Short & Sweet (Episode 6)

Naive Bayes Algorithm

Naive Bayes is a classification algorithm that is based on **Bayes' theorem**, a fundamental concept in probability theory. The algorithm calculates the **probability** of a data point belonging to a certain class, based on the values of its features. The "**naive**" part of the name comes from the assumption that each feature is independent of every other feature, given the class variable. This simplifying assumption makes the algorithm easier to implement, but it can sometimes lead to inaccurate results. Despite this limitation, Naive Bayes has several **advantages -**

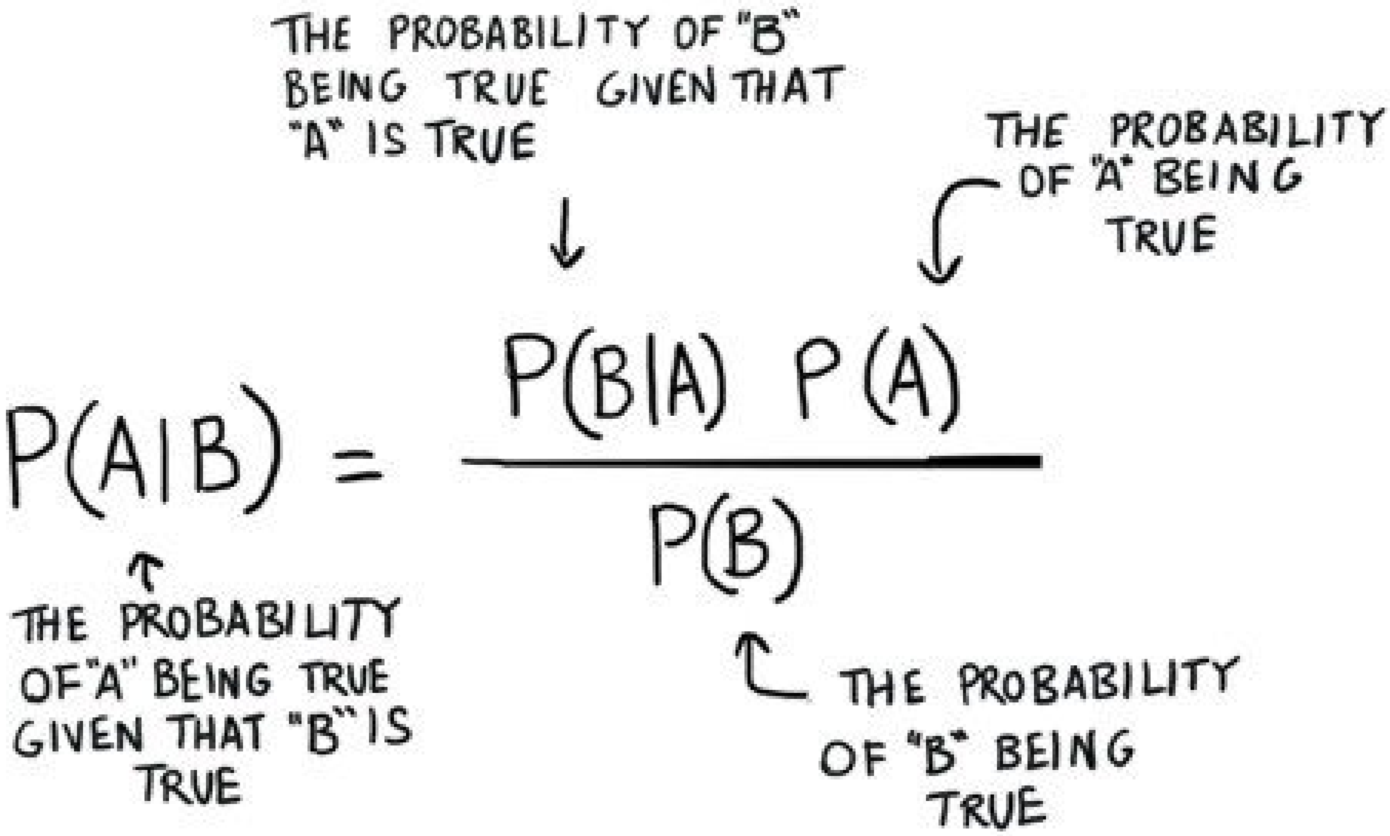
- It is simple and easy to implement, requiring only a few lines of code.
- It can also handle both continuous and discrete data, making it a versatile tool for a wide range of classification problems.
- Naive Bayes is **relatively fast** and scalable, meaning it can handle large datasets with ease.
- It is also robust to irrelevant features, meaning it can maintain good performance even when some features are not informative.

Naive Bayes can be used for both binary and multi-class classification, making it a flexible choice for a wide variety of applications.

What is Bayes Theorem?

Bayes' theorem is a fundamental concept in probability theory that describes the probability of an event based on prior knowledge of conditions that might be related to the event. It is stated as: **P(A|B) = [P(B|A) * P(A)] / P(B)**, where:

- **P(A|B)** is the posterior probability of A given B.
- **P(B|A)** is the likelihood of B given A.
- **P(A)** is the prior probability of A.
- **P(B)** is the marginal likelihood of B.

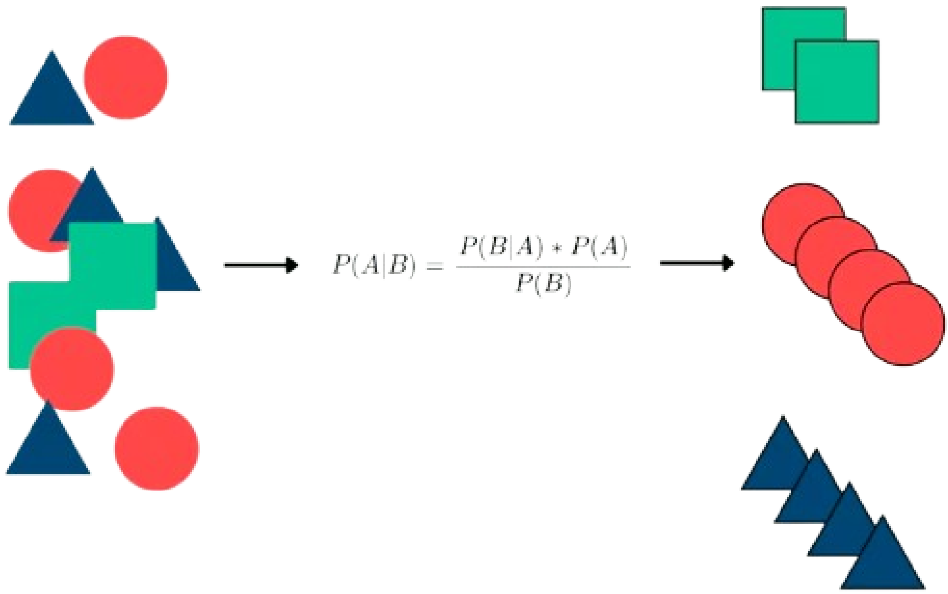


Credit goes to [pathminds]

In the context of **Naive Bayes**, **Bayes' theorem** is used to calculate the probability of a data point belonging to a certain class, based on the values of its features. Specifically, the algorithm calculates the likelihood of each feature given the class variable, and then combines these likelihoods using Bayes' theorem to calculate the posterior probability of the class variable.

How It Works?

The **Naive Bayes** algorithm works by calculating the **probability** of each class given the input features, and then assigning the data point to the class with the highest probability. The algorithm uses **Bayes' theorem** to calculate the **posterior probability** of the class variable (target variable), which is the probability of the class variable given the input features. To do this,



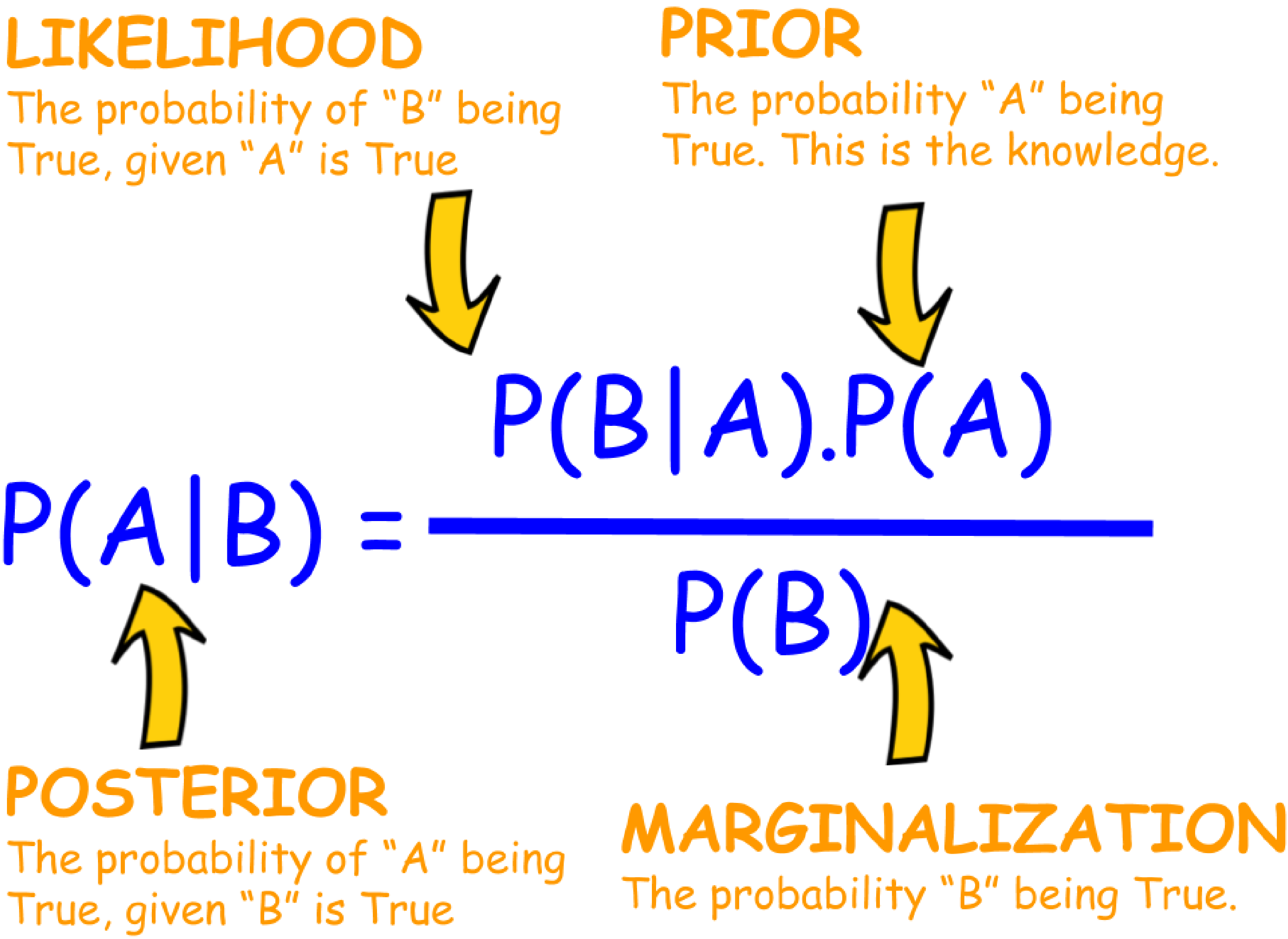
- The algorithm first calculates the **prior probability** of the class variable, which is the probability of the class variable before seeing any input features.
- Then, for each feature, the algorithm calculates the **likelihood** of the feature given the class variable. The algorithm assumes that each feature is independent of every other feature, given the class variable, which simplifies the calculation of the likelihood.
- Finally, the algorithm combines the prior probability and the likelihoods of each feature using Bayes' theorem to calculate the posterior probability of the class variable. The data point is then assigned to the class with the highest posterior probability.

What is posterior and prior probability?

You can clear get a brief through this illustration -

- **Posterior probability** - is the **updated probability** of a class occurring after considering the evidence. It's calculated using Bayes' theorem, which combines the prior probability with the probability of observing the given evidence given each class **(likelihood)**, to determine the probability of the class given that evidence.

- **Prior probability** - It refers to the **initial belief** or probability assigned to a particular class before considering any evidence. It represents what we know about the probability of a class occurring based on previous information or assumptions.



Assumptions of Naive Bayes Algorithm

The **Naive Bayes** algorithm makes several assumptions as below:

- The most important assumption is that each feature is independent of every other feature, given the class variable. This assumption can sometimes lead to inaccurate results, especially when the features are correlated or dependent.
- Another assumption of Naive Bayes is that the data distribution is **normal** or **Gaussian**. This assumption is used to calculate the likelihood of each feature, and it can lead to inaccurate results when the data distribution is not normal.
- Finally, Naive Bayes assumes that there is **zero variance** in the data distribution. This assumption is used to calculate the prior probability of the class variable, and it can lead to inaccurate results when the prior probability is very small or very large.