Naive-Bayes Classification Algorithm

Naive-Bayes is a family of probabilistic algorithms based on Bayes' Theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. It is particularly suited for large datasets and is often used for text classification (for example, spam detection).

Types of Naive-Bayes Classifiers

1. **Gaussian Naive-Bayes**: Assumes that the continuous values associated with each feature are distributed according to a Gaussian (normal) distribution.
2. **Multinomial Naive-Bayes**: Used for discrete data (for example, word counts in documents).
3. **Bernoulli Naive-Bayes**: Used for binary/boolean features.

Bayes' Theorem

Bayes' Theorem provides a way of calculating the posterior probability, P(Y | X), from P(X), P(Y), P(X | Y):

$$P(Y|X) = \frac{P(X \mid Y) * P(Y)}{P(X)}$$

where:

- P(Y | X) is the posterior probability of each class in Y given predictor X.
- P(Y) is the prior probability of label Y.
- P(X | Y) is the likelihood which is the probability of predictor X given a class in Y.
- P(X) is the prior probability of predictor X.

Naive-Bayes Assumptions

The primary assumption of Naive-Bayes is that features are conditionally independent:

$$P(X1, X2, ..., Xn \mid Y) = P(X1 \mid Y) * P(X2 \mid Y) * ... * P(Xn \mid Y)$$

Mathematics of Naive-Bayes

Given a set of features X=(X1, X2, ..., Xn), we want to determine the probability of a class in target variable Y given these features. Using Bayes' theorem, we can write:

$$P(Y \mid X) = \frac{P(X \mid Y) * P(Y)}{P(X)}$$

Since P(X) is constant for all classes, we can maximize P(Y | X) by maximizing the numerator:

$$P(Y \mid X) \propto P(X \mid Y) \cdot P(Y)$$

Using the independence assumption, we can write:

$P(X \mid Y) = P(X1 \mid Y) * P(X2 \mid Y) * ... * P(Xn \mid Y)$

The label(class) Y that maximizes this posterior probability is the predicted class.

Training the Naive-Bayes Classifier

1. **Calculate Priors**: Estimate the prior probability P(Y) for each class in Y.
2. **Calculate Likelihoods**: Estimate the likelihood $P(Xi \mid Y)$ for each feature Xi given each class in Y.