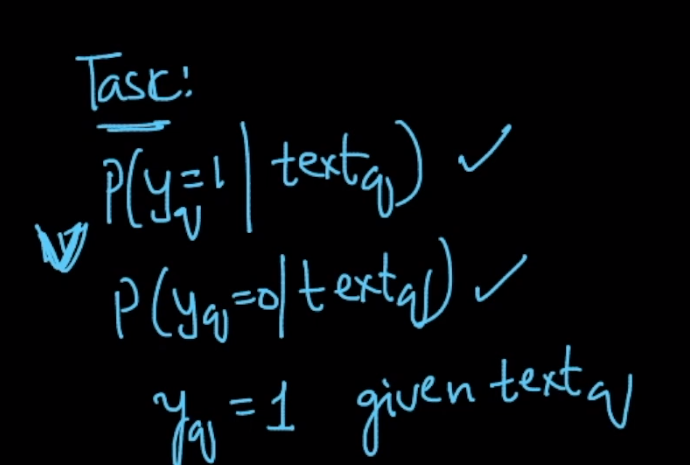
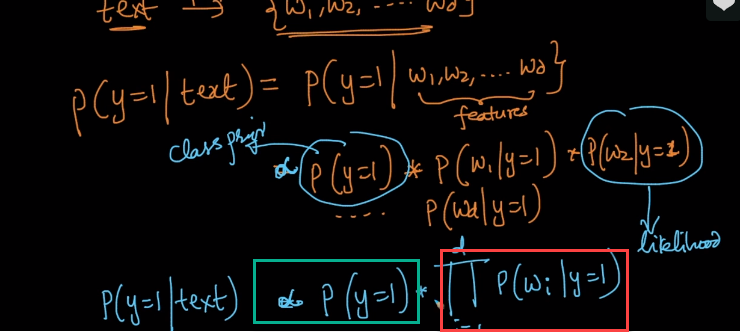
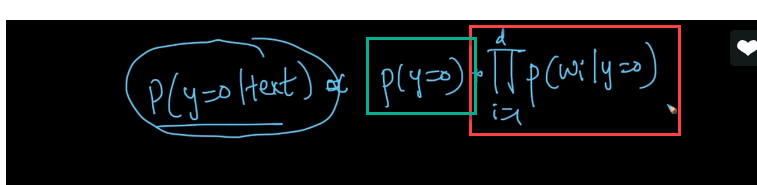
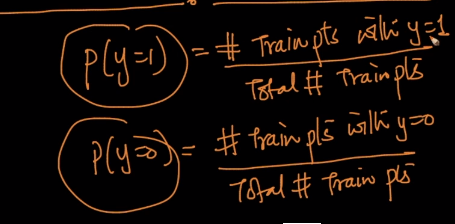
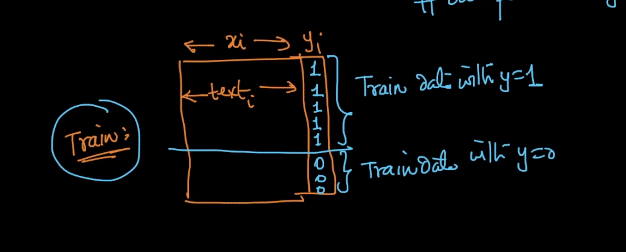
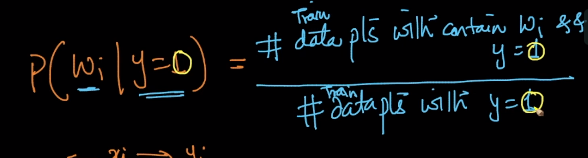
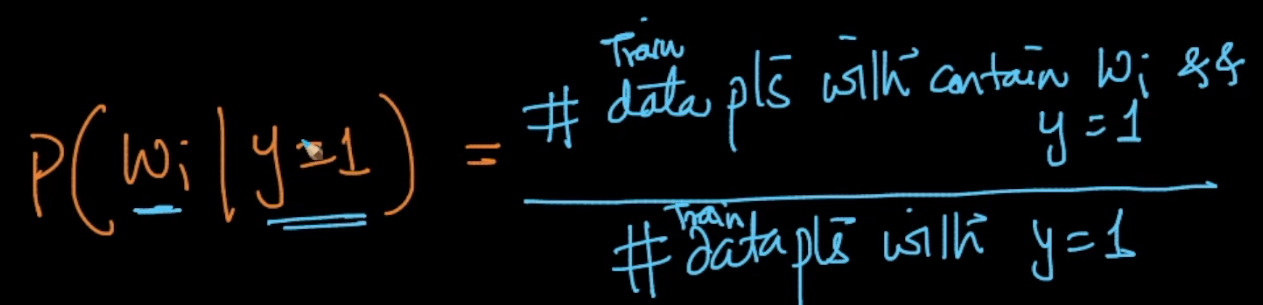
**Naive Bayes on text data:**  
- For text classification , NB can be a simple method to use.   
 For example - To know whether review is +ve or -ve  
 Spam filter: To identify given email is spam or not  
- For a two-class classification problem (0 or 1), for a given task we will calculate probability of Yq = 1 for given text q and probability of Yq = 0 for given text q. we will decide the class of the given text based on the value. High value will be the winner.  
- NB is considered as base line(benchmark) model for text classifications.  
**Task:**



- Just like other techniques, here we will pre-process data like remove stop words, stemming and sometimes taking n-grams of a corpus = {w1,w2,w3,..wd}. d is the dimensionality of our corpus   
- Once we have it, for a given text, we will calculate the binary BOW(vector representation, which has 0- word not present in corpus or 1 – word present in corpus)  
- Once we have Binary BOW of given text, we need to calculate P(Y=1/text) meaning probability of Y=1 of a given text.   
  
- similar we can calculate P(Y=0/text) by using above steps, the final formulae will be  
  
Now the question, we have two parts(Green – Part#1(Also called them as **Class** **Prior**), Red – Part#2(Also called them as **likelihood**)) in above, how to compute it ?  
  
**Calculating Part#1**   
  
By above formula, we can calculate the **Class Priors**

**Calculating Part#2:**  
Here, we will split Train data set in to two parts.   


**Formulae:** 1st image with all values Y=1 and 2nd image with all values Y = 0 of data set **By Above, we calculate Part#2,**   
Total number of data points which contain word Wi and with class label Y =1 divided by Total data points with Y = 1.  
Total number of data points which contain word Wi and with class label Y =0 divided by Total data points with Y = 0.

Once we have the prior class and likelihood, we can calculate the our Task