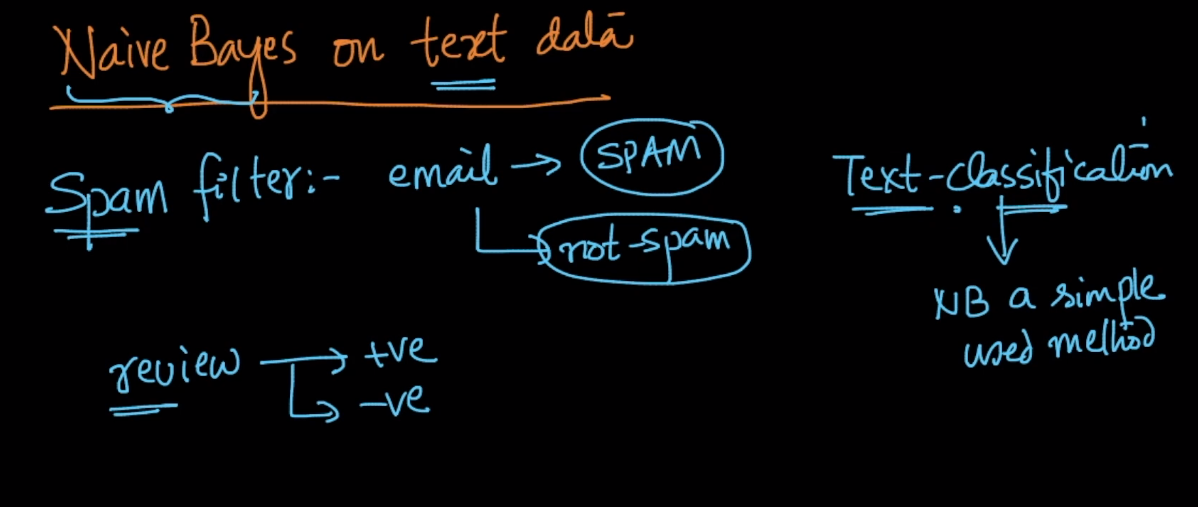
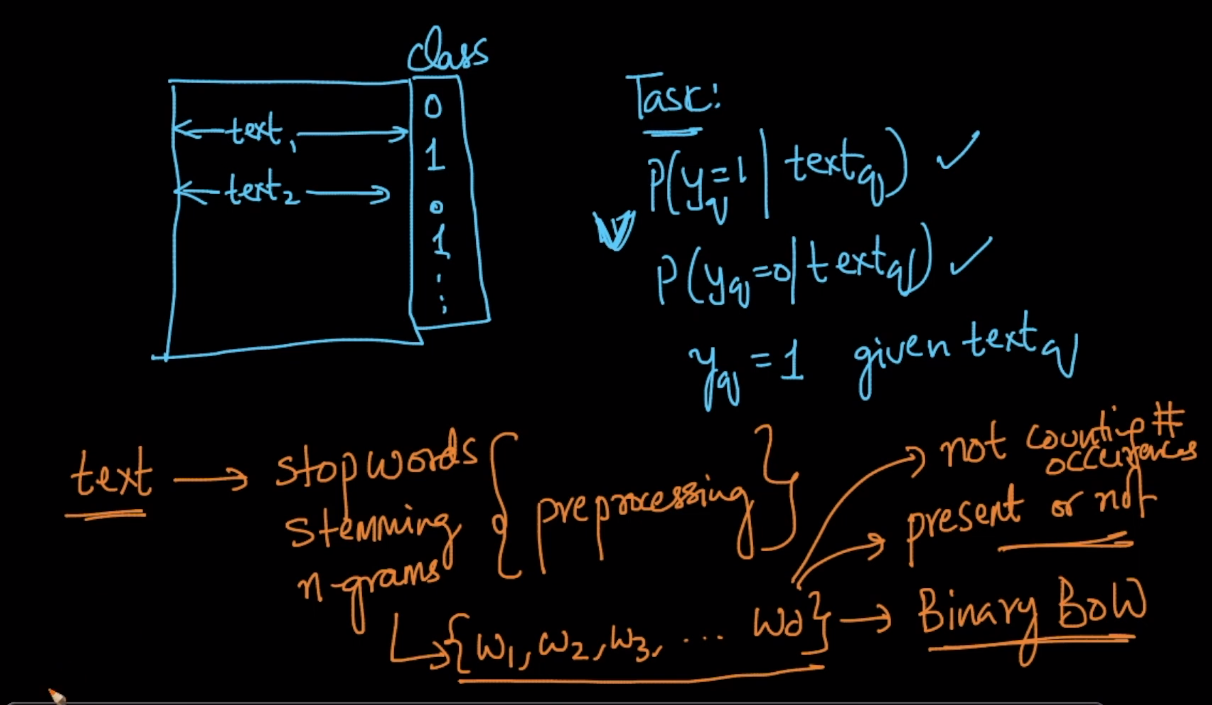
**Naïve Bayes in text classification:**

It is really very popular algorithm in text classification and may be first algorithm used to classify spam emails from non – spam emails

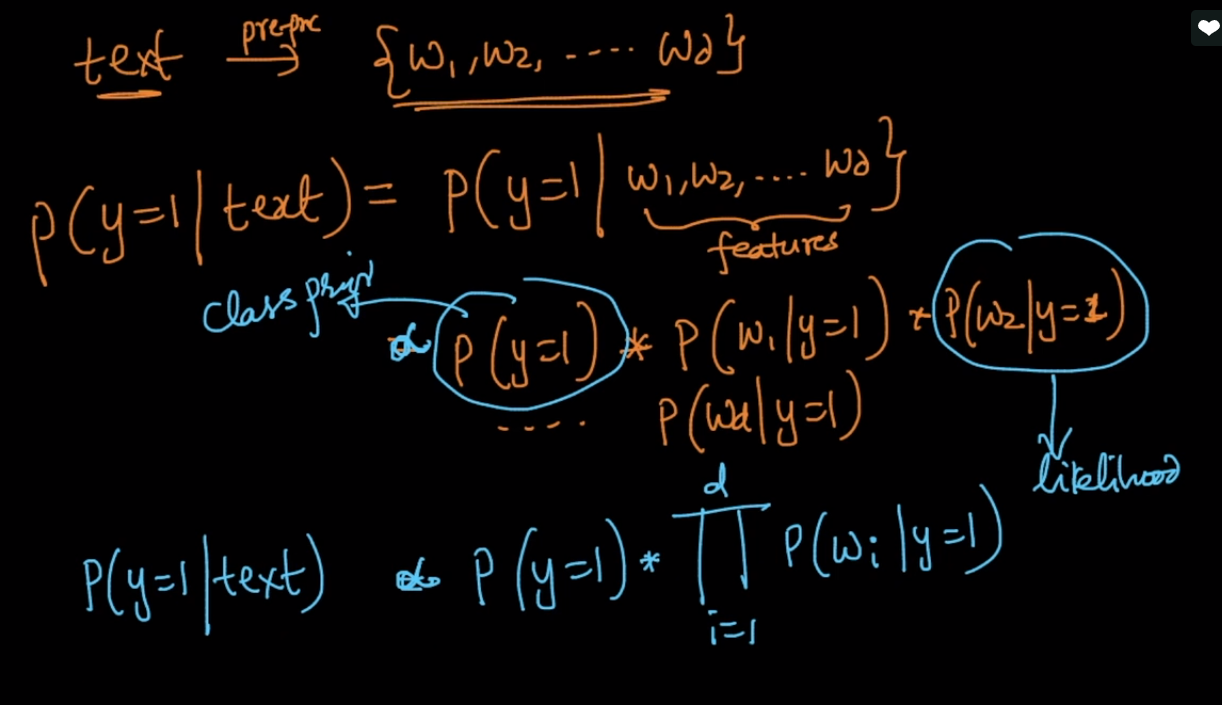
\

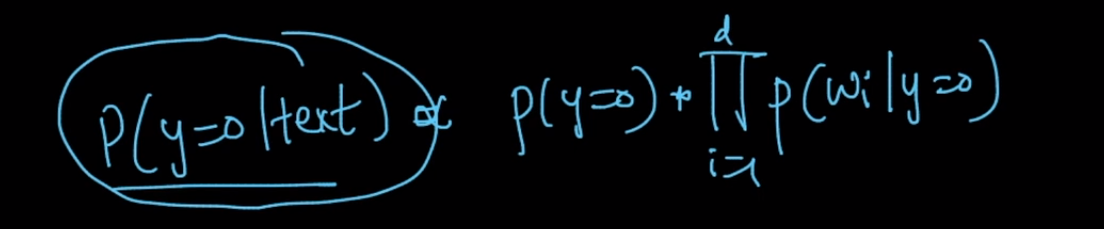
Taking an example of binary classification lets see how NB performs in text classification



Just like in any text classification we do text pre-processing depending upon requirement and so lets say we use Binary BOW where we will get d words in the end and instead of counting occurrence of word we just check if word is present or not.

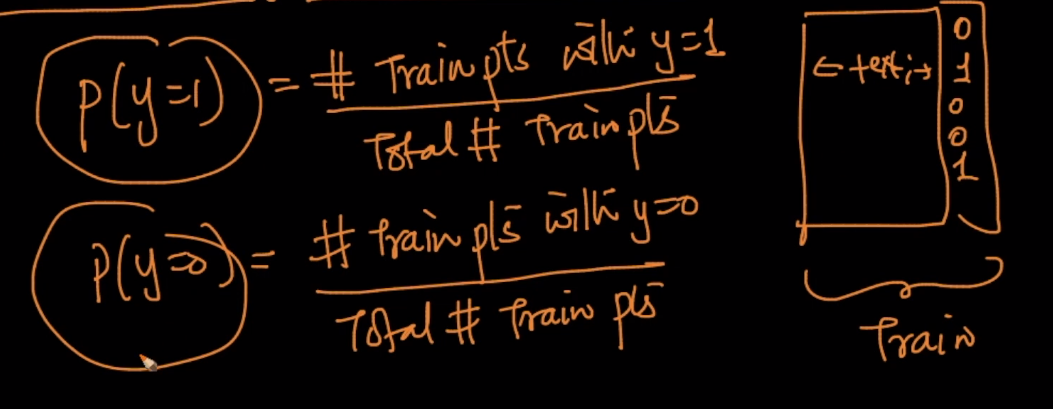
Now after pre-processing what we have got is



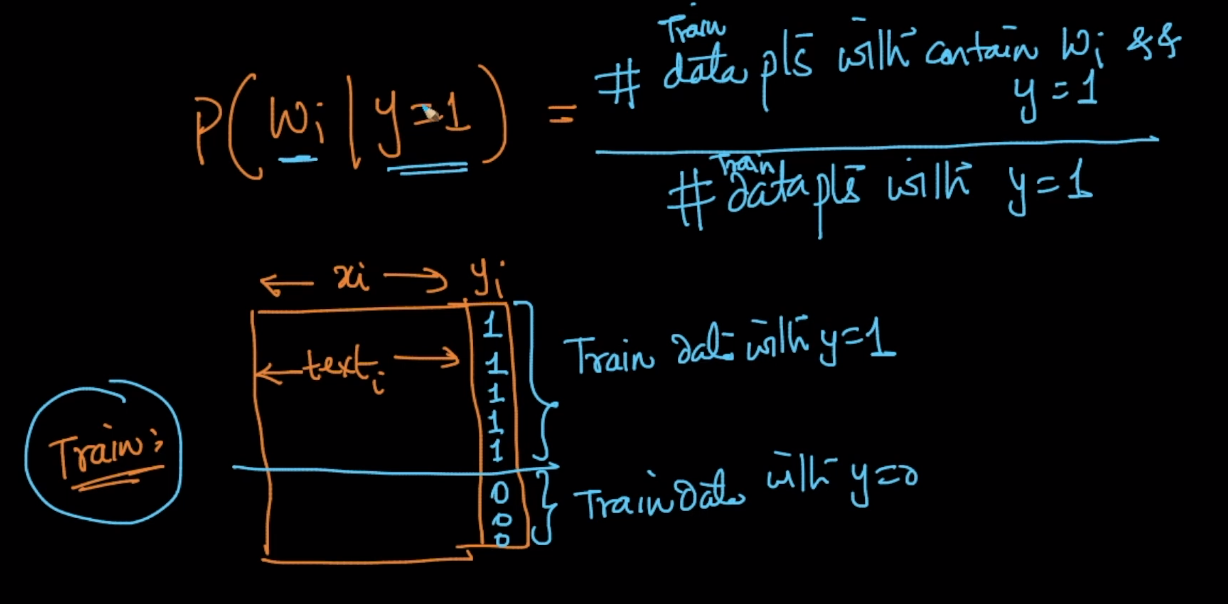


After we have got this values

We will calculate P(y = 0) and P( y = 1)



Now we will compute likelihoods:



And for y = 0 we can same thing by replacing y = 0 in place y = 1. And remember this is all done on just training data.

This is way it do Text classification.

Naïve Bayes is often times used as benchmark which means taking its performance measure as benchmark we compare it with other algorithms and check how much improvement or depreciation in performance they are giving.

