Naive Bayes Algorithm

- Naive Bayes is a powerful algorithm used for classification tasks in machine learning.
- It is based on applying Bayes' theorem with a crucial assumption of strongindependence between features.
- The algorithm leverages Bayes' theorem to make predictions.
- Bayes' theorem calculates the probability of a hypothesis based on priorprobabilities.
- Naïve Bayes models are also known as simple Bayes or independent Bayes.

Naive Bayes algorithm intuition

- ✓ Naïve Bayes Classifier uses the Bayes' theorem to predict 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 the **Maximum A Posteriori (MAP)**.

Formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

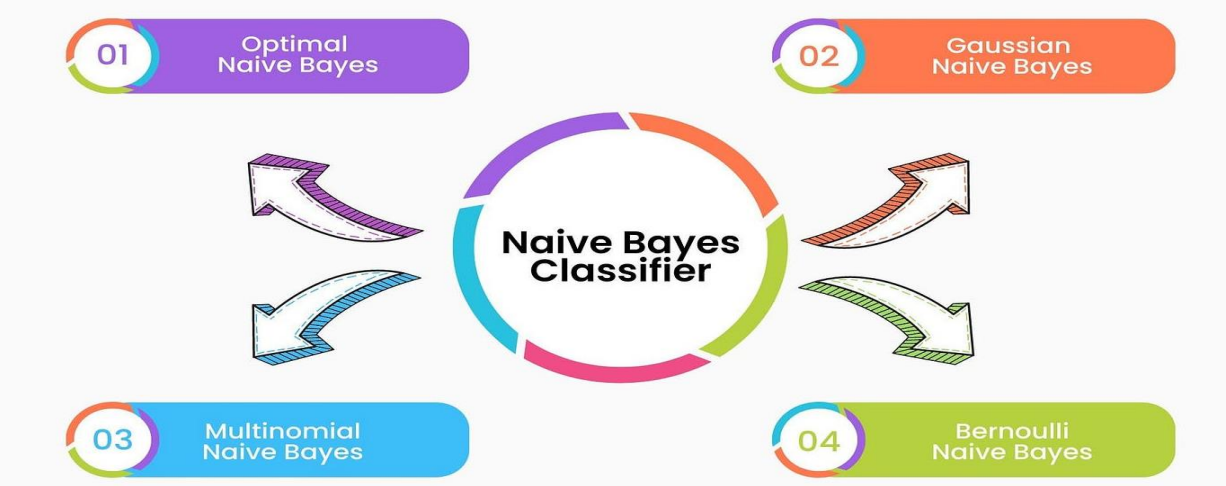
A, B = events

$P(A|B)$ = probability of A given B is true

$P(B|A)$ = probability of B given A is true

$P(A), P(B)$ = the independent probabilities of A and B

Types of Naive Bayes algorithm

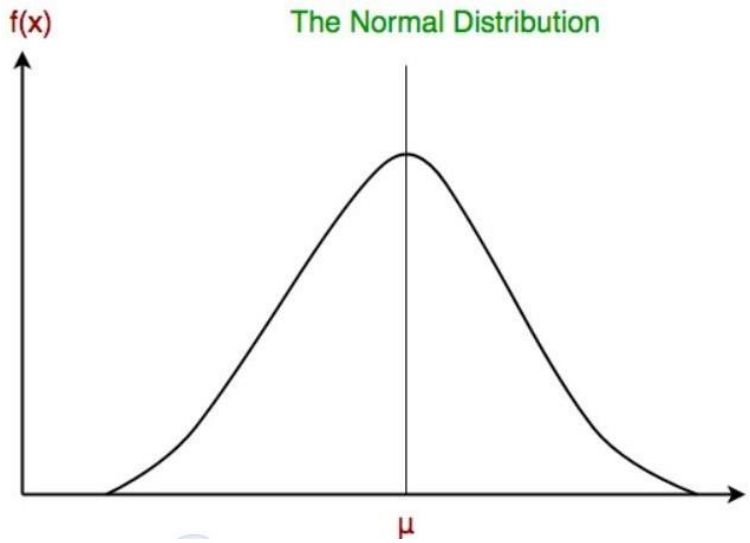


1. Gaussian Naïve Bayes algorithm

- When we have continuous attribute values, we made an assumption that the values associated with each class are distributed according to Gaussian or Normal distribution.
- Machine learning algorithm for classification.
- Probabilistic approach based on Bayes' theorem.
- "Naive" due to strong independence assumption between features.
- Suitable for problems with continuous features.
- Estimate probability distribution of each feature for each class:
- Calculate mean and standard deviation of each feature for each class.
- Represent mean value and variability of the feature within a class.
- Calculate probability of each class label given the observed features.
- For example,Training data with continuous attribute x

- Segmented data by class
  - Mean **μi**
  - variance **σi** of attribute **x** for each class i
  - Observation value **xi**

$$p(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}}$$

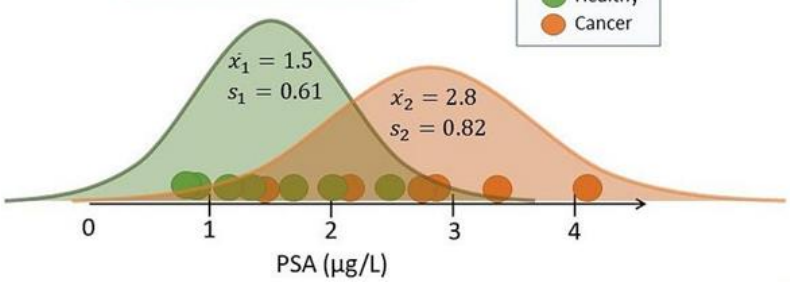


Example:

Gaussian Naive Bayes

Status	PSA
Cancer	4.1
Cancer	3.4
Cancer	2.9
Cancer	2.8
Cancer	2.7
Cancer	2.1
Cancer	1.6
Healthy	2.5
Healthy	2.0
Healthy	1.7
Healthy	1.4

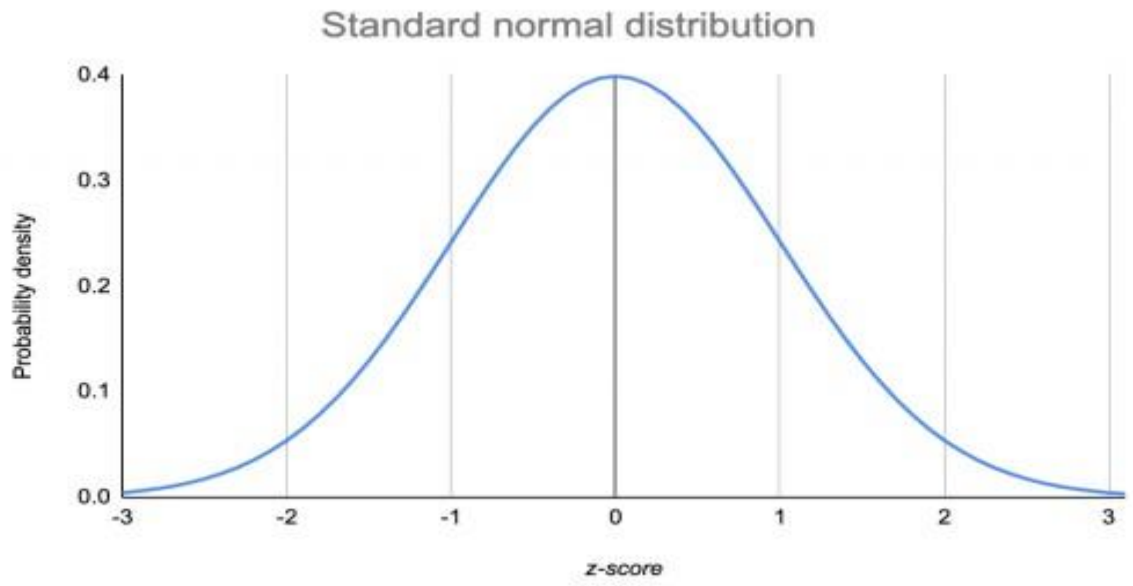
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



2. Multinomial Naïve Bayes algorithm

- With a Multinomial Naïve Bayes model, samples (feature vectors) represent the frequencies with which certain events have been generated by a multinomial (p1, . . .
  ,pn) where pi is the probability that event i occurs.
- Multinomial Naive Bayes is a variant of the Naive Bayes algorithm specifically designed for classification tasks with features that represent counts or frequencies, commonly encountered in text classification tasks.
- Multinomial Naive Bayes calculates the likelihood of observing each feature value in each class.
- It estimates the probabilities of each feature's value given each class based on the training data.
- During prediction, it computes the likelihood of observing the feature values in the new instance given each class.
- Finally, it combines the likelihood with the prior probabilities of classes to compute the posterior probabilities and selects the class with the highest probability as the predicted class.
- Multinomial Naïve Bayes algorithm is preferred to use on data that is multinomiallydistributed.
- It is one of the standard algorithms which is used in text categorization classification.

$$P(X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_k = x_k) = \frac{n!}{\prod_i x_i!} \prod_i {p_i}^{x_i}$$



3. Bernoulli Naive Bayes

- Classification algorithm based on Bayes' theorem.
- Predicts probability of a sample belonging to a specific class.
- Used for binary classification problems (target variable: 0 or 1).
- Independence of features: Each feature independent of all others.
- Bernoulli distribution: Each feature has two possible values (0 or 1) with equalprobability across samples.
- Bernoulli Naive Bayes is another variant of the Naive Bayes algorithm, specifically tailored for binary feature vectors, where features represent presence or absence of certain terms or attributes.
- Bernoulli Naive Bayes calculates the likelihood of observing each feature value (presence or absence) in each class.
- It estimates the probabilities of each feature's value given each class based on the training data.
- During prediction, it computes the likelihood of observing the feature values in the new instance given each class.

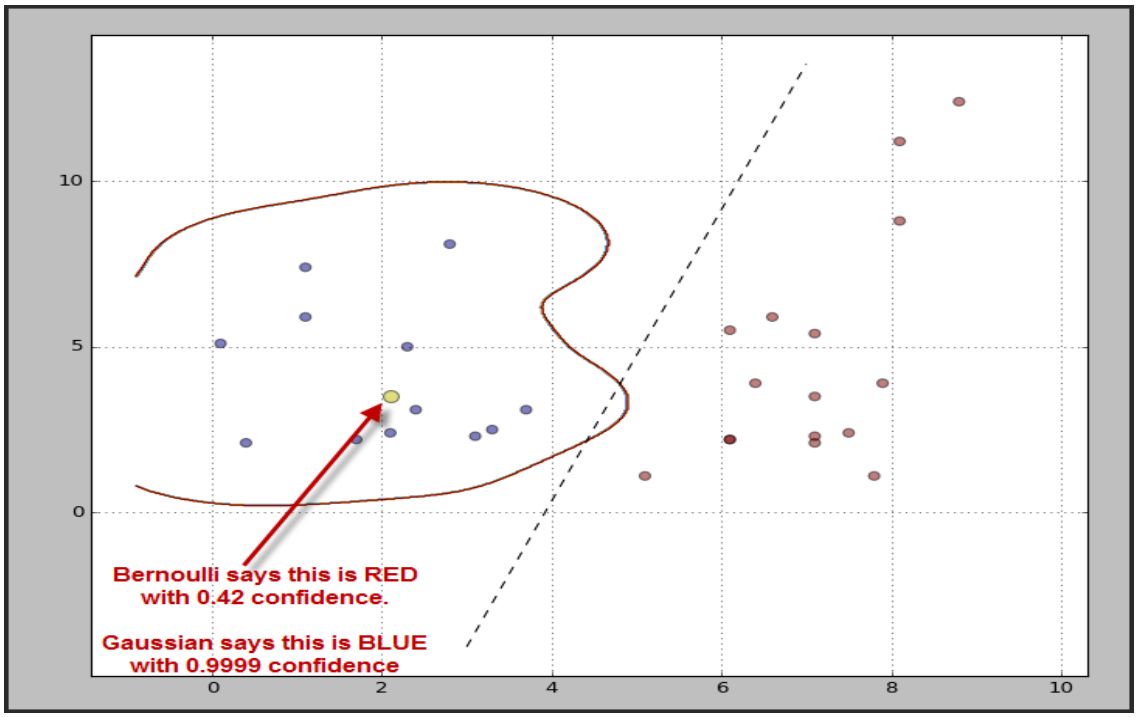
- Finally, it combines the likelihood with the prior probabilities of classes to compute the posterior probabilities and selects the class with the highest probability as the predicted class.

The mathematical formula for Bernoulli Naive Bayes is as follows:

P(y|x) = P(y) * ∏ P(xi|y)

Where:

- P(y|x) is the posterior probability of class y given the features x
- P(y) is the prior probability of class y
- P(xi|y) is the likelihood of feature i given class y
- ∏ is the product symbol



Advantages of Naive Bayes Classifier

- ✓ Naive Bayes is one of the most fast-moving and effortless Machine Learningalgorithms to predict the class of a dataset.
- ✓ It can manage to perform Binary as well as Multi-class Classifications.
- ✓ It is best suited for multi-class predictions as compared to any otheralgorithm.
- ✓ The most popular practical implementation of this classifier is for textclassification problems.

Disadvantages of Naive Bayes Classifier

- ✓ The data may not always be independent of each other.
- ✓ This algorithm can not be used for an imbalanced dataset.
- ✓ When we encounter a response vector in the test data for a particular class absent in the training data, we might end up with zero class probabilities; this is known as the Zero probability problem.

Applications of Naive Bayes Classifier

- o Spam filtration
- o Text classification
- o Sentiment analysis
- o Recommendation System
- o Multi-class prediction