Bayesian Classification

Naive Bayes Classifier

➢Naive Bayes is a probabilistic machine learning algorithm that can be used in a wide variety of classification tasks.

➢Bayesian classifier can **predict the probability** that a given tuple belongs to a particular class

➢Naive Bayes classifiers are based on Bayes' Theorem.

➢Bayesian classifier gives more speed and accuracy as compared to Decision tree classifier.

Bayes' theorem

➢Bayes' theorem provides a way to calculate the probability of an event, based on prior knowledge related to that event.

➢It's expressed as:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

➢P(A|B) is the **posterior probability** (Probability of hypothesis A on the observed event B.)

➢P(B|A) is the **likelihood probability** (Probability of the evidence given that the probability of a hypothesis is true)

➢P(A) and P(B) are the probabilities of events A and B.

Finding the probability of having a disease given you were tested positive
Finding the probability of liking Harry Potter given we know the person likes fiction

Why is it called 'Naive'?

➢The name naive is used because it assumes the features that go into the model is independent of each other.

➢That is changing the value of one feature, does not directly influence or change the value of any of the other features used in the algorithm.

Naive Bayes Classifiers

➢ The dataset is divided into two parts, namely

➢ Feature matrix

➢ Response vector

➢ **Feature matrix** contains all the vectors (columns) of dataset in which each vector consists of the value of dependent features.

➢ **Response vector** contains the value of class variable for each row of feature matrix.

Working of Naive Bayes Classifiers

➢Convert the given dataset into frequency tables.

➢Generate Likelihood table by finding the probabilities of given features.

➢Now, use Bayes theorem to calculate the posterior probability.

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Advantages of Naive Bayes Classifiers

➢Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.

➢It can be used for Binary as well as Multi-class Classifications.

➢It performs well in Multi-class predictions as compared to the other Algorithms.

➢It is the most popular choice for text classification problems.

Disadvantages of Naive Bayes Classifiers

➢Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

Applications of Naive Bayes Classifiers

➢It is used for Credit Scoring.

➢It is used in medical data classification.

➢It can be used in real-time predictions because Naïve Bayes Classifier is an eager learner.

➢It is used in Text classification such as Spam filtering and Sentiment analysis.