NAÏVE BAYES

Naïve Bayes Theory

Naïve Bayes is a classification technique based on Bayes theorem with an assumption of independence among features. It is a way to figure out conditional probability.

Conditional Probability is the probability of an event happening given that it has some relationship to one or more other events.

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

$$P(H|E) = P(E|H).\ P(H) / P(E)$$

Prior Probability P(H): How probable was our hypothesis before observing the evidence?

Posterior Probability P(H|E): How probable is our hypothesis, given the observed evidence?

Likelihood: How probable is the evidence given that our hypothesis is true?

Marginal: How probable is the new evidence under all possible hypothesis?

Types of Naïve Bayes Model

Gaussian: It is used in classification and it assumes that features follow a normal distribution.

Multinomial: It is used for discrete counts.

Bernoulli: The binomial model is useful if your feature vectors are binary (i.e., zeros and ones).

Naïve Bayes with Python

Data Source: https://www.kaggle.com/kumargh/pimaindiansdiabetescsv

This dataset describes the medical records for Pima Indians and whether or not each patient will have an onset of diabetes within +ve years.

Fields description follow:

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

BloodPressure: Diastolic blood pressure (mm Hg)

SkinThickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction: Diabetes pedigree function

Age: Age (years)

Class: Class variable (1: tested positive for diabetes, 0: tested negative for diabetes)