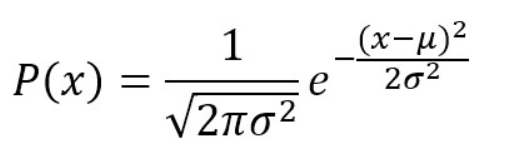
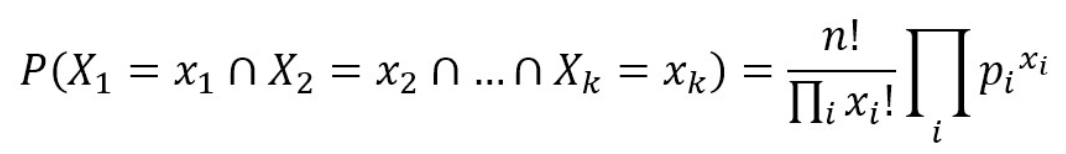
1. **Write the difference between the following:**
2. **Gaussian Naive Bayes**

Gaussian Naive Bayes is useful when working with continuous values which probabilities can be modeled using a Gaussian distribution:



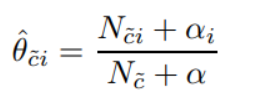
1. **Multinomial Naive Bayes**

A multinomial distribution is useful to model feature vectors where each value represents, for example, the number of occurrences of a term or its relative frequency. If the feature vectors have n elements and each of them can assume k different values with probability pk, then:



1. **Complement Naive Bayes**

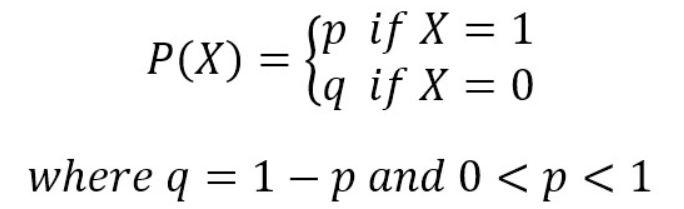
This approach is almost the same as the Multinomial, though now we count the occurrences of a word in the complement to the class. For example, for the spam message we will count the repetitions of each word in all the non-spam messages:



Nc — total number of words in the opposite class (for the spam parameter — number of non-spam words), Nci — repetitions of a word in the opposite class (for a word from spam message — the number of repetitions in all non-spam messages).

1. **Bernoulli Naive Bayes**

If X is random variable Bernoulli-distributed, it can assume only two values (for simplicity, let’s call them 0 and 1) and their probability is:



1. **Categorical Naive Bayes**

Categorical Naive Bayes is suitable for the categorical values — if the example has the set of features or not.All formulas are the same as for the multinomial approach but with the occurrences instead of repetitions.

1. **Out-of-core naive Bayes model fitting**

Naive Bayes models can be used to tackle large scale classification problems for which the full training set might not fit in memory. To handle this case, MultinomialNB, BernoulliNB, and GaussianNB 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. All naive Bayes classifiers support sample weighting.

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

**Jaccard Similarity :**

Jaccard similarity measures the similarity between two nominal attributes by taking the intersection of both and dividing it by their union. In terms of the above definitions this gives



A11 = total number of binary values where both vectors have the value 1.

A01 = total number of binary values where first vector has value 1, other has value 0.

A10 = total number of binary values where the first vector has value 0, other has value 1.

A00 = total number of binary values where both vectors have the value 0.

**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.

