Naive Bayes Classifier

Dipon Talukder
Assessment Developer (Data and AI), Workera.ai
Research Assistant, Universal Machine.io
Adjunct Faculty, East Delta University

Introduction

- Also known as a probabilistic classifier since it is based on Bayes' Theorem.
- Allows us to "invert" conditional probabilities.
- Assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. (V.V.I.)

Bayes Theorem Example

Bag 1 contain 2 Red and 3 Black balls. Bag 2 contain 3 Red and 4 Black Balls. One ball is drawn at random from one of these bags and its red. Find the probability that it is drawn from Bag 1.

Note: Do we know how to calculate P(Picking Red Ball) ?

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Actual Naive Bayes Formula

$$P(Y \mid X_1, X_2, \ldots\ldots X_n) = \frac{P(X_1 \mid Y) * P(X_2 \mid Y) * P(X_3 \mid Y) \ldots\ldots P(X_n \mid Y) * P(Y)}{P(X_1) * P(X_2) * P(X_3) \ldots\ldots P(X_n)}$$

Naive Bayes Example

Problem: If the weather is sunny, then the Player should play or not?

Dataset:

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Likelihood table:

Weather	No	Yes	
Overcast	0	5	5/14= 0.35
Rainy	2	2	4/14=0.29
Sunny	2	3	5/14=0.35
All	4/14=0.29	10/14=0.71	

Target: P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)

P(Sunny|Yes)= 3/10= 0.3
P(Sunny)= 0.35
P(Yes)=0.71
So P(Yes|Sunny) = 0.3*0.71/0.35= 0.60

Advantages

- It is easy and fast to predict class of test data set. It also perform well in multi class prediction
- When assumption of independence holds, the classifier performs better compared to other machine learning models like logistic regression or decision tree, and requires less training data.
- It perform well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

Disadvantages

- If categorical variable has a category (in test set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as “Zero Frequency”. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
- On the other side, Naive Bayes is also known as a bad estimator, so the probability outputs from predict_proba are not to be taken too seriously.
- Another limitation of this algorithm is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

Tips for Performance Improvement

- If continuous features do not have normal distribution, we should use transformation or different methods to convert it in normal distribution.
- If test data set has zero frequency issue, apply smoothing techniques “Laplace Correction” to predict the class of test data set.
- Remove correlated features, as the highly correlated features are voted twice in the model and it can lead to over inflating importance.

Variations of Naive Bayes

There are three types of Naive Bayes Model, which are given below:

- **Gaussian:** The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.
- **Multinomial:** The Multinomial Naïve Bayes classifier is used when the data is multinomial distributed. It is primarily used for document classification problems, it means a particular document belongs to which category such as Sports, Politics, education, etc. The classifier uses the frequency of words for the predictors.
- **Bernoulli:** The Bernoulli classifier works similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

Let’s Code!!!