Naïve Bayes Classifier

Naive Bayes

- Naive Bayes is a probabilistic technique for constructing classifiers.
- The characteristic assumption of the naive Bayes classifier is to consider that the value of a particular feature is independent of the value of any other feature, given the class variable.

Application – Naiye Bayes

- To mark an email as **spam**, or **not spam** ?
- Classify a news article about **technology**, **politics**, or **sports** ?
- NLP(textual data analysis) kind of task.
- Check a piece of text expressing **positive** emotions, or **negative** motions?

Bayes theorem

- It works on conditional probability.
- Conditional probability is the probability that something will happen, given that something else has already occurred.
- Using the conditional probability, we can calculate the probability of an event using its prior knowledge.

conditional probability.

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Bayes theorem

where

- P(H) is the probability of hypothesis H being true. This is known as the prior probability.
- P(E) is the probability of the evidence(regardless of the hypothesis).
- P(E|H) is the probability of the evidence given that hypothesis is true.
- P(H|E) is the probability of the hypothesis given that the evidence is there.

Example-

Problem

- A Lab is performing a Test of disease say “D” with two results “Positive” & “Negative.”
- They guarantee that their test result is 99% accurate: if you have the disease, they will give test positive 99% of the time.
- If you don’t have the disease, they will test negative 99% of the time. If 3% of all the people have this disease and test gives “positive” result, what is the probability that you actually have the disease?

Bayes theorem

- calculating the probability that the patient actually have the disease i.e, P(D | Pos) we will use Bayes theorem

$$P(D \mid Pos) = \frac{P(Pos \mid D) * P(D)}{P(Pos)}$$

Bayes theorem

- Probability of people suffering from Disease D, P(D) = 0.03 = 3%
- Probability that test gives “positive” result and patient have the disease, P(Pos | D) = 0.99 =99%

- Probability of people not suffering from Disease D, P(~D) = 0.97 = 97%
- Probability that test gives “positive” result and patient does have the disease, P(Pos | ~D) = 0.01 =1%

Bayes theorem

We have all the values of numerator but we need to calculate P(Pos):

P(Pos) = P(D, pos) + P(~D, pos)

= P(pos|D)*P(D) + P(pos|~D)*P(~D)

= 0.99 * 0.03 + 0.01 * 0.97

= 0.0297 + 0.0097

= 0.0394

Let’s calculate, P(D | Pos) = (P(Pos | D) * P(D)) / P(Pos)

= (0.99 * 0.03) / 0.0394

= 0.753807107

So, Approximately 75% chances are there that the patient is actually suffering from disease.

Naïve Bayes Classifier

Naive Bayes is a kind of classifier which uses the Bayes Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class.

The class with the highest probability is considered as the most likely class. This is also known as **Maximum A Posteriori (MAP)**.

Naïve Bayes Classifier Algorithm

The MAP for a hypothesis is:

MAP(H)

= max(P(H|E))

= max((P(E|H)*P(H))/P(E))

= max(P(E|H)*P(H))

P(H|Multiple Features) = P(E1| H)* P(E2|H) ……*P(En|H) * P(H) / P(Multiple Features)

Advantages

- Naive Bayes Algorithm is a fast, highly scalable algorithm.
- Naive Bayes can be use for Binary and Multiclass classification. It provides different types of Naive Bayes Algorithms like Gaussian NB, Multinomial NB, Bernoulli NB.
- It is a simple algorithm that depends on doing a bunch of counts.
- Great choice for Text Classification problems. It’s a popular choice for spam email classification.
- It can be easily train on small dataset

Disadvantages

- It considers all the features to be unrelated, so it cannot learn the relationship between features.
- E.g., Let’s say Remo is going to a part. While cloth selection for the party, Remo is looking at his cupboard. Remo likes to wear a white color shirt. In Jeans, he likes to wear a brown Jeans, But Remo doesn’t like wearing a white shirt with Brown Jeans.
- Naive Bayes can learn individual features importance but can’t determine the relationship among features.