes) * Pr(Hot/Yes)}{\cancel{Pr(Sunny) * Pr(Hot)}}$$

$$= 9/14 * 2/9 * 2/9$$

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

$$\frac{0.031}{0.031 + 0.085}$$

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

$$= 5/14 * 3/5 * 2/5$$

$= 0.085$.

$$Pr(Yes/(Sunny, hot)) = \frac{0.031}{0.031 + 0.085} = 0.27 = 27\%$$

$$Pr(No/(Sunny, hot)) = \frac{0.085}{0.031 + 0.085} = 0.73 = 73\%$$

[Sunny, hot] => No

=> Outlook → Rain

Temprature → Mild

Variants of Naive Baye's

1. Burnoulli Naive Baye's
2. Multinomial Naive Baye's
3. Gaussian Naive Baye's

1. Bernoulli Naive Baye's

Whenever your features are a following Bernoulli's Distribution, the we use Burnoulli Naive Baye's

f1	f2	f3	f4	O/p
Yes	Male	Pass		
No	Female	Fail		

② **Multinomial Naive Bayes** ⇒ I/p ⇒ Text

Dataset ⇒ Sentiment Analysis, Spam ham Classification.

Text ⇒ Numerical Values ⇒ NLP

① BOW
② TFIDF
③ Word2Vec

③ **Gaussian Naive Bayes**

Independent features ⇒ Gaussian Distribution

Age Height Weight O/P

② **K Nearest Neighbour** (KNN)

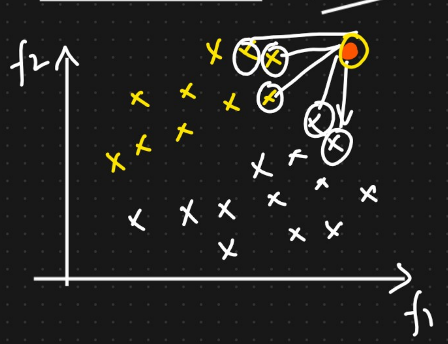
① Classification
② Regression

K=10

K=5 ⇒ Hyperparameter

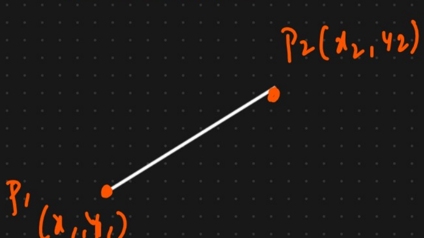
Classification

1 Million Dataset

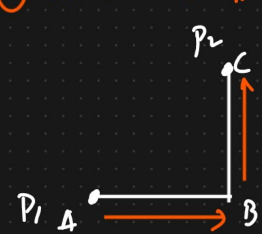


Distance Metrics

① Euclidean Distance

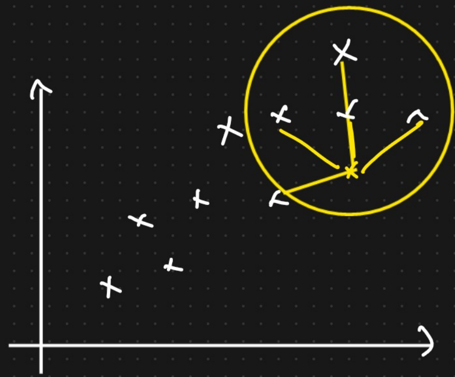


② Manhattan Distance



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

$$\boxed{|x_2-x_1|+|y_2-y_1|}$$



⇒ Continuous value..

Variants of KNN

Two Variants

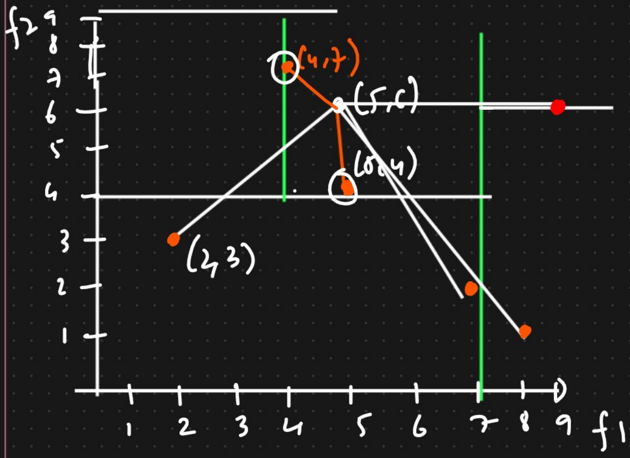
① K Dimension Tree

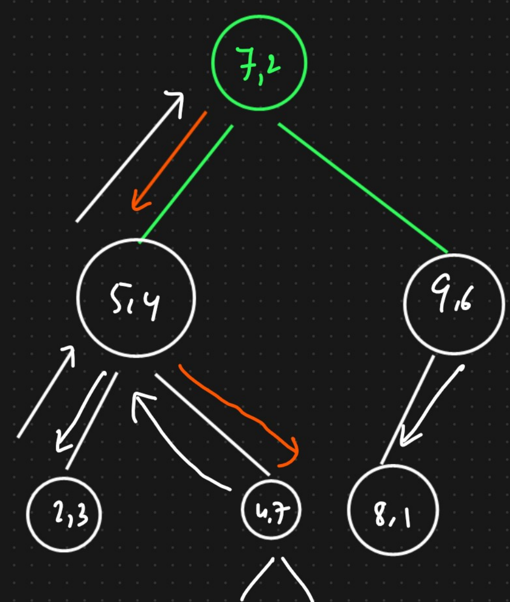
② Ball Tree

K=4

K=5

① K-D Tree





(2,3) (4,7) (5,4) (7,2) (8,1) (9,6)

2, 4, $\boxed{5, 7}$, 8, 9 $\qquad \frac{5+7}{2} = 6$

3, 7, 4, 2, 1, 6

1, 2, 3, 4, 6, 7 $\qquad \frac{3+4}{2} = 3.5$

② Ball Tree