Machine Learning

Naive Bayes

Agenda

- Joint probability and conditional probability
- Conditional probability and Bayes Theorem
- What is Bayes Theorem?
- Example question & solution
- Naive Bayes
- Steps of a Naive Bayes Classifier
- Types of Naive Bayes Classifier
- Assumptions in Naive Bayes
- Summary

Let’s try to solve a problem

- A Lab is performing a test for a disease say “D” with two outputs “Positive” & “Negative”
- Lab ensures that the given test results have 99% accuracy
- They will give the test result as positive 99% of the time, if a patient has the disease
- They will give the test result as negative 99% of the time, if a patient does not have the disease
- Question: It is given that 3% of all the people have this disease in the city. A new patient goes to the Lab for test and his test result is “positive”
    - What is the probability that New Patient actually has the disease? (Note:- it depends on their positivity rate and his test results)
    - How can doctors use this information in their inferences?

Joint Probability

- Statistical measure to compute likelihood of two events happening together.
- It is defined as the probability of two events occurring together (we are not talking of causality here i.e. one event leads to another)
- Formula for joint probability of two independent events :

$$P(A \cap B) = P(A) * P(B)$$

- Example:
  - Event-1: Probability of drawing a king from a deck of cards is 4/52
  - Event-2: Probability of drawing a red colour card from a deck of cards is 26/52
  - Joint Event: **Probability of drawing a red colour king = (4/52)*(26/52) = (2/52)**

Conditional Probability

- Some events influence the occurrence of other events. So, if we know probability of occurrence on one event then we can compute probability of occurrence of another event using conditional probability.
- It is defined as the probability of an event (event not yet observed) given another event has occurred.
- Example: given the card drawn is red (an event has occurred) from a deck of cards
    - What is the probability it is a king (event not yet observed)?
    - Since the card is red, there are 26 likely values. Of these 26 possible values we are interested in king which is 2 (king of diamonds & King of heart)
    - **Thus the conditional probability that the card is a king given red card is 2 /26**

Conditional Probability and Bayes Theorem

- Bayes theorem describes the probability of an event which is based on the preceding events. The bayes theorem is the extended version of conditional probability.
- Using conditional probability, Bayes theorem provides a way to calculate posterior probability $P(A|B)$ from $P(A)$, $P(B)$ and $P(B|A)$
- The probability of event A given B is:

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

- The probability of event B given A is:

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

- Using above two equations,

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

, known as Bayes theorem.

Bayes Theorem

- Given a hypothesis H and evidence E, Bayes theorem states that the relationship between the probability of hypothesis P(H) before getting the evidence and the probability of P(H/E) of the hypothesis after getting the evidence is,

$$P(H_i/E) \quad \frac{P(E/H_i)*P(H_i)}{P(E/H_1)P(H_1)+P(E/H_2)P(H_2)+...+P(E/H_k)P(H_k)}$$

Where,

H = i^{th} event of K mutually exclusive and collectively exhaustive events

E = new event that might impact P(H)

Let’s try to solve our example

- Probability of people suffering from disease D:
  - P(D) = 0.03 = 3%
- Probability that test gives ‘positive’ result and patient has the disease:
  - P(Positive | Disease) = 0.99 = 99%
- Probability of people not suffering from disease:
  - P(~Disease) = 0.97 = 97%
- Probability that test gives ‘positive’ result and patient does not have the disease,
  - P(Positive | ~ Disease) = 0.01 = 1%
- The probability that the patient actually has the disease given that he is tested positive:
  - P(Disease|Positive) = P(Positive|Disease) * P(Disease) / P(Positive)
- We have all the values of the previous equation except P(Positive), let’s calculate it.

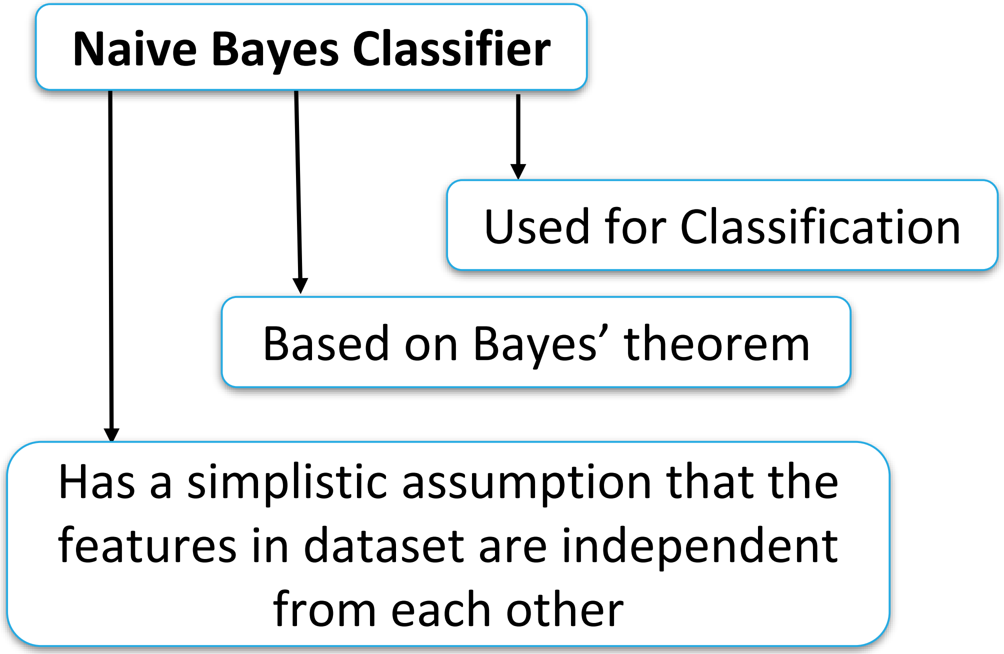
Example Solution

- P(Positive) = P(D∩Positive) + P((~D)∩Positive))
- P(Positive) = P(Positive|D) ∗ (PD) + P(Positive|(~D)) ∗ P(~D)
  - P(Positive) = 0.99 ∗ 0.03 + 0.01 ∗ 0.97
  - P(Positive) = 0.0297 + 0.0097 = 0.0394
- P(Disease|Positive) = P(Positive|Disease) ∗ P(Disease) / P(Positive)
  - P(Disease|Positive) = 0.99 ∗ 0.030.0394 = 0.753807107

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

Naive Bayes

- It estimates conditional probability which is the probability that event A will happen, given that event B has already occurred.
- Hence, used to classify target column based on given features
- It assumes that the features are independent.
- It is easy to implement, fast, accurate, and can be used to make real time predictions.



Naive Bayes

- The Bayes rule computes the probability of class Y for given X features. In real life problems, the target class Y depends on multiple X variables. So, the formula of bayes rule can be extended for multiple input features like,

$$P(Y|X1,X2,X3,....Xn) = \frac{P(X1, X2, X3,....Xn|Y) * P(Y)}{P(X1, X2, X3,.....Xn)}$$

According to the naive bayes assumption of input features are independent,

$$P(Y|X1,X2,X3,....Xn) = \frac{P(X1|Y) * P(X2|Y) *......P(Xn|Y) * P(Y)}{P(X1) * P(X2) * P(X3)*.....P(Xn)}$$

Steps of a Naive Bayes Classifier

Compute the prior probabilities for given class labels.

↓

Compute the Likelihood of evidence with each attribute for each class.

↓

Calculate the posterior probabilities using Bayes rule.

↓

Select the class which has higher probability for given inputs.

Types of Naive Bayes Classifier

Gaussian NB	Used when features are continuous. Assumes that variable x is normally distributed which is used to calculate probability density.
MultiNomial NB	Good for text classification. When data is multinomial distributed. This classifier treats each occurrence of a word as an event.
Bernoulli NB	Used when you have multiple features that are assumed to be binary.

Naive Bayes Classification- Assumptions

- **Assumption 1:** The predictors are independent of each other.

  Example: Consider the example of classifying a patient who has the disease or not, with attributes gender, age, blood pressure, and sugar level.

  The gender being “male” has nothing to do with age, blood pressure or sugar level. Hence the predictors are assumed to be independent.

- **Assumption 2:** All the predictors have an equal effect on the outcome.

  Example: each predictor is given the equal importance. For example knowing only the gender and age cannot predict the outcome perfectly. Hence the predictors are assumed to have an equal effect on the outcome.

Advantages

- Simple, fast in processing and effective.
- Does well with noisy data and missing data.
- Requires few examples for training (assuming the data set is a true representative of the population).
- Easy to obtain estimated probability for a prediction.

Disadvantages

- Relies on and often incorrect assumption of independent features.
- Not ideal for data sets with large number of numerical attributes.
- Estimated probabilities are less reliable in practice than predicted classes.
- If rare predictor value is not captured in the training set but appears in the test set the probability calculation will be incorrect.

Summary

In this module we discussed:

- Joint probability and conditional probability
- Bayes Theorem
- Naive Bayes and connection with conditional probability
- Types of Naive Bayes Classifier
- Assumptions in Naive bayes
- Advantages & disadvantages

THANK YOU
Happy learning 😊