Naive Bayes

Kristian Wichmann

May 18, 2017

1   Example: Document classification

Assume the corpus C is divided into k of disjoint categories $C_1, C_2, \ldots, C_k$. We now want to develop a way to evaluate in which category a new document d should be classified.

For the *naive Bayes* classifier, we assume that the occurence of a word w_i in a document in C_j is stochastic, and happens independently of the occurence of other words. It is naive in the sense, that this is obviously not true. But it may still yield good results.

1.1   Theory

The new document d can be written $d = (w_1, \ldots, w_n)$, where all the words w_i are in the vocabulary V. Now, consider the probability that a random document from C_j contains all the words of d. According to the assumptions of naive Bayes, this is simply a product of the probabilities of each individual word being in a random document from C_j:

$$p(d|C_j) = \prod_{i=1}^{n} p(w_i|C_j) \tag{1.1}$$

Of course, the probability we're interested in is $p(C_j|d)$: The probability of being in C_j assuming that the words in d are all present. To calculate this, *Bayes' theorem* is used. The general version is:

$$p(A|B) \propto p(B|A)p(A) \tag{1.2}$$

Since we're only interested in relative probabilities (i.e. which j is most likely?), the normalization constant is irrelevant here. In this case:

$$p(C_j|d) \propto p(d|C_j)p(C_j) = p(C_j) \prod_{i=1}^{n} p(w_i|C_j) \tag{1.3}$$