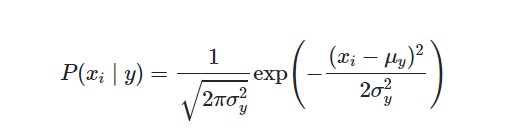
1. **Write the difference between the following:**

i. Gaussian Naive Bayes:

This type of Naive Bayes is built on the assumption of a normal distribution of probabilities. A Gaussian distribution is also called Normal distribution. When plotted, it gives a bell-shaped curve which is symmetric about the mean of the feature values.

When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed to be-



Sometimes assume variance

* is independent of Y (i.e., σi),
* or independent of Xi (i.e., σk)
* or both (i.e., σ)

Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution.

An approach to create a simple model is to assume that the data is described by a Gaussian distribution with no co-variance (independent dimensions) between dimensions. This model can be fit by simply finding the mean and standard deviation of the points within each label, which is all what is needed to define such a distribution.

ii. Multinomial Naive Bayes:

Multinomial classification suits best for the discrete values like word counts. It ignores the non-occurrence of the features. So, if you have frequency 0 then the probability of occurrence of that feature will be 0 hence multinomial naive Bayes ignores that feature.

iii. Complement Naive Bayes:

This approach is almost the same as the Multinomial, though now we count the occurrences of a word in the complement to the class. Multinomial Naive Bayes is not able to do very well with unstable data.

iv. Bernoulli Naive Bayes:

Bernoulli formula is close to the multinomial one, this is used when features are binary. So, instead of using the frequency of the word, if you have discrete features in 1s and 0s that represent the presence or absence of a feature. In that case, the features will be binary and we will use Bernoulli Naive Bayes.

v. Categorical Naive Bayes:

The categorical Naive Bayes classifier is suitable for classification with discrete features that are categorically distributed. The categories of each feature are drawn from a categorical distribution.

vi. Out-of-core naive Bayes model fitting:

Various naïve Bayes models can be used to tackle large scale classification problems for which the full training set might not fit in memory. So for this problem most of the types expose a partial fit method that can be used incrementally as done with other classifiers as demonstrated in Out-of-core classification of text documents.

1. **What is Jaccard and Cosine Similarity?**

Cosine similarity measures the similarity between two vectors by taking the cosine of the angle the two vectors make in their dot product space. If the angle is zero, their similarity is one, the larger the angle is, the smaller their similarity. The measure is independent of vector length (the two vectors can even be of different length), which makes it a commonly used measure for high-dimensional spaces.

Jaccard similarity measures the similarity between two nominal attributes by taking the intersection of both and divide it by their union.

1. **Cosine similarity** measures the similarity between two vectors by taking the cosine of the angle the two vectors make in their dot product space. If the angle is zero, their similarity is one, the larger the angle is, the smaller their similarity.

**Jaccard similarity** measures the similarity between two nominal attributes by taking the intersection of both and divide it by their union.

1. In calculating the similarity using the **cosine similarity** calculation done for one title with another title.

In calculating the similarity using the **Jaccard similarity** calculation done for one title with another title.

1. Results of cosine similarity has the highest value in comparison with Jaccard similarity and the joint between Cosine and Jaccard similarity.
2. Jaccard similarity is good for cases where duplication does not matter, cosine similarity is good for cases where duplication matters while analysing text similarity.