

Naïve Bayes Classifier

BCSE 0105 MACHINE LEARNING

Naïve Bayes Classifier Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.**
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles.**

Bayes' Theorem:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Where,

$P(A | B)$ is Posterior probability

$P(B | A)$ is Likelihood probability

$P(A)$ is Prior Probability

$P(B)$ is Marginal Probability

Working of Naïve Bayes' Classifier:

- Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions.

Steps:

1. Convert the given dataset into **frequency tables**.
2. Generate **Likelihood table** by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the **posterior probability**.

Problem: If the weather is sunny, then the Player should play or not?

- **Solution:**

To solve this, first consider the below dataset:

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

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

Frequency table for the Weather Conditions:

Weather	No	Yes
Overcast	0	5
Rainy	2	2
Sunny	2	3

Likelihood table weather conditions:

Weather	No	Yes	
Overcast	0	5	$5/14 = 0.35$
Rainy	2	2	$4/14 = 0.29$
Sunny	2	3	$5/14 = 0.35$
All	$4/14 = 0.29$	$10/14 = 0.71$	

Applying Bayes'theorem:

- $P(\text{Yes} | \text{Sunny}) = P(\text{Sunny} | \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$
- $P(\text{Sunny} | \text{Yes}) = 3/10 = 0.3$
- $P(\text{Sunny}) = 0.35$
- $P(\text{Yes}) = 0.71$
- So $P(\text{Yes} | \text{Sunny}) = 0.3 * 0.71 / 0.35 = \mathbf{0.60}$

- $P(\text{No} | \text{Sunny}) = P(\text{Sunny} | \text{No}) * P(\text{No}) / P(\text{Sunny})$
- $P(\text{Sunny} | \text{NO}) = 2/4 = 0.5$
- $P(\text{No}) = 0.29$
- $P(\text{Sunny}) = 0.35$
- So $P(\text{No} | \text{Sunny}) = 0.5 * 0.29 / 0.35 = \mathbf{0.41}$
- So as we can see from the above calculation that $P(\text{Yes} | \text{Sunny}) > P(\text{No} | \text{Sunny})$
- Hence on a Sunny day, Player can play the game.

Advantages of Naïve Bayes Classifier:

- 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 Naïve Bayes Classifier:

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

Applications of Naïve Bayes Classifier:

- 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**.