

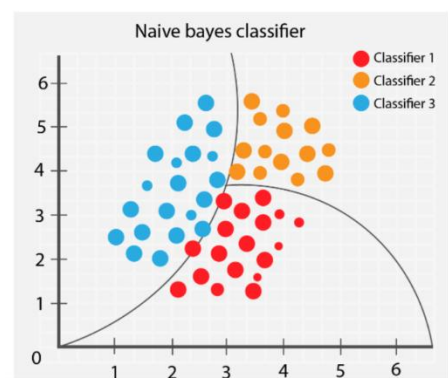
EXPERIMENT NO: 6

Aim: Apply the Naive Bayes machine learning algorithm for classification tasks and assess accuracy, precision, and recall.

Software Used: Google Colab

Theory:

The main idea behind the Naive Bayes classifier is to use Bayes' Theorem to classify data based on the probabilities of different classes given the features of the data. It is used mostly in high-dimensional text classification. The Naive Bayes Classifier is a simple probabilistic classifier and it has very few numbers of parameters which are used to build the ML models that can predict at a faster speed than other classification algorithms. It is a probabilistic classifier because it assumes that one feature in the model is independent of existence of another feature. In other words, each feature contributes to the predictions with no relation between each other. Naïve Bayes Algorithm is used in spam filtration, Sentimental analysis, classifying articles and many more.



Bayes' Theorem provides a principled way to reverse conditional probabilities. It is defined as:

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)}$$

Where:

- $P(y|X)$: Posterior probability, probability of class y given features X
- $P(X|y)$: Likelihood, probability of features X given class y
- $P(y)$: Prior probability of class y
- $P(X)$: Marginal likelihood or evidence

Types of Naive Bayes Model:

1. Gaussian Naive Bayes

In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution. A Gaussian distribution is also called Normal distribution. When plotted, it gives a bell shaped curve which is symmetric about the mean of the feature values as shown below:

2. Multinomial Naive Bayes

Multinomial Naive Bayes is used when features represent the frequency of terms (such as word counts) in a document. It is commonly applied in text classification, where term frequencies are important.

3. Bernoulli Naive Bayes

Bernoulli Naive Bayes deals with binary features, where each feature indicates whether a word appears or not in a document. It is suited for scenarios where the presence or absence of terms is more relevant than their frequency. Both models are widely used in document classification tasks

Advantages of Naive Bayes Classifier

- Easy to implement and computationally efficient.
- Effective in cases with a large number of features.
- Performs well even with limited training data.
- It performs well in the presence of categorical features.
- For numerical features data is assumed to come from normal distributions

Disadvantages of Naive Bayes Classifier

- Assumes that features are independent, which may not always hold in real-world data.
- Can be influenced by irrelevant attributes.
- May assign zero probability to unseen events, leading to poor generalization.

Applications of Naive Bayes Classifier

- Spam Email Filtering: Classifies emails as spam or non-spam based on features.
- Text Classification: Used in sentiment analysis, document categorization, and topic classification.
- Medical Diagnosis: Helps in predicting the likelihood of a disease based on symptoms.
- Credit Scoring: Evaluates creditworthiness of individuals for loan approval.
- Weather Prediction: Classifies weather conditions based on various factors.

Conclusion: Hence, we learned about apply the Naive Bayes machine learning algorithm for classification tasks and assess accuracy, precision, and recall.