Naïve Bayes

Epoch IIT Hyderabad

Arin Aggarwal
MA22BTECH11006

1 Introduction

Naïve Bayes is an ML algorithm used for classification problems. Naïve because it assumes independence in its features and bayes because it is based on the bayes theorem. The naivety gives it a high bias but because of its effectiveness it has low variance.

Bayes Theorem:-

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

2 Multinomial Naïve Bayes

It is used when features have discrete values. Let there be 2 classes, A & B and there be 2 features x & y. Then to find if an observation containing features x & y will belong to which class, we compute for A

$$p(A) * p(x|A) * p(y|A)$$

where $p(A)$ is called **prior probability** and the rest are called **posterior probability**. Similar calculation is done for B. Upon comparison the class with larger value is the one with greater possibilty for containing features x & y.

Now, it may be the case that $p(y|A) >> p(y|B)$ and $p(A) > p(B)$ and intuitively we should predict A as the most probable case but $p(x|A) = 0$. This skews the whole probability. Thus to solve this, α number of observations are added to all features in A & B. This way no probability equals to 0. This is known as **Laplace Smoothing**. Most common value for α is 1 but optimal values can be found using Cross-Validation.

As we are multiplying many probabilities, it may happen that the number turns out so small that computer approximates it as 0. That is why sum of $log(probability)$ are also sometimes used for comparisons.

3 Gaussian Naïve Bayes

This is used when continuous numerical features are distributed normally. It is similar to multinomial but the likelihood of the features is assumed to be Gaussian, hence, conditional probability is given by:

$$p(x|A) = \frac{1}{\sqrt{2}\pi\sigma_A^2} e^{-\frac{(x-\mu_A)^2}{2\sigma_y^2}}$$

where

$$\sigma = Standard\ Deviation$$

$$\mu = Mean$$