

In Naive Bayes, Bayes comes from Bayes Theorem from Probability.

Bayes Theorem states that,

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

where,

$P(A|B)$ – the probability of event A occurring, given event B has occurred

$P(B|A)$ – the probability of event B occurring, given event A has occurred

$P(A)$ – the probability of event A

$P(B)$ – the probability of event B

So what is the problem that is generally given?

In a classification problem, we are generally given an x, lets say one test data, we are supposed to predict the value of y.

Lets say the value of Y can be a or b. For Naive Bayes we can do as many classes as possible. Its not really a Binary classification. It can handle as many classes as we have.

So we have to find out whether the class of this x is a or b.

We can never be sure what class is going to be. Its basically based on the training data whether we believe the class is goign to be a or b.

Lets say the x is the input we recieved.

And we want to oredict y values out of these classes.

In another way we can say we want to find out what is the probability that y is a given that input is x which can be written as $P(y = a/x = x)$.

Along with these, we can find out these values $P(y = b/X = x)$, along with lets say another class $P(y = c/X = x)$, what these values mean is given this input is already decided, what is the probability that the output class is the y = a, y=b or y = c.

And sum of these three should be 1. Coz these three are the only possible outputs.

Lets say if we find the probablities of these three terms, we can find the out the predictions that should be the maximum of these three.

So in other ways lets say the classes are $Y = a_1, a_2, a_3. a_k$ say we have k possible classes.

What we really waant to find is $P(y = a_i/X = x)$ and then maximize this with $i = 1 - to - n$. So the answer will be the maximum a_i means maximum a will be the answer.

SO lets just focus on one particular value.

Lets say we want to find out what is the probability $P(y = a_i/X = x)$ seems like this might be really hard to find out given the training data as well.

What we are going to try is we will try the Bayes theorem and look at the other side of it.

Applying Bayes theorem our equation looks like

$$P(y = a_i/X = x) = \frac{P(X = x/y = a_i) * P(y = a_i)}{P(X = x)}$$

If we put this in $P(y = a_i/X = x)$, and we try to maximize over i , we will realize the denominator is same for all terms. So,

$$P(y = a_i/X = x) = \frac{P(X = x/y = a_i) * P(y = a_i)}{P(X = x)}$$

Lets say for a_1 ,

$$P(y = a_1/X = x) = \frac{P(X = x/y = a_1) * P(y = a_1)}{P(X = x)}$$

.

Now if we do it for class a_2 , the numerator will be different because they depend upon class, but the denominator is going to be exactly the same for any class.

So what we are going to do is we are going to get rid of denominator.

As soon as we get rid of denominator, THE SUM IS NOT GOING TO BE 1, ITS NOT GOING TO BE PROBABLITY ANYMORE.

But the point is we want to maximize the $P(y = a_i/X = x)$, we want to find that particular i which is max, removing the denominator in $\frac{P(X=x/y=a_1)*P(y=a_1)}{P(X=x)}$ will not change the maximum it will not change the i for which this

term $P(X = x/y = a_1) * P(y = a_1)$ going to be maximized.It is going to be the exactly the same i .

So we can very easily get rid of the denominator, the problem that we are now going to solve is we want ot maximize ($P(y = a_i/X = x)$). This is basically same as *Find – i – such – that*

$$max_{i=1}^k (P(y = a_i/X = x))$$

This is basically same as

$$max_{i=1}^k P(X = x/y = a_i) * P(y = a_i)$$

because the denominator is exactly same in all the terms.

So we are going to maximize $P(X = x/y = a_i) * P(y = a_i)$.

It is not too hard to find these values, lets say we have 3 classes, a1, a2, a3. and a1 = 40, a2 = 30, a3 = 30.

So $P(y = a_i)$ of $a_1 = 0.4$, $a_2 = 0.3$ and $a_3 = 0.3$,

Finding $P(X = x/y = a_i)$ will be interesting.