DATA SCIENCE WITH PYTHON : NAIVE BAYES CLASSIFIER #396

PRESENTED BY:

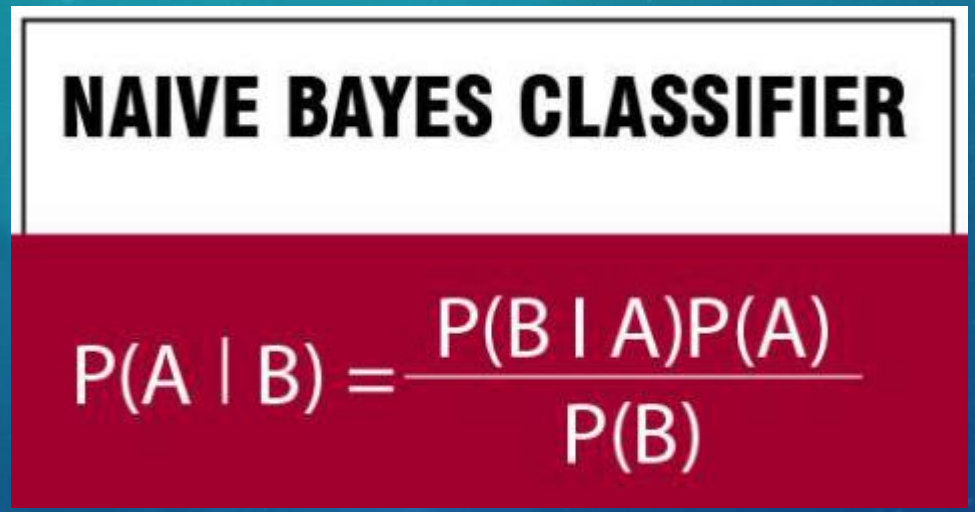
DEEPTHI M

BATCH NUMBER: 05

SERIAL NUMBER: 172

Naive Bayes Algorithms

- Naïve Bayes algorithm is a supervised machine learning algorithm.
- It is used for solving classification problems.
- It uses Bayes theorem
- It is a probabilistic classifier which means it predicts on basis of the probability.



Why it is called as Naive Bayes Algorithms?

- It is called as Naive Bayes because it assumes that occurrence of a certain feature is independent of the occurrence of other features.

- For example:
  If a fruit is recognized on basis of color, shape and taste, then apple is identified on basis of red, spherical, and sweet fruit. Hence, each feature contributes independently.
  It is called as Bayes because it uses Bayes theorem.

Bayes theorem formula:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B	= events
$P(A\|B)$	= probability of A given B is true
$P(B\|A)$	= probability of B given A is true
$P(A), P(B)$	= the independent probabilities of A and B

Bayes Theorem

Likelihood
Probability of collecting this data when our hypothesis is true

Prior
The probability of the hypothesis being true before collecting data

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Posterior
The probability of our hypothesis being true given the data collected

Marginal
What is the probability of collecting this data under all possible hypotheses?

Types of naïve Bayes algorithm:

1. Gaussian Naïve Bayes algorithm:
- It assumes that features is normally distributed. This means if predictions take continuous values instead of discrete, then model assumes that these values are sampled from the Gaussian distribution.

2. Multinomial Naive Bayes algorithm:
- It is used when the data is multi-nominal categories.

3. Bernoulli Naïve Bayes algorithm:
- It is used when predictor variables are independent Boolean variables.

Advantages :

- It is one of the fast machine learning algorithm.
- It is one of the easy machine learning algorithm.
- It is used for binary and multi-class classification.
- It performs well in multi-class predictions.
- It is the most popular text classification problems.

Disadvantages:

- It cannot learn relationship between features as it assumes all features are independent

I hope now we have an overview on the topics:

- Naïve Bayes algorithm

Thank you I hope you have enjoyed!!!