Naive Bayes Report

Donal Loitam

July 20, 2022

Contents

1 Introduction

- A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task
- The crux of the classifier is based on the **Bayes theorem**
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Explaining the terms:
Naive : It assumes the occurence of a certain feature is independent of the occurence of other features i.e any pair of features are independent

$$p(X_2|X_1, y) = p(X_2|y) \qquad \because \text{[X1,X2 independent]} \tag{1}$$

Bayes Theorem : Now we will understand this rule, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.

$$P(H|E) = \frac{P(H).P(E|H)}{P(E)} \tag{2}$$

where , $P(H|E)$ = Posterior probability: Probability a hypothesis H is true given some evidence E.
$P(E|H)$ = Likelihood : Probability of the evidence if the hypothesis is true,
$P(H)$ = Prior : Prob a hypothesis is True (before any evidence)
$P(E)$ = Probability of seeing the evidence

Note: The assumptions made by Naïve Bayes are generally not correct in real-world situations. The independence assumption is never correct but often works well in practice

2 Key Points of the Algorithm

Let us take an example to get some better idea about the algorithm. Consider the following example,which is essentially a Car theft problem with attributes Color, Tyoe, Origin, and the target(Stolen) can either be Yes/No.

Illustration table:

Color	Type	Origin	Stolen ?
red	sports	Domestic	Yes
red	sports	Domestic	No
red	sports	Domestic	Yes
yellow	sports	Domestic	No
yellow	sports	Imported	Yes
yellow	SUV	Imported	No
yellow	SUV	Imported	Yes
yellow	SUV	Domestic	No
red	SUV	Imported	No
red	sports	Imported	Yes

Let us now understand about the assumptions made with the help of this example :

- No pair of features are dependent : example , the color being "yellow" has nothing to do with the origin of the car

GOAL: to classify a Red Domestic SUV is getting stolen or not
Note that there is no example of a red domestic SUV in our example.
Solution : According to the example, we can rewrite Bayes Theorem as :

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \tag{3}$$

$$\text{where, } X = (x_1, x_2, ..., x_n) \tag{4}$$

The variable y is the class variable(stolen?), which represents if the car is stolen or not given the conditions. In our example, x_1, x_2, x_3 = Color, Type, Origin. Also,

$$\begin{aligned}
P(X|y) &= P(x_1, x_2, x_3|y) & (5)\\
&= P(x_1|y)P(x_2|x_1, y)P(x_3|x_1, x_2, y) & (6)\\
&= P(x_1|y)P(x_2|y)P(x_3|y) \qquad [\because x_i's \text{ are independent}] & (7)\\
P(X) &= P(x_1, x_2, x_3) = P(x_1)P(x_2)P(x_3) & (8)
\end{aligned}$$

Hence eq 3 can be rewritten as :-

$$P(y|X) = \frac{P(x_1|y)P(x_2|y)P(x_3|y)P(y)}{P(x_1)P(x_2)P(x_3)} \tag{9}$$

$$\tag{10}$$

Our aim is to find the maximum between $P(Y = Yes|X)$ and $P(Y = \text{No}|X)$. From eq 9 , we can deduce that $P(y|X) \propto P(y)\prod_{i=1}^{3} P(x_i|y)$

- First, we create a frequency table for eaxh attribute against the target.
- Then , we mold the freq tables to Likelihood Tables

Below are the Frequency and likelihood tables for all three predictors.

Frequency Table

		Stolen?	
		Yes	No
Color	Red	3	2
	Yellow	2	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Color	Red	3/5	2/5
	Yellow	2/5	3/5

Figure 1: Frequency and Likelihood tables of 'Color'

Frequency Table

		Stolen?	
		Yes	No
Type	Sports	4	2
	SUV	1	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Type	Sports	4/5	2/5
	SUV	1/5	3/5

Figure 2: Frequency and Likelihood tables of 'Type'

Frequency Table

		Stolen?	
		Yes	No
Origin	Domestic	2	3
	Imported	3	2

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Origin	Domestic	2/5	3/5
	Imported	3/5	2/5

Figure 3: Frequency and Likelihood tables of 'Origin'

So in our example, we have 3 predictors X.

Color	Type	Origin	Stolen
Red	SUV	Domestic	?

As per the equations discussed above, we can calculate the posterior probability $P(Yes|X)$ as :

P(Yes | X) = P(Red | Yes) * P(SUV | Yes) * P(Domestic | Yes) * P(Yes)

= ⅗ * ⅕ * ⅖ * 1
= 0.048

and, $P(No|X)$:

P(No | X) = P(Red | No) * P(SUV | No) * P(Domestic | No) * P(No)

= ⅖ * ⅗ * ⅗ * 1
= 0.144

Since $0.144 > 0.048$, Which means given the features Red SUV and Domestic, our example gets classified as 'NO' the car is not stolen.

3 Some Questions

1. What are some benefits of Naive Bayes?
Ans. It works better than simple algorithms like logistic regression etc. It also works well with categorical data and with numerical data as well

- It performs well with both clean and noisy data.
- Training takes a few samples, but the fundamental assumption is that the training dataset is a genuine representation of the population.
- Obtaining the likelihood of a forecast is simple.

2. What are the cons of Naive Bayes classifier?
Ans. Some of the cons are :

- Naive Bayes classifiers suffer from **"Zero Frequency"** problem. This happens when a category is not present in the training set. It will give it 0 probability.
- The consideration of features as independent of each other because in real life it is impossible to get independent features

3. Is Naive Bayes is a discriminative classifier or generative classifier?
Ans Naive Bayes is a generative classifier.
It learns from the actual distribution of the dataset by performing operations on it.
It does not create a decision boundary to classify data.

4. While calculating the probability of a given situation, what error can we run into in Naïve Bayes and how can we solve it?
Ans. We might encounter the zero division error when the probability for a particular scenario in the numerator is zero. Possible solution is **Laplace Smoothing** which basically does :

$$\text{numerator} \rightarrow \text{numerator} + 1 \tag{11}$$

$$\text{denominator} \rightarrow \text{denomintor} + 1 \tag{12}$$

5.What is the best dataset scenario for the Naïve Bayes Classifier?
Ans If the training data is smaller or if the dataset has fewer number of observations (samples) and a high number of features. Naïve Bayes works well on this data because of its High bias – Low variance trade off.
6. How does Naïve Bayes treat numerical and categorical values?
Ans

- For the categorical features, we can estimate our probability using a distribution such as multinomial or Bernoulli.
- For the numerical features, we can estimate our probability using a distribution such as Normal or Gaussian.