12.

Naive Bayes classifier are a collection of classification algorithms based on Baye's Theorem

It is not a single algorithm but a family of algorithms where all of them share a common principle.

Naive Bayes Assumption

Attributes that describe data instances are conditionally independent given the classification hypothesis.

$$P(d|h) = P(a_1, \ldots, a_j | h) = \prod_t P(a_t | h)$$

→ It is a simplifying assumption, obviously it may be violated in reality.

→ Inspite of that, it works well in practice

→ The Bayesian classifier that uses the Naive Baye's Assumption and computes the MAP hypothesis is called Naive Bayes classifier.

→ Successful Applications :-

→ Medical Diagnosis

→ Text classification

Example :- play Tennis Data

Day	outlook	Temperature	Humid-ity	Wind	play Tennis
Day 1	Sunny	Hot	High	Weak	No
Day 2	Sunny	Hot	High	Strong	No
Day 3	Overcast	Hot	High	Weak	Yes
Day 4	Rain	Mild	High	Weak	Yes
Day 5	Rain	Cold	Normal	Weak	Yes
Day 6	Rain	Cold	Nomal	Strong	No
Day 7	Overcast	Cold	Normal	Strong	Yes
Day 8	Sunny	Mild	High	Weak	No
Day 9	Sunny	Cold	Nomal	Weak	Yes
Day 10	Rain	Mild	Normal	Weak	Yes
Day 11	Sunny	Mild	Nomal	Strong	Yes
Day 12	Overcast	Mild	High	Strong	Yes
Day 13	Overcast	Hot	Nomal	Weak	Yes
Day 14	Rain	Mild	High	Strong	No

Using above data set

problem : players will play if whether is sunny. Is this statement correct ?

$$P(Yes/sunny) = P(sunny/Yes) * P(Yes) / P(sunny)$$

Here $P(sunny/Yes) = 3/9 = 0.33$,

$P(sunny) = 5/14 = 0.36$

$P(Yes) = 9/14$

Now, $p(yes/sunny) = \frac{\frac{3}{9} \times \frac{9}{14}}{\frac{5}{14}}$

$= 0.60$

Naive Bayes algorithm is mostly used in text classification and with problems having multiple classes.

The dataset is divided into two parts, feature matrix, response vector.

- Feature matrix consists of all the vectors, of dataset in which each vector consists of value of dependent features.
  Here features are 'Outlook', 'Temperature', 'Humidity', 'windy'
- Response vectors contains the values of class variable for each row of feature matrix.

Here, the class variable is play Tennis

The fundamental Naive Bayes Assumption is that each feature makes an

- independent
- equal

contribution to the outcome.