Naive Bayes Algorithm: (Binary and Multiclass Classification)

Independent Events:

- Rolling A Dice

Dependent Events: A bag full of different color marbles. If we pick one, then ~~we can~~ and compute probability, then the next probability will be changed for any marble because the number of marble in the box has been changed.

~~Bayes Theorem~~, $P(A \text{ and } B) = Pr(A) * P(B|A)$

Bayes theorem:

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

$$\Rightarrow P(A) * P(B|A) = P(B) * P(A|B)$$

$$\Rightarrow P(A|B) = \frac{P(A) * P(B|A)}{P(B)} \longrightarrow \text{Bayes Theorem}$$

Suppose, for a dataset, Bayes Theorem would be →

x_1	x_2	x_3	O/P (Y)
–	–	–	Yes
–	–	–	No
–	–	–	Yes
–	–	–	Yes
–	–	–	No
–	–	–	No

$$\boxed{P(y|(x_1,x_2,x_3)) = \frac{P(Y) * P((x_1,x_2,x_3)|y)}{P(x_1,x_2,x_3)}}$$

Now, $$P(y|(x_1,x_2,x_3)) = \frac{P(y) * P(x_1,x_2,x_3|y)}{P(x_1,x_2,x_3)}$$

$$= \frac{P(y) * P(x_1|y) * P(x_2|y) * P(x_3|y)}{P(x_1) * P(x_2) * P(x_3)}$$

if, y = yes

$$P(yes|(x_1,x_2,x_3)) = \frac{P(yes) * P(x_1|yes) * P(x_2|yes) * P(x_3|yes)}{P(x_1) * P(x_2) * P(x_3)}$$

$$P(No|(x_1,x_2,x_3)) = \frac{P(No) * P(x_1|No) * P(x_2|No) * P(x_3|No)}{P(x_1) * P(x_2) * P(x_3)}$$

Suppose $P(Yes|(x_1,x_2,x_3)) = 0.60$ and $P(No|(x_1,x_2,x_3)) = 0.40$

for a new data point using the context of (x_1,x_2,x_3) actual points decision would come Yes because Yes has the greater probability.

Let's take an example with the real dataset:

Day	Outlook	Temp	Humid	Wind	Play Tennis
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

E = sunny, Overcast, Rain

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

For simplicity, let's consider only these two are our ~~de~~ independent feature and play tennis is our target feature.

Total Yes = 9 in Play tennis feature $P(Yes) = 9/14$

Total No = 5 " " " " $P(No) = 5/14$

Now,

$$P(Yes \mid (Sunny, Hot)) = \frac{P(Yes) * P(Sunny \mid Yes) * P(hot \mid Yes)}{\cancel{P(Sunny) * P(hot)}}$$ (This won't be needed)

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

$$= 0.031$$

$$P(No \mid (sunny, hot)) = \frac{P(No) * P(sunny \mid No) * P(hot \mid No)}{\cancel{P(Sunny) * P(hot)}}$$ (This won't be needed)

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

$$= 0.085$$

Finally,

$$P(\text{Yes}|(\text{sunny},\text{hot})) = \frac{0.031}{0.031+0.085} = 0.27 = 27\%$$

[We did that to equalize
Means to make percentage
values]

$$P(\text{No}|(\text{sunny},\text{hot})) = \frac{0.085}{0.031+0.085} = 0.73 = 73\%$$

So, for sunny and hot value probability of No is greater

For a new data contains [hot, sunny] probability will be "No".
or "0"

Means → Person will not play

Variants of Naive Bayes:

(i) Bernouli Naive Bayes
(ii) Multinomial Naive Bayes
(iii) Gaussian Naive Bayes

(1) Bernouli Naive Bayes:

Whenever your features are following a Bernouli Distribution, then use Bernouli Naive Bayes.

If your input features and target feature are in Binary kind of form
Look next page

Dataset:

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

⇒

Bernouli → 0,1

f_1	f_2	f_3	O/P
1	1	1	
0	0	0	
1	1	1	
0	1	0	
1	1	1	

} Sparse Matrix

In this type of dataset, we will use Bernouli Naive Bayes. Multinomial Naive Bayes can also be used here. (Because of being a sparse matrix)

Multinomial Naive Bayes:

When our input data is in $\boxed{\text{text format}}$, we have to use this techniques

Example, Dataset:

Review Message	O/P
Product is good	Positive
Product is bad	Negative
Product is missing	Negative

We will convert these text based input feature into numerical values by using NLP Techniques –
1. BOW
2. TF-IDF
3. Word2vec

And for prediction we will use Multinomial Naive Bayes.

③ Gaussian Naive Bayes:

If the features are following Gaussian (Normal) Distribution, we will use Gaussian ~~Dis~~ Naive Bayes.

In Gaussian Distribution feature are $\boxed{\text{continuous.}}$

We have ~~of~~ to think the maximum number of features are following which distribution and act according to that.