Naive Baye's Algorithm (Classification)

① Probability [Independent And Dependent Events]

② Baye's theorem

③ Naive Baye's Math Intuition.

① Probability

Independent Events

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

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

Dependent Events

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

① $\rightarrow P(O) = 3/5 \rightarrow 1^{st}$ Event

② $\rightarrow P(Y) = \frac{2}{4} \rightarrow 2^{nd}$ Event

$\Downarrow$

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

$Pr(O \text{ and } Y) = P(O) * \boxed{P(Y/O)}$ ⇒ Conditional Probability

$= 3/5 * 2/4 = \boxed{\frac{3}{10}}$

$\Downarrow$

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

Baye's Theorem

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

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

$$\boxed{Pr(A/B) = \frac{Pr(A) * Pr(B/A)}{Pr(B)}} \Rightarrow \text{Baye's Theorem}$$

$P(A/B)$ = Probability of Event A given B has Occured

$P(A)$ = Probability of Event A

$P(B)$ = Probability of Event B

$P(B/A)$ = Probability of Event B given A has Occured.

DATASET

x_1	x_2	x_3	O/P ⇒ y (↓ Predict)
–	–	–	Yes
–	–	–	No
–	–	–	Yes
–	–	–	No
–	–	–	Yes

$$Pr(\underbrace{Y}_{A}/\underbrace{(x_1, x_2, x_3)}_{B}) = \frac{Pr(y) * P(x_1, x_2, x_3/y)}{Pr(x_1, x_2, x_3)}$$

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

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

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

DATASET

x_1	x_2	x_3	O/P ⇒ (Pred.)
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	No
—	—	—	Yes

$$\Pr(y = Yes / (x_1, x_2, x_3)) = \frac{\Pr(Yes) * \Pr(x_1/Yes) * P(x_2/Yes) * \Pr(x_3/Yes)}{\cancel{\Pr(x_1) * \Pr(x_2) * \Pr(x_3)}} = 0.60 \Rightarrow 1$$

Remove

$$\Pr(Y = NO / (x_1, x_2, x_3)) = \frac{\Pr(NO) * \Pr(x_1/NO) * \Pr(x_2/NO) * \Pr(x_3/NO)}{\cancel{\Pr(x_1) * \Pr(x_2) * \Pr(3)}} = 0.40$$

Naive Bayes is a family of probabilistic algorithms based on Bayes' Theorem, widely used for classification tasks in machine learning. It's called "naive" because it assumes that the features in a dataset are mutually independent, which is a simplifying assumption that often does not hold true in real-world data. Despite this, Naive Bayes can perform surprisingly well, especially for text classification problems like spam detection and sentiment analysis.

Lets Solve This Problem

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

(Sunny, Hot) = O/P

Yes/No

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

Yes 9 $\quad P(Y) = 9/14$

No 5 $\quad Pr(No) = 5/14$

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

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

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

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

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

$= 0.085$

Finally

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

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

NEW DATA [Sunny, Hot] $\Rightarrow$ $\boxed{73\%}$ $\Rightarrow$ NO $\Rightarrow$ 0

27% $\Rightarrow$ Yes

$\Downarrow$

$\boxed{\text{Person will Not play}}$

Variants of Naive Baye's

1. Bernoulli Naive Bayes
2. Multinomial Naive Bayes
3. Gaussian Naive Bayes

1. Bernoulli Naive Bayes

indep features

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

Bernoulli $\longrightarrow$ 0, 1

$P(\text{Success}) = 1 = p$

$P(\text{Fail}) = 0 = 1 - p$

Dataset

f_1	f_2	f_3	O/P
Yes	Pass	Male	Yes
No	Fail	Female	No
Yes	Pass	Male	Yes
No	Pass	Female	No.
Yes	Pass		

2. Multinomial Naive Bayes $\Rightarrow$ I/P = Text

Dataset: Sentiment Analysis

I/P

Review Message	O/P
The product is really good	Positive
The product is bad	Negative

Numerical Values ⇒ Natural language Processing

① BOW

② Tf-IDF

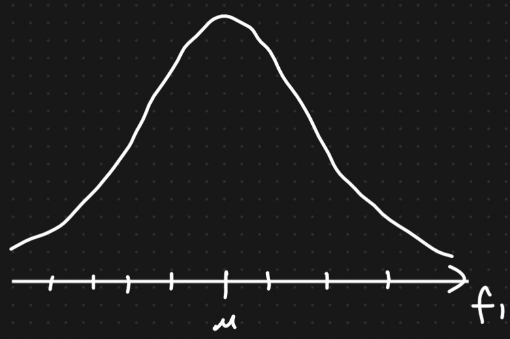
③ word2vec

③ Gaussian Naive Bayes

If the features are following Gaussian Distribution then we use Gaussian Naive Baye's Algorithm to solve Classification problem

IRIS Dataset

[continuous features]



Age	Height	Weight	Yes/No
25	170	78	
28	160	75	
32	170		
34	140		

Types of Naive Bayes Classifiers
Gaussian Naive Bayes:

Assumes that the continuous features follow a normal (Gaussian) distribution.
Suitable for cases where features are continuous and normally distributed.
Multinomial Naive Bayes:

Assumes that features are multinomially distributed.
Commonly used for text classification and document classification problems.
Bernoulli Naive Bayes:

Assumes that features are binary (0 or 1).
Also used in text classification, especially when the presence or absence of a particular feature (word) is more relevant than its frequency.