⇒ Naive Baye's Algorithm:

- Only used for classification

① Probability

② Baye's Theorem

- **Independent events:**

One event don't affect the probability of other

e.g.,

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

$Pr(1) = \frac{1}{6}$

After getting one

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

No effect

- **Dependent events:**

One event affect the probability of other event

e.g.:

We have a jar of marbles have 3 white 2 yellow marbels

- After getting a white marble what is the probability of getting yellow marble.

$P(\text{white}) = \frac{3}{5} \Rightarrow$ **1st event**

$\boxed{P\left(\frac{\text{Yellow}}{\text{white}}\right) = \frac{2}{4}}$ (As one marble removed)

(So outcome/probability of event affected by 1st event)

$$\boxed{P(\text{white and Yellow}) = P(\text{white}) * P\left(\frac{\text{Yellow}}{\text{white}}\right)}$$

$\downarrow$

Conditional Probability

$$P(\text{white and Yellow}) = \frac{3}{5} \times \frac{2}{4}$$

$$\boxed{P(\text{white and Yellow}) = \frac{3}{10}}$$

Now in generalized form

$$\boxed{P(A \text{ and } B) = P(A) \times P\left(\frac{B}{A}\right)}$$

$\downarrow$

Conditional Probability

(2) Baye's Theorem:

- **Derivation:**

We know

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

Expand equation

$P(A) \times P(\frac{B}{A}) = P(B) \times P(\frac{A}{B})$

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

$\Downarrow$

Baye's Theorem

$P(\frac{A}{B}) \Rightarrow$ Probability of A given B has occured

$P(\frac{B}{A}) \Rightarrow$ Probability of event B given A has occured

$P(A) \Rightarrow$ Probability of event A

$P(B) \Rightarrow$ Probability of event B

Now let we have a dataset

$\underbrace{\text{Independent features}}$

x_1	x_2	x_3	y → Dependent feature
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	No

We have to predict y
Means probability of y given x_1, x_2, x_3 so from Baye's Theorem

$$\boxed{P\left(\frac{y}{(x_1, x_2, x_3)}\right) = \frac{P(y) \times P\left(\frac{x_1, x_2, x_3}{y}\right)}{P(x_1, x_2, x_3)}}$$

⇓

Naive Baye's
Machine Learning equation

Now expand it

$$P\left(\frac{y}{x_1, x_2, x_3}\right) = \frac{P(y) \times P\left(\frac{x_1}{y}\right) \times P\left(\frac{x_2}{y}\right) \times P\left(\frac{x_3}{y}\right)}{P(x_1) \times P(x_2) \times P(x_3)}$$

Now according to our problem statement Yes|No

$$P\left(\frac{Yes}{x_1, x_2, x_3}\right) = \frac{P(Yes) \times P\left(\frac{x_1}{Yes}\right) \times P\left(\frac{x_2}{Yes}\right) \times P\left(\frac{x_3}{Yes}\right)}{P(x_1) \times P(x_2) \times P(x_3)}$$

$$P\left(\frac{No}{x_1, x_2, x_3}\right) = \frac{Pr(No) \times P\left(\frac{x_1}{No}\right) \times P\left(\frac{x_2}{No}\right) \times P\left(\frac{x_3}{No}\right)}{P(x_1) \times P(x_2) \times P(x_3)}$$

Now we can see that in both $P\left(\frac{Yes}{x_1, x_2, x_3}\right)$ and $P\left(\frac{No}{x_1, x_2, x_3}\right)$ I am getting same thing constant in remaindar which we can ignore

$$P\left(\frac{Yes}{x_1, x_2, x_3}\right) = P(Yes) \times P\left(\frac{x_1}{Yes}\right) \times P\left(\frac{x_2}{Yes}\right) \times P\left(\frac{x_3}{Yes}\right)$$

$$P\left(\frac{No}{x_1, x_2, x_3}\right) = P(No) \times P\left(\frac{x_1}{No}\right) \times P\left(\frac{x_2}{No}\right) \times P\left(\frac{x_3}{No}\right)$$

So, when we get a new test data we will calculate both Yes and No probabilities using above formula and assign class with most probability.

Let's solve problem

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Now we have to find some things for every relevant feature

Total No. of Yes =15

1. 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	3/5

2. Temprature:

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

[redacted] Now let for practice let only take these two features

Output: Play

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

Now let we get new test data

Outlook	Temprature
Sunny	Hot

So we will write

$$P\left(\frac{Yes}{Sunny, Hot}\right) = P(Yes) \times P\left(\frac{Sunny}{Yes}\right) \times P\left(\frac{Sunny}{Hot}\right)$$

$$P\left(\frac{Yes}{Sunny, Hot}\right) = \frac{\cancel{9}}{14} \times \frac{2}{\cancel{9}} \times \frac{2}{9} = \frac{2}{63}$$

$$\boxed{P\left(\frac{Yes}{Sunny, Hot}\right) = 0.031}$$

$$P\left(\frac{No}{Sunny, Hot}\right) = P(No) \times P\left(\frac{Sunny}{No}\right) \times P\left(\frac{Hot}{No}\right)$$

$$P\left(\frac{No}{Sunny, Hot}\right) = \frac{5}{14} \times \frac{3}{5} \times \frac{2}{5} = \frac{3}{35}$$

$$P\left(\frac{No}{Sunny, Hot}\right) = 0.085$$

Now we will get in %age

$$\boxed{\begin{aligned} P\left(\frac{Yes}{Sunny, Hot}\right) &= \frac{0.031}{(0.031 + 0.085)} = 27\% \text{ or } 0.27 \\ P\left(\frac{No}{Sunny, Hot}\right) &= \frac{0.085}{(0.031 + 0.085)} = 0.73 \text{ or } 73\% \end{aligned}}$$

As Probability of No is greater so our outcome will be

Outlook	Temprature	Play
Sunny	Hot	$\boxed{\text{No}}$

⇒ Variants of Naive Bayes:

① Bernoulli Naive Bayes.

When your features are following Bernoulli distribution (output will be 0 or 1) then we need to use Bernoulli Naive Bayes Algorithm

e.g

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

As all features have two outputs ⇒ **Bernoulli Distribution**

So we will use Bernoulli Naive Bayes

② Multinomial Naive Bayes

If input data is in the form of text then we will use Multinomial Naive Bayes for classification.

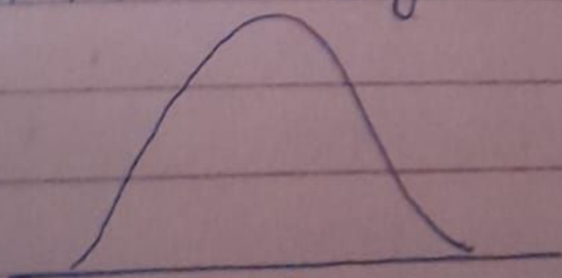
e.g Spam Classifier

Email	Spam / Not Spam
You won million Dollar lottery	Spam
You got job	Not Spam

- We convert these sentences into numerical values (vectors) using Natural Language Processing (NLP)

③ Gaussian
Naive Bayes:

If the features are following gaussian distribution (Bell curved), then we use Gaussian Naive Bayes



e.g IRIS dataset

- Features are continuous e.g Age, Height, weight etc.

Age	Height	Weight	Yes/No
25	170	78	Yes
28	160	75	No
22	150	65	Yes

* In Bernoulli Naive Bayes also we convert features to numerical called **Sparse Matrix**
* So in Natural Language Processing (NLP) we use both Bernoulli Naive Bayes and Multimonial Naive Bayes

f_1	f_2	f_3	Output
Yes	Fail	Male	Yes
No	Pass	Female	No
Yes	Pass	Male	No
No	Fail	Female	Yes

$\Downarrow$

f_1	f_3	f_3	Output
1	0	1	1
0	1	0	0
1	1	1	0
0	0	0	1

Sparse Matrix