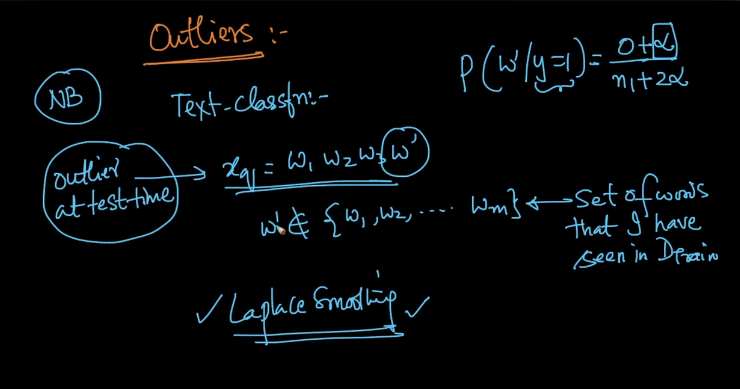
**Outliers:**

What are outliers in NB.

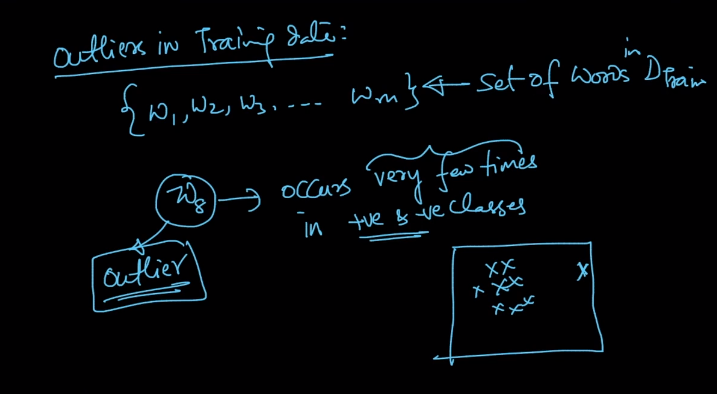
The point wx we got in our test data which was not present in training data, it can be handled by Laplace Smoothing but since we have not seen the point or word in training data

It is said to be an outlier.



Than what are outliers in training data?

So the answer is that the word which occur very less frequently in both +ive and –ive class are called outliers in training data.

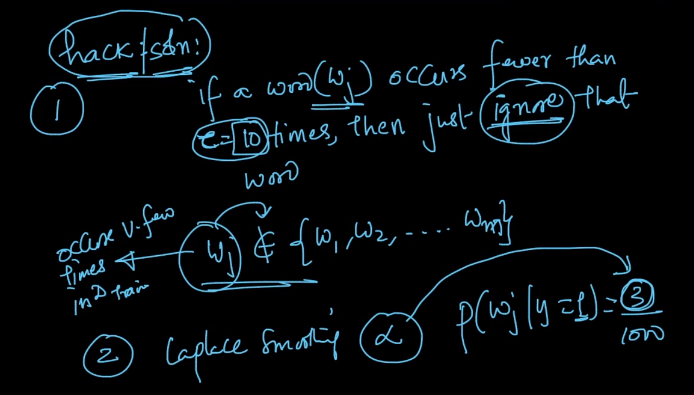


And to handle outliers in training data we have some methods.

The first way is to set a threshold for minimum occurrence of the word and below that threshold don’t consider the point.

This is very simple hack or solution.

Another way is to use Laplace Smoothing and since the numerator while calculating probability for outlier will be very low so the value of Alpha will impact more to it.



Comments:

