Date.....

Working of Naive Bayse Algo

NAIVE Bayse

CSK [Chennai Super Kings]

Toss	Venue	Outlook	Result
won	Mumbai	Overcast	Won
lost	Chennai	Sunny	Won
won	Kolkata	Sunny	Won
won	Chennai	Sunny	Won
Lost	Mumbai	Sunny	Lost
won	Chennai	Overcast	Lost
won	Kolkata	Overcast	Lost
won	Mumbai	Sunny	Won

CSK_Data

To predict ⇒ Whether or the Input data given, CSK will Won or Lost the Match.

Now :- A B

$P(W|Won \cap Mumbai \cap Sunny) = ?$

OR

A B

$P(L|Won \cap Mumbai \cap Sunny) = ?$

Input ⇒ {won, Mumbai, Sunny}

Predict ⇒ Won or Lost..

As per the Bayse theorem...

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

Accordingly :-

$$P(W|Won \cap Mumbai \cap Sunny) = \frac{P(Won \cap Mumbai \cap Sunny | Won) * P(W)}{P(Won \cap Mumbai \cap Sunny)} \Rightarrow (1)$$

Let's say the value will be ⇒ 0.90

$$P(L|Won \cap Mumbai \cap Sunny) = \frac{P(Won \cap Mumbai \cap Sunny | Loose) * P(Loose)}{P(Won \cap Mumbai \cap Sunny)} \Rightarrow (2)$$

Let's say the value will be ⇒ 0.10

Ans Probability which is greater is predicted and tells us the answer by the Naive Bayse Algorithm...

From (1) and (2) We can Observe that the Denominator is same in both the cases... So, we can remove the denominator..

Taking the Numerators of (1) and (2)

Date.....

$P(\text{Won} \cap \text{Mumbai} \cap \text{Sunny} \mid \text{Won}) * P(\text{Won}) \rightarrow 5/8$

$P(\text{Won} \cap \text{Mumbai} \cap \text{Sunny} \mid L) * P(\text{Loose}) \rightarrow 3/8$

This is for ① Numerator

To Calculate the $P(\text{Won} \cap \text{Mumbai} \cap \text{Sunny} \mid \text{Won})$ = It is too tough to go too Conditional and Specific. Then

Here Naive Bayes put a Assumption as :-

$= P(\text{Won} \mid W) * P(\text{Mumbai} \mid W) * P(\text{Sunny} \mid W)$

It means that the P(Won the toss given the match Won).

It means P(Venue is Mumbai given Match Won).

It means P(OutLook is Sunny given Match Won)

$\Rightarrow \frac{4}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{5}{8} \Rightarrow \frac{4}{25} \Rightarrow \boxed{0.160}$ - ①

For ② Numerator also, same process :-

To Calculate the $P(\text{Won} \cap \text{Mumbai} \cap \text{Sunny} \mid \text{Loost}) * P(\text{Loose})$ done $\Rightarrow 3/8$

It it too tough because it is too much specific and conditional. To solve this, Naive Bayes puts the Assumption and the equation will be changed as :-

$$\boxed{P(\text{Won} \mid \text{Loost}) * P(\text{Mumbai} \mid \text{Loost}) * P(\text{Sunny} \mid \text{Loost})}$$

$\Rightarrow \frac{2}{3} \times \frac{1}{3} \times \frac{1}{2} \times \frac{3}{4} \Rightarrow \frac{1}{36} \Rightarrow \boxed{0.028}$ - ②

From ① and ② values.., Which is greater Value ①

Ans The Predicted Answer would be Won the Match.. this Analysis of which is greater is **Maximum a posteriori Rule** or MAP Rule...

Spiral

Date.....

Note:- The Assumption of Naive Bayes is termed as "Conditional Independence". It means the Input Data variables all are Independent with each other and only related with the Output Variable [Y]. i.e; "Result" as per this example...

Handling Numerical DATA

Q Till Now, we have looked for the Categorical Data. But, How Naive Bayes handles Numerical Data we see in this example...

Height	Weight	Gender
172	150	M
180	170	M
165	140	M
190	200	M
139	100	F
145	120	F
160	140	F
172	150	F

Input Data $\Rightarrow \{H=185, W=170, G=?\}$

Predict using Naive Bayes...

Sol: To Calculate $\Rightarrow P(M|H=185 \cap W=170)=?$

OR

$P(F|H=185 \cap W=170)=?$

$P(M|H=185 \cap W=170) = P(H=185|M) * P(W=170|M) * P(M) \Rightarrow 1/2$

To Calculate the $P(H=185|M) * P(W=170|M) * \frac{1}{2}$

This is Calculated as per the data.

Here as per the data there are 3 types of Distribution of Naive Bayes which we can apply:-

1. Bernoulli Distribution...
2. Multinomial Distribution...
3. Gaussian Distribution...

Spiral