Naïve Bayes

Stephanie Okosa[1]

Contents

Abstract: Naïve Bayes is a simple, agile supervised learning algorithm used in machine learning, for the purpose of classifying very often high-dimensional data set. Its is a long used approach in information retrieval. As a result of its fast and few tunable parameters, it is being used mostly as a baseline for classification problems. In this paper, naïve bayes is described operating under the principle of the Bayesian method. This text giving a brief introduction into the Bayes theorem, and foundation of naïve bayes learning with explanation of some of its use case implementation, its violations and successes.

[1] stephanie-chinenye.okosa@stud.hshl.de

1 Motivation

When computers are instructed on how to learn in order to solve problems with given data inputs, they have the ability to progress beyond what they are told through diverse data models. Data models include descriptors in a table form which are; the attribute in the column of the data table called features and the examples specified in the rows of the data table. Numerous learning systems that are applied in today's world work with various learning algorithms such as, image recognition systems in self driving cars, business learning systems that enable optimization of business processes on different sites, learning systems in the medical field for example, in triage, and many other sectors using observed data. The process of building data models with tunable parameters that can conform to different observed data is known as Machine learning [Va22]. There are two main classification of machine learning: unsupervised learning and supervised learning.
Unsupervised learning entails modelling the features of a dataset without making any reference to the examples of the data model. These type of learning can be applied in pattern recognition.
Supervised learning includes modelling the relationship between observed features of data and some examples associated within the data. This modelling category is used in Naive bayes.

2 Foundation of Naïve Bayes

Naïve Bayes is known as the probabilistic classifier as it operates on the Bayesian learning using the Bayes's rule. Bayesian learning which is the Bayes Theorem is a statistical-based method that depicts a detailed equation of the conditional probabilities of statistical quantities [art]. Bayes's rule states that the probability of x given y is the probability that they happen together relative to the probability that r happens at all. It permits the transpose of conditional probability, which represents the probability of an event given that some other event has occurred as shown below.

$$p(x|y) = p(x) \ or \ p(y/x) = p(y) \tag{1}$$

$$p(x|y) = (p(y|x))/(p(y)) \tag{2}$$

$$p(y) = p(y|x) * p(x) + p(y|x^c) * p(x^c) \tag{3}$$

where $x^c = x_2 \cup x_3 \cup \ldots x_n$, and redefining $x = x_1$

The conditional probability p(x|y) is the joint probability of x and y divided by the probability of y as shown in equation 4. Assuming that $x \neq 0$ and $y \neq 0$,

$$p(x|y) = p(x \cap y)/p(y) \tag{4}$$

From equation 4, it is obvious that,

$$p(x \cap y) = p(x|y) * p(y) = p(y|x) * p(x) \tag{5}$$

Therefore,

$$p(x|y) = \frac{p(y|x) * p(x)}{p(y)} \tag{6}$$

which is the Bayes theorem. If a sample space, can be divided into finite number of mutually exclusive events $x_i = x_1, x_2, \ldots, x_n$, and if y an event with p(y) > 0 is a subset of the union of all events x_i, then Bayes Formula would be:

$$p(x_i|y) = \frac{p(y|x_i) * p(x_i)}{\sum_{j=1}^{n} p(y/x_i) * p(x)} \tag{7}$$

Employing Bayes theorem stemming from equation 7 and 3, the posterior probability of a hypothesis given observed data is

$$p(hypothesis|data) = \frac{p(data|hypothesis) * p(hypothesis)}{p(data|hypothesis) * p(hypothesis) + p(data|hypothesis^c) * p(hypothesis^c)} \tag{8}$$

where $p(hypothesis|data)$ represents the posterior probability which is the probability that an event occurs when more information or knowledge of any assumption is provided for a problem context; $p(hypothesis)$ is the prior probability that no prior information or knowledge of assumption is provided for the problem context; $p(data|hypothesis)$ is the likelihood of the data given the hypothesis; $p(data)$ is the probability of the data regardless of the given hypothesis i.e. $p(data)$ is usually constant.

3 Naïve Bayes

Observing from section 1, data models have tunable parameter. However, one of the characteristics of naive Bayes models is that they have few tunable parameters which enables them to be extremely fast and are used as a baseline for classification problems.

3.1 Naïve Bayes Classification

Naïve Bayes classifiers operate under the following assumptions:

- Each event are conditionally independent in a model
- All features contribute equally to the outcome in a model

.
They vary depending on their feature value distributions. They follow the crisp classification rule i.e. Bayes Theorem [8], which assigns each instance to exactly one class. The word "*naive*" is prescribed from the simplified assumption of the generative model used in specifying the hypothetical random process that generates the data of a model [Va22]. All types of naive bayes classifiers have different naive assumption about the data of the model. Some examples of these classifiers would be explained in this section.
The Gaussian Naïve Bayes as the name implies is applied in Gaussian distributions i.e. continuous and normal distributions by, finding the mean and standard deviation of points within each example [art]. The assumption is that data from each example is acquired from a simple Gaussian distribution [Va22]. Examples of where this method is applied is in Iris classification in section 3.2. For explicit explanation refer to [Va22].
Multinomial Naïve Bayes employs discrete data to describe the probability of observing counts of different categories. Here features are assumed to be produced from a simple multinomial distribution [Va22]. It is most appropriate for features depicting counts or count rates such as frequency counts. It can be applied in areas such as, natural processing language [art], text classification such as spam filtering [see section 3.2].
Another variant of the Naïve Bayes classifier is the Bernoulli Naïve Bayes classifier which is used with Boolean variables i.e. 1's, 0's or true,false.
The features of values with the Naïve Bayes classifier are easy and fast to estimate. Thus, Naïve Bayes is presented as a simple classifier that provides straightforward and interpretable probabilistic predictions [Va22] which are applied in both in data science and machine learning. However, the most known downside of this method is that it the conditional independence assumption does not always hold, leading to false classification [art].

3.2 Use case Implementation

Iris classification, employs the Gaussian classification method to predict flowers based on their specific features such as sepal length, sepal width, petal length and, petal width.
Another use case implementation can be seen in spam filtering which uses the multinomial distribution classification method to classify words in order to filter spam emails.

4 Conclusion

The independence assumptions in naive bayes rarely holds true for natural data sets. This has resulted in further research being carried out to produce better modifiers by easing the independent assumptions, modify feature sets in order to validate the independence assumptions, and to undertake explaining why the independence assumption is inessential.

5 Declaration of Originality

I, Stephanie Okosa, herewith declare that I have composed the present paper and work by myself and without the use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned. The paper and work in the same or similar form have not been submitted to any examination body and have not been published. This paper was not yet, even in part, used in another examination or as a course performance. I agree that my work may be checked by a plagiarism checker.

Date & Place - Stephanie Okosa

Bibliography

[art] What are Naïve Bayes classifiers? IBM.

[Va22] VanderPlas, Jake: Python Data Science Handbook, 2nd Edition. December, 2022.