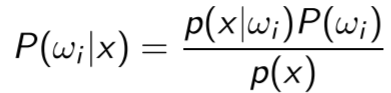
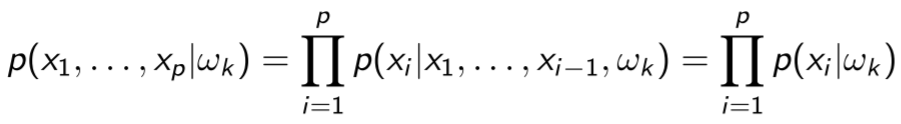
1.

The Bayes classifier is a solution to a classification problem where you have x features and want to know which class they belong. It is based on the calculation of the a posteriori distribution of the classes by using the Bayes law. From a train dataset we can obtain and and By applying Bayes law we obtain:

By calculating this probability, we can predict what class the element belongs to based on its features. If we associate a loss function ) to penalize a difference between the target (y) and predicted class ) with the probability ) we obtain the cost function:

This is an optimal classifier in the sense that it minimizes the risk of a general loss function. This is very easy to implement for a problem that only has 1 feature but it becomes quite difficult when we have many features since it is harder to calculate the conditional distribution P(x |).

This is where the Naïve Bayes classifier comes in, it simplifies the problem by assuming that features are independent from each other which allows us to do simplify the previous calculation:

2.

Sketch of the scatter plot of the training data, training data is not sketched since that is not asked for.

Green: Class = 1 ; Yellow: Class = 2 ; Red: Class = 3

3.

Test set error rate: 2,67%

4.

The test set error rate is very low which concludes that the obtained Naïve Bayes Classifier gives a reliable classification.

Part 2

Section 1 code:

scores(languageindex) = 0;

Section 2 code:

scores(languageindex) = scores(languageindex) + trigramcount;

Section 3 code:

scores(languageindex) = scores(languageindex)/total\_counts(languageindex);