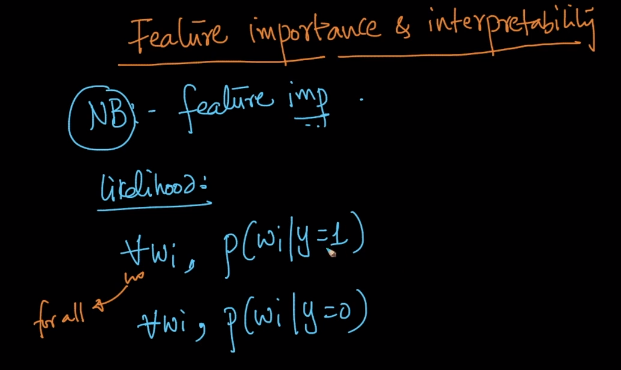
**Feature importance in Naïve Bayes:**

Just like we saw in KNN we can check feature importance using forward selection and other techs.

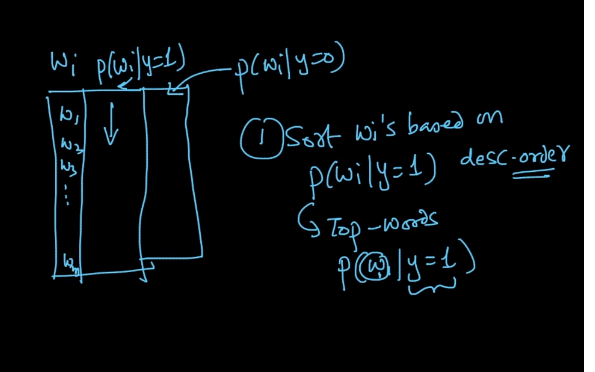
How can we check Feature importance in NB.

SO in NB we have likelihood probabilities for every W in corpus.

And we have it like for word W we have P(W|Y=1) and P(W|Y = 0)

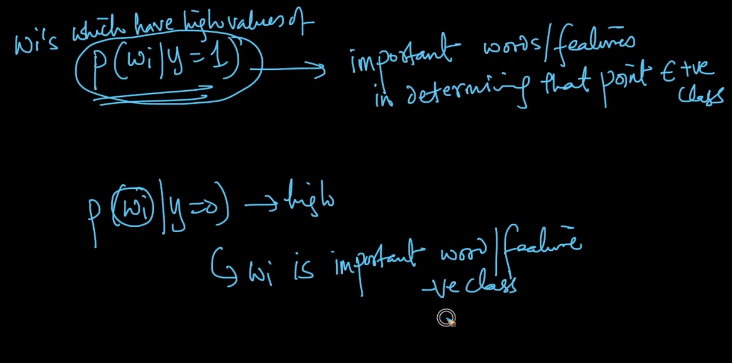


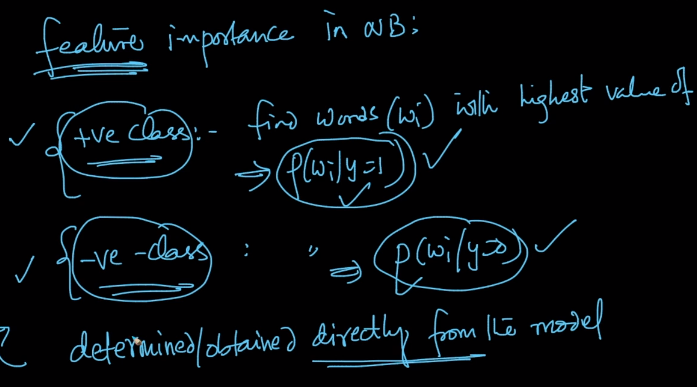
Now since we have this for every word in table ,we will sort it on P(W|Y=1)



Which means our top words after sorting occurs very frequently in that class.

So on the same basis we can say that the word or the feature is how much important.





Since we are computing all the likelihoods in Naïve Bayes in training phase itself so feature can be directly determined by model.

Where as in KNN it was not possible.

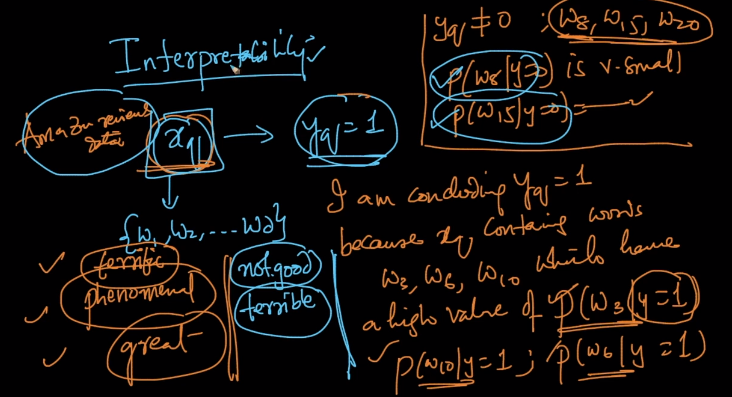
NOW let’s look at another term called **Interpretability**.

As the word suggest how can we interpret if the text or sentence belong to positive class or negative class.

So this is also done by looking at the values of likelihood.

We take some words from text or reviews and say that P(W|Y=1 ) and P(W1|Y=1 ) are very high and hence we are declaring it as positive class

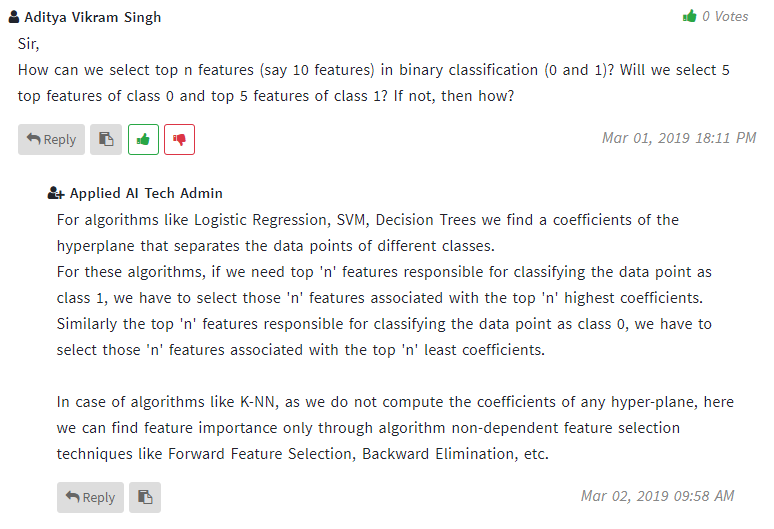
And simultaneously we can also say P(W3|Y=0) and P(W5|Y=0) are very low and hence it cannot be negative class

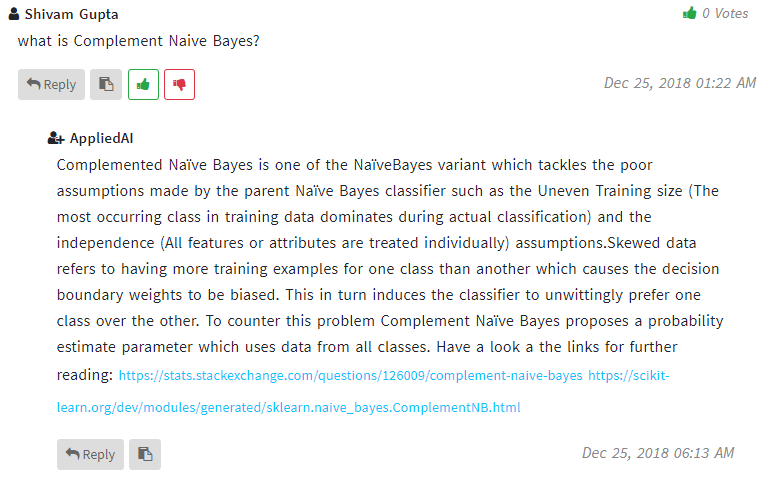


When we go to our Amazon food reviews dataset we get words like “terrific” , “Phenomenal” or “Great” and so the value of likelihoods for this words in positive class

Will be very high and in negative class it will be very low and hence it will always got to positive class and not negative class.

Comments:





**if you need total top n features you can n/2 from each class or you can combine all and you can take top n features**.

