

Introduction to Naive Bayes for Classification



Srishti Sawla

Jun 9, 2018 · 4 min read

If you would have read my earlier blogs, I have covered two most popular classification algorithms Linear Discriminant Analysis and K-Nearest Neighbors.

Another very popular Supervised Classification algorithm is Naive Bayes.

Before diving deep let's understand What does **Naive** and **Bayes** signify.

This algorithm is called “Naive” because it makes a naive assumption that each feature is independent of other features which is not true in real life.

As for the “Bayes” part, it refers to the statistician and philosopher, Thomas Bayes and the theorem named after him, Bayes’ theorem, which is the base for Naive Bayes Algorithm.

What is Naive Bayes Algorithm?

On Summarizing the above mentioned points Naive Bayes algorithm can be defined as a supervised classification algorithm which is based on **Bayes theorem** with an assumption of **independence among features**.

A Brief look on Bayes Theorem :



Bayes Theorem helps us to find the probability of a hypothesis given our prior knowledge.

As per wikipedia, In probability theory and statistics, **Bayes' theorem** (alternatively **Bayes' law** or **Bayes' rule**, also written as **Bayes's theorem**) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Lets look at the equation for Bayes Theorem,

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

Diagram illustrating the components of Bayes' Theorem:

- $P(A|B)$: THE PROBABILITY OF "A" BEING TRUE GIVEN THAT "B" IS TRUE
- $P(B|A)$: THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE
- $P(A)$: THE PROBABILITY OF "A" BEING TRUE
- $P(B)$: THE PROBABILITY OF "B" BEING TRUE

Where,

- $P(A|B)$ is the probability of hypothesis A given the data B. This is called the **posterior probability**.
- $P(B|A)$ is the probability of data B given that the hypothesis A was true.
- $P(A)$ is the probability of hypothesis A being true (regardless of the data). This is called the **prior probability of A**.
- $P(B)$ is the probability of the data (regardless of the hypothesis).

If you are thinking what is $P(A|B)$ or $P(B|A)$? These are conditional probabilities having formula :



If you still have confusion, this image summarizes Bayes Theorem-



How does Naive Bayes Algorithm work?

Let us take an example to understand how does Naive Bayes Algorithm work.

Suppose we have a training dataset of 1025 fruits. The features in the dataset are these : Yellow_color, Big_Size, Sweet_Taste. There are three different classes apple, banana & others.

Step 1: Create a frequency table for all features against all classes



What can we conclude from the above table?

- Out of 1025 fruits, 400 are apples, 525 are bananas, and 100 are others.
- 175 of the total 400 apples are Yellow and the rest are not and so on.

- 400 fruits are Yellow, 425 are big in size and 200 are sweet from a total of 600 fruits.

Step 2: Draw the likelihood table for the features against the classes.



In our likelihood table Total_Probability of banana is maximum(0.1544) when the fruit is of Yellow_Color,Big in size and Sweet in taste. Therefore as per Naive Bayes algorithm a fruit which is Yellow in color,big in size and sweet in taste is Banana.

In a nutshell, we say that a new element will belong to the class which will have the maximum conditional probability described above.

Pros and Cons of Naive Bayes Algorithm:

Pros :

1. It is easy to understand.
2. It can also be trained on small dataset.

Cons :

1. It has a '*Zero conditional probability Problem*', for features having zero frequency the total probability also becomes zero. There are several sample correction techniques to fix this problem such as "Laplacian Correction."
- Another disadvantage is the very strong assumption of independence class features that it makes. It is near to impossible to find such data sets in real life.

Applications of Naive Bayes Algorithm :

1. Naive Bayes is widely used for text classification

2. Another example of Text Classification where Naive Bayes is mostly used is Spam Filtering in Emails

3. Other Examples include Sentiment Analysis ,Recommender Systems etc

Hope you are liking my blogs.Feel free to reach me if you have any doubts.

Happy Learning !!!

Machine Learning

Data Science

Classification

About Help Legal