Naive Bayes

Sylesh Suresh

March 2018

1 Introduction

Naive Bayes is a simple supervised classifier based on Bayes' theorem that, despite its assumption that there is independence between every pair of features of the input given its classification, tends to perform quite well.

2 Bayes' Theorem

Bayes' Theorem is as follows:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

This means, then, that for a class y and for categorical features $x_1, ..., x_n$:

$$P(y|x_1, ..., x_n) = \frac{P(y)P(x_1, ..., x_n|y)}{P(x_1, ..., x_n)}$$

The goal is to find the class y that maximizes this expression, which will be our class prediction.

The expression above can be simplified to:

$$P(y|x_1, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, ..., x_n)}$$

$P(x_1, ..., x_n)$ does not depend on y, so we can simply maximize $P(y)\prod_{i=1}^{n} P(x_i|y)$ with respect to y, which will yield our class prediction. The probabilities in the expression can be calculated off the training data.

3 Continuous Features

On the other hand, for continuous features, we must define $P(x_i|y)$ with a probability distribution. One way to define this probability distribution is with the Gaussian distribution. If the training data has a continuous feature x_i, first, we will separate the training data samples based on its class y. For each set of samples with class y, we must calculate the mean μ_y and the variance σ_y^2. We can now define $P(x_i|y)$ as:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{\frac{-(x_i-\mu_y)^2}{2\sigma_y^2}}$$

This allows us to solve the previously defined maximization problem to calculate our class prediction.