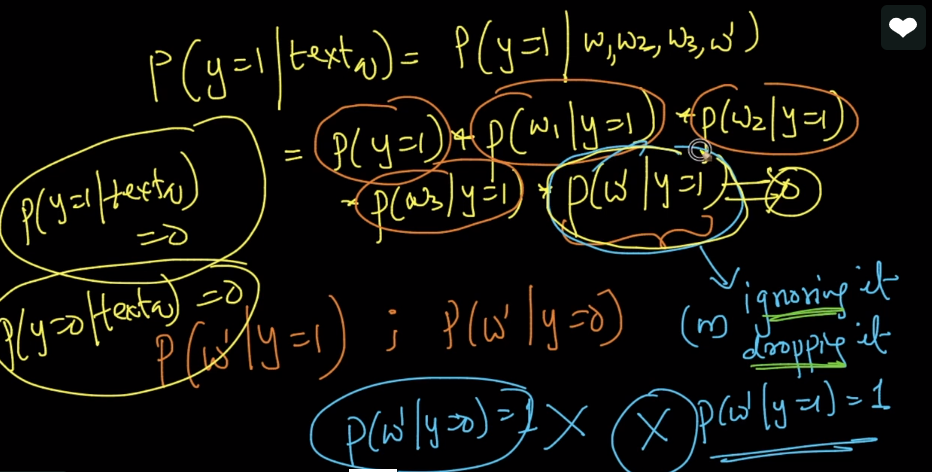
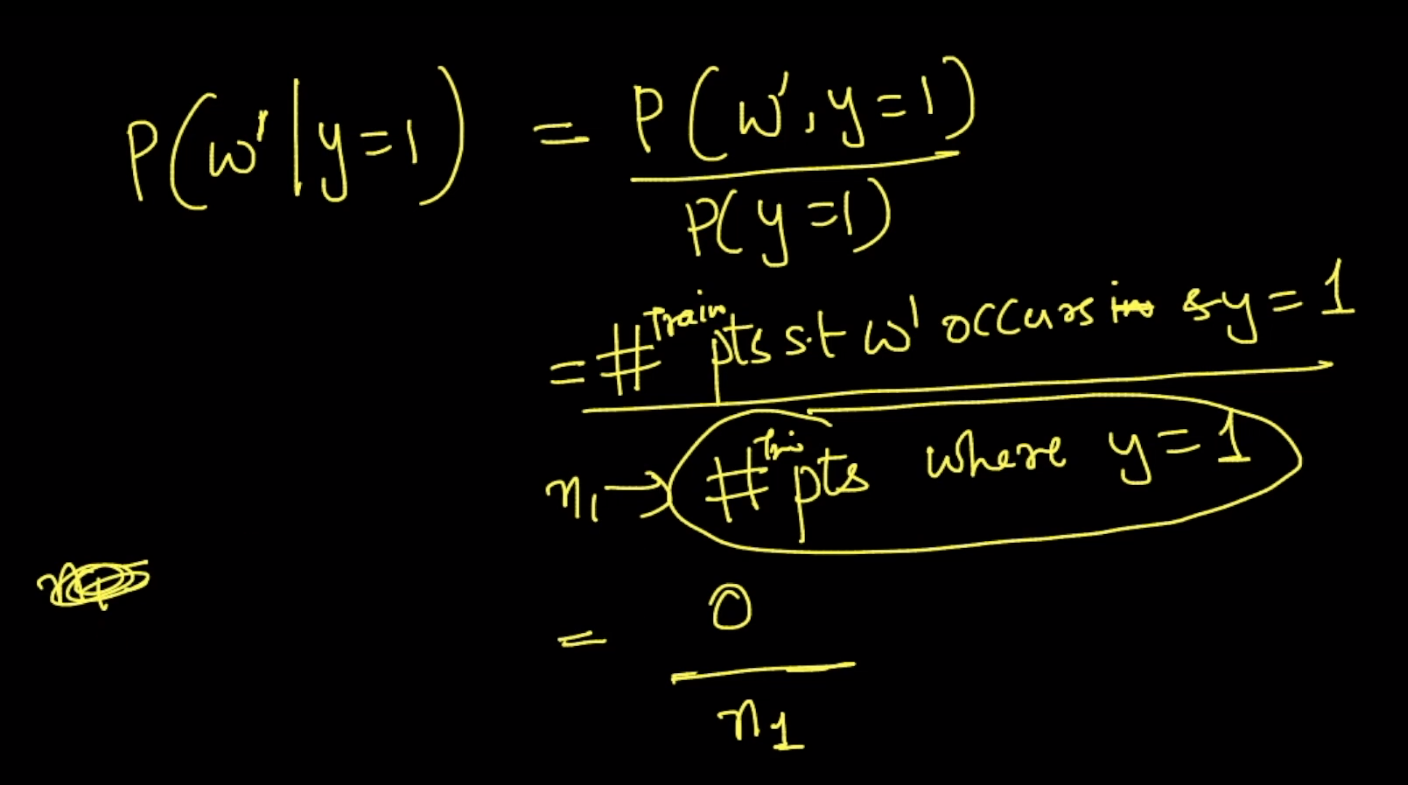
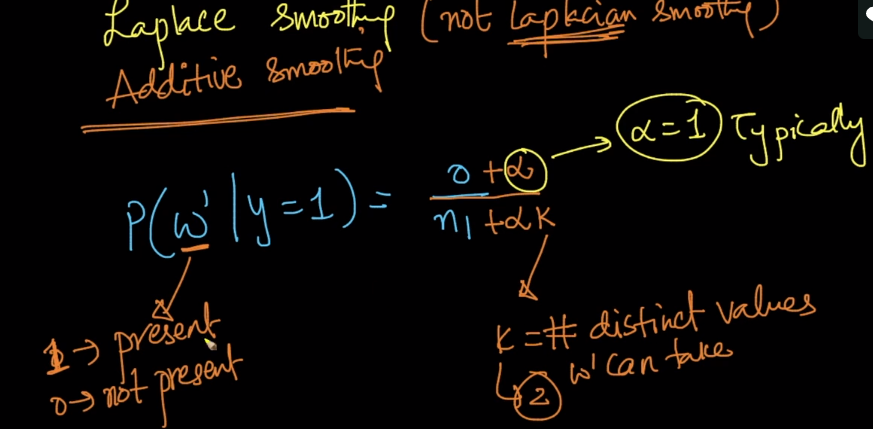
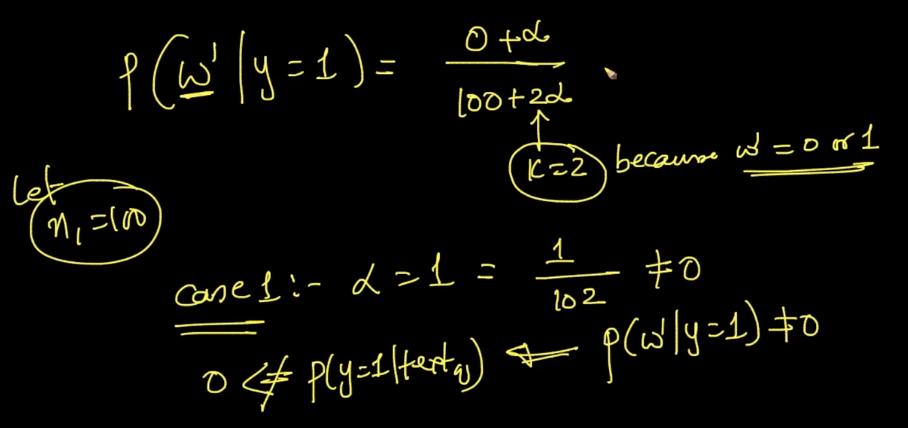
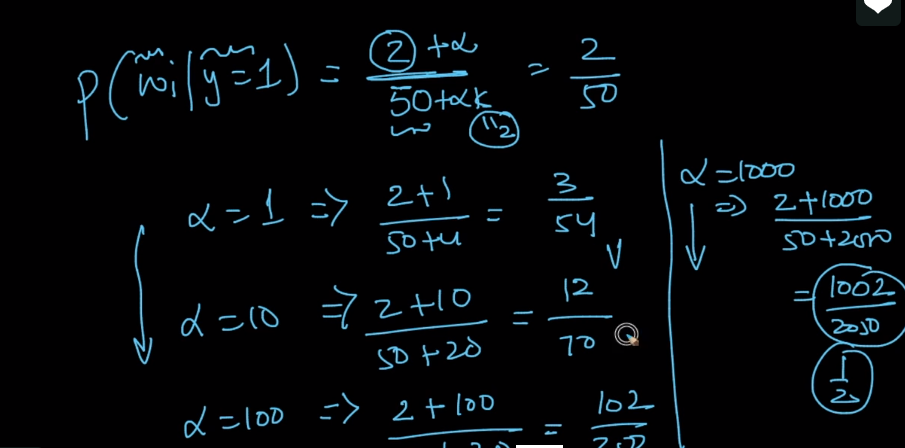
**Laplace Smoothing Or additive smoothing :**

Let say, training data we pre-processed it and prior class and likelihood calculated.   
Now let’s say we have a test text and we have new word in the test text.  
we cannot put it as “0” while calculating the prior and likelihood on test data because, enter output will become zero or dropping which is not correct. So, we use Laplace smoothing technique to avoid this issue.   
Example: If w’ is not in train data set and we compute our probability for that,   
  
  
To avoid this issue, we have method called Laplace smoothing. Here, we have formulae

Here K is number of w’ values we can take. Here, the k value = 2 because the possible values of w’ is 2 i.e., 1 - present,0-not present and we are looking on binary BOW and not count vectorized BOW. And alpha value is typically 1 and n1 is number of words.  
Another example below:  
  
Here n1 =100,k=2 as we using binary BOW  
The final out is not 0 here and the final output mentioned in image 1 is also not 0  
Now let’s take what happens when we increase the alpha value  
  
As we increase the alpha value, we’re moving towards uniform distribution i.e, ½   
Meaning half probability distributed equally when y=1 and y = 0. When we do it? Explained below  
When num and denominator are small, then we will have less confidence. In that case, we going to increase the alpha value and moving it to uniform distribution  
   
So in general, mostly we will add Laplace smoothing even if we didn’t get the output = 0   
Typically, alpha will be 1 in all functions and we will increase if we have less confidence ratio