Supervised Learning: Naive Bayes Theorem

April 16, 2021

1 Motivation

We need to determine the likelihood of pulling a certain color ball from a bag. Without any prior information on the frequency of colors are inside the bag, it is reasonable to assume, with this information, that for a known number of colors, n, we can say the probability of pulling any given color $c_i \in C$ is $\frac{1}{n}$, where C is the set of possible colors in the bag. If we are given information on the frequency of colors in the bag, like, the frequency of colors on different days, can we determine the likelihood of a ball of color c_i being pulled out, given that we are doing the event on a Tuesday? Can we generalize this formula?

2 Bayes' Theorem

The probability of some event A happening, given the prior information that B happens, is $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$. This theorem is quite simple and to the point, but it has many implications. For two sets, A, A^c $P(A|B) = \frac{P(B|A)P(A)}{P(A)P(B|A)+P(A^c)P(B|A^c)}$. This is better covered in a statistics book.

2.1 False Positives

Bayes' Theorem helps especially with misleading data. If you have a virus that occurs every 1 in n ($n >>> 1$) people, and a test for the virus that is p% accurate at predicting $p = 100 + \epsilon, \epsilon << 1$, testing more people will make it such that there will be more false positives than actual positives (try with odds of 1 in 10,000 and 99% accurate test for virus, testing on 1 million people). So, a positive test does not tell us that you have the virus, because the chance is still in your favor to not have the virus, it just says you need to be further tested. A negative test, however, means it is very likely you do not have the virus. While we are talking givens, I feel it is apt to point out that if having the virus (V) shows symptoms (S), then our $P(V|S)$ becomes much higher, but the experiment earlier assumes no symptoms or a harmless virus.

3 Algorithm

This isn't going to have pseudo-code because it's a lot more about processing the data you have than an algorithm.

3.1 The Naive Assumption

The naive assumption consists of two parts:
$P(A \cap B) = P(A)P(B)$ (The events are independent)
$P(A|B)P(B) = \frac{P(B|A)P(A)}{P(B)} \rightarrow P(A|B) \propto \frac{P(B|A)P(A)}{P(B)}$

3.2 Calculation

If we are interested in the probability of some event E happening, given some set $s_i \in S$ is occurring, $P(E|S) \propto P(S|E)P(E) = \Pi P(s_i|E) * P(E)$, where all values to the right of the proportional to symbol are calculated using training data.