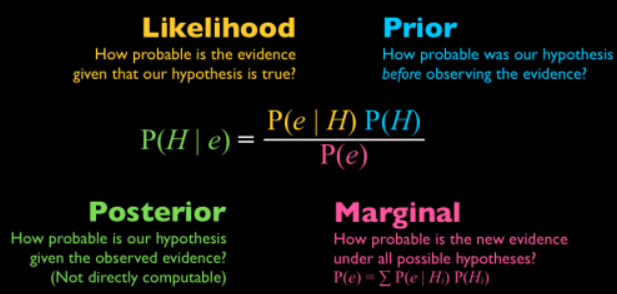
Naïve Bayes

Introduction:

It makes a naïve assumption that each feature is independent of other feature which is not true in reality that’s why its called Naïve and bayes refers to the statistician Thomas Bayes and the theorem named after him.

Bayes theorem describes the probability of an event given our prior knowledge.

Algorithm:



* P(A|B): Posterior Probability: It is the probability of our hypothesis after given the observed evidence.
* P(B|A): Data Likelihood: Probability of our evidence given our hypothesis is true.
* P(A): Prior Probability: It is the probability of our Hypothesis before observing the evidence.
* P(B): Posterior Probability: It is the probability of evidence.

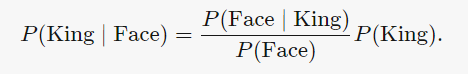
Conditional Probability with an example:

Conditional probability of A given B, it denotes the probability of A occurring given that B has already occurred.

Mathematically, Conditional probability of A given B can be computed as: P(A|B) = P(A AND B) / P(B)

Example:  To Calculate the probability of getting a king given the card is a face card.

*P*(King)= 4/52 ​=1/13​.



P(Face|King) = 1

P(king) = 1/13

P(Face) = 3/13

Applying bayes theorem gives P(King| Face) = = .

Types of conditional probability:

* Gaussian Naïve Bayes: It is simplest naïve Bayes classifier having the assumption that data is from each label is drawn from a simple gaussian distribution
* Multinomial Naïve Bayes: Here features are assumed to be drawn from simple multinomial distribution it is more appropriate for the features that represents discrete counts.
* Bernoulli Naïve Bayes: Here features are 0s and 1s . Text classification with ‘bag of words’ model can be an application of Bernoulli Naïve Bayes.

Applications:

* Real time prediction
* Multi-class prediction
* Text-classification

Advantages:

* It is easy to implement and fast.
* It requires less training data
* It can handle continuous as well as discrete data.
* It can be used for both binary and multi-class classification problems.

Disadvantages:

* Important cons of naïve Bayes classification is its strong feature independences however in real time it is almost impossible to have a set of features which are completely independent.
* Another con in naïve Bayes classification is its ‘Zero frequency’ i.e.., if a categorical variable has a category but not being not being observed in training data set. for features having zero frequency the total probability also becomes zero.