**Naive Bayes Theory in Machine Learning**

**1. Introduction to Naive Bayes**

Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes that the features are conditionally independent given the class label. This strong independence assumption simplifies computations and makes the algorithm efficient.

**2. Bayes' Theorem**

Bayes' Theorem describes the probability of a hypothesis A given new evidence B. It is formulated as:

P(A∣B)=P(B∣A)⋅P(A)P(B)/P(A|B)

Where:

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

**3. Conditional Independence Assumption**

The naive assumption is that features x1,x2,...,xnx\_1, x\_2, ..., x\_nx1​,x2​,...,xn​ are conditionally independent given the class C:

P(x1,x2,...,x(n)∣C)=P(x1∣C)⋅P(x2∣C)⋅...⋅P(x(n)∣C)

This simplifies the calculation of the posterior probability.

**4. Naive Bayes Classification**

**1. Gaussian Naive Bayes**

**Use Case**: Continuous features.

**Assumption**: The features follow a normal (Gaussian) distribution.

**Formula Explanation**:

* To calculate the likelihood of a feature given a class, you determine the likelihood of that feature's value assuming the values follow a bell-curve distribution.
* This involves calculating how many standard deviations the feature's value is away from the class's mean and then adjusting for the distribution's spread.

**2. Multinomial Naive Bayes**

**Use Case**: Discrete features, such as word counts in text classification.

**Assumption**: Features are distributed according to a multinomial distribution.

**Formula Explanation**:

* To calculate the likelihood of a feature given a class, you count how often that feature occurs in all instances of the class, then add a smoothing parameter (usually 1) to avoid zero probabilities.
* You then divide this count by the total count of all features in that class, also adjusted by the smoothing parameter multiplied by the number of possible features (e.g., words in vocabulary).

**3. Bernoulli Naive Bayes**

**Use Case**: Binary/boolean features, indicating the presence or absence of a feature.

**Assumption**: Features follow a Bernoulli distribution.

**Formula Explanation**:

* To calculate the likelihood of a feature given a class, you determine the probability that the feature is present in instances of the class.
* If the feature is present, you use the probability directly; if the feature is absent, you use one minus the probability of the feature being present.