Theory of Naïve Bayes

Chris Miles

January 14, 2021

1 Overview

Suppose you have a dictionary of words which are indexed in alphabetical order. Let w_i be the ith word in the dictionary. We are interested in creating a feature vector $\mathbf{x}$ which has elements x_i where $x_i = 1$ if w_i is in the document or $x_i = 0$ if w_i in the document. We are interested in classifying a document with the probability:

$$p(\mathbf{x}|y) = p(x_1, x_2, \ldots, x_{10000}|y) = p(x_1|y)p(x_2|y)\ldots p(x_n|y)$$

The above expression for the conditional joint probability is not true in general. The core idea of Naïve Bayes is to make the assumption that the conditional probability of seeing a word w_i and the w_j only depends on the class y.