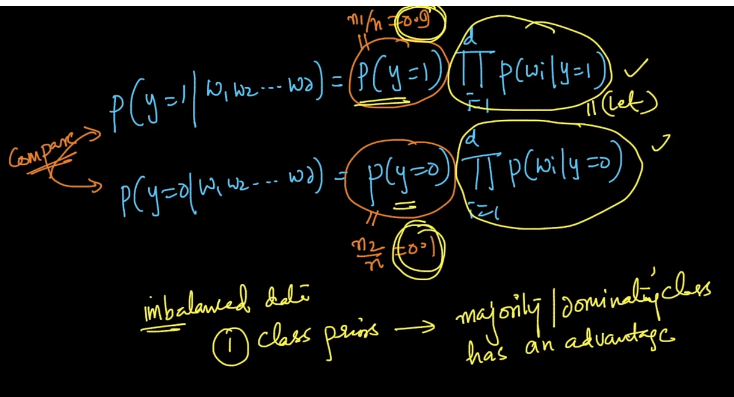
So as we have much more number of points of class than other class i.e., n1>>>n2 in imbalanced data.

It will affect a lot to Naïve Bayes formula as we are using P(Y=1) and P(Y=0) in formula of Naïve Bayes.

So suppose 90% of points are +ive i.e. n1 = 90% and n2 = 10%

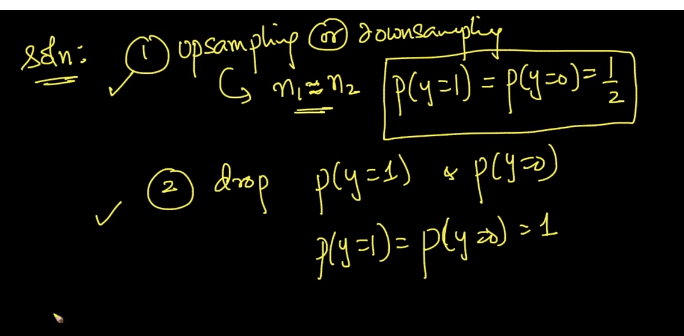
Than P(Y=1) = 0.9 and P(Y=0) = 0.1 and this is fairly unfair.



So how can we overcome this problem?

We can use methods of upsampling and down sampling what we learnt in KNN.

And another way is to drop P(Y=1) and P(Y=0)



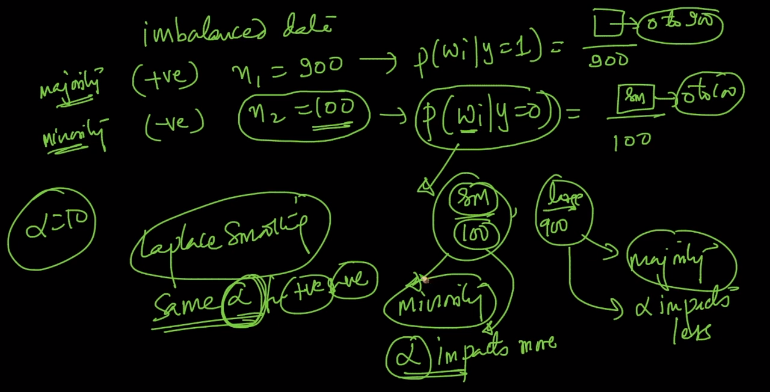
Both the methods are fine as it will cancel out each other.

There is also a third technique for which some scientist have modified NB formula itself but they are just some research papers and are seldom used.

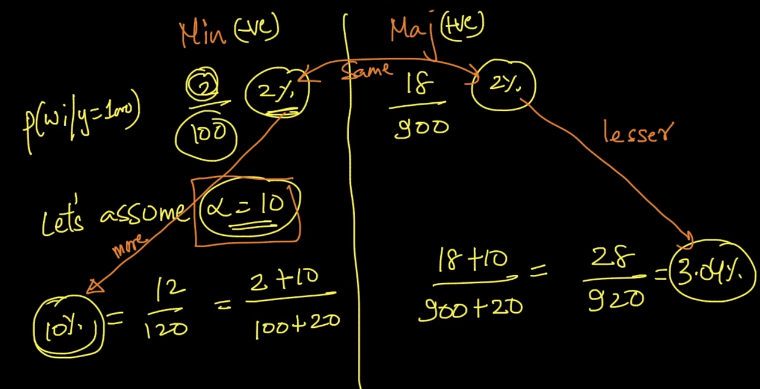
Bur there is another problem with imbalanced data.

Suppose we apply Laplace Smoothing to data and since we apply same value “Alpha” to both majority and minority class.

It will affect less to Majority and will impact more to Minority class.



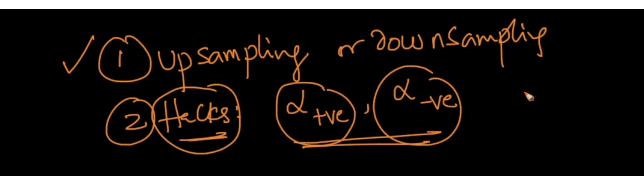
Let’s see this with an example.



So initially both the minority and majority class were giving probability of 2% but after applying Laplace smoothing the minority class become 10% and majority become 3.04%.

The change in probability % is very much for minority class and for majority class it is very less.

So to overcome this we can do up sampling and down sampling or we can use a hack that we use different “Alpha” for both positive and negative class.

.