

Before Implementing, we have to look at one more thing which might cause us a lot of problem.

Lets say the data that we have the same table.

salary	Loan_approved
High	Yes
High	Yes
Mid	Yes
Low	Yes
Mid	Yes
High	Yes
Low	No
Low	No
Low	No
Mid	No
Low	No

So what is the probability that $P(\text{salary} = \text{Low} / \text{approved} = \text{Yes})$,

So we look at only those which have approved yes.

Then within that, how many have salary as Low?, so 1/6. Seems like we are getting the probability.

Now how about $P(\text{salary} = \text{High} / \text{approved} = \text{No})$

Seems like none of the No entry have the salary High.

So we will end up writing the probability as 0/5.

This means that if the approval is No, the salary cannot be high.

So if the example comes up, which has high salary, there is no way you can predict that it can get approved.

That sounds like a Bad classification.

So it is possible that a high salary person has done fraud that's why he's not getting approval.

In a nutshell, we do not want our probabilities to be zero.

What we generally do is we correct the values. What we are gonna do is salary has three different values.

So we will add a little data to each of these. That means we will look at main formula

$$P(X^j = x^j / y = a_i) = \frac{\text{countof}(y = a_i \text{ and } X^j = x^j)}{\text{countof}(y = a_i)}$$

Now we are saying that the numerator count can be zero.

What we want is we want Probability never to be zero.

So let's add +1 in numerator.

$$P(X^j = x^j / y = a_i) = \frac{\text{countof}(y = a_i \text{ and } X^j = x^j) + 1}{\text{countof}(y = a_i)}$$

Adding a +1 creates a problem that it does not represent probability anymore.

The sum of all such problem will not be 1 anymore.

Lets say for $P(X^j = x^j / y = a_i)$, X^j has two possible values let's say h and l, then the probability equation will be like,

$P(X^j = h / y = a_i) + P(X^j = l / y = a_i) = 1$, because X^j can either take the value h or l.

So what we will do is we should add to denominator as well.

We should add possible no of values X^j can take.

$$P(X^j = x^j / y = a_i) = \frac{\text{countof}(y = a_i \text{ and } X^j = x^j) + 1}{\text{countof}(y = a_i) + \text{Possible values of } |X^j|}$$

So for eg, let's take the above table, values in X^j is 3, high, mid and low, so $|X^j|$ will be 3.

This correction is called *Laplace - Correction*

Laplace - Correction says that to avoid zero probability, whatever probability you are calculating, just add +1 in numerator, just assume the hypothetical training datapoint 1 we are adding. Now as we add 1, and there are let's say 5 different possible values of X^j , and we are adding 1 on each of them, we need to add 5 to the denominator as well.