**(B)In a word document,**

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

**I.** Gaussian Naive Bayes,

**II.** Multinomial Naive Bayes,

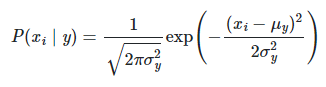
**III.** Complement Naive Bayes,

**IV**. Bernoulli Naive Bayes,

**V.** Categorical Naive Bayes,

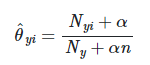
**VI.** Out-of-core naive Bayes model fitting

**I. Gaussian Naive Bayes:** This approach is built on the assumption of a normal distribution of probabilities. It means, that spam and not-spam classes of messages have frequencies of the words from vocabulary distributed by the Gaussian law:

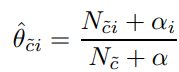


The formula is based on the mean (**μ**) and Bessel corrected variance (**σ**) of the frequency of each word in the class of messages. In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a **Gaussian distribution**. A Gaussian distribution is also called Normal distribution. When plotted, it gives a bell-shaped curve which is symmetric about the mean of the feature values.

**II. Multinomial Naive Bayes**: Feature vectors represent the frequencies with which certain events have been generated by a multinomial distribution. This is the event model typically used for document classification. Multinomial classification suits best for the discrete values like word counts. So, we expect it to show the best accuracy. In this case distribution of probabilities for each event bases on the formula:



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.

****

**IV. Bernoulli Naive Bayes:** Bernoulli formula is close to the multinomial one, though the input is the set of Boolean values (the word is present in the message or not) instead of the set of frequencies.



So, the algorithm explicitly penalizes the non-occurrence of a feature (word in the message is absent in the vocabulary) while the multinomial approach uses the smoothing parameter for the absent values.

**V**. **Categorical Naive Bayes:** Categorical Naive Bayes is suitable for the categorical values — if the example has the set of features or not. In our case, it means, that the vocabulary is treated as the set of features, and the occurrence of a word in the message is treated as the matching with the feature. All formulas are the same as for the multinomial approach but with the occurrences instead of repetitions.

**VI. Out-of-core naive Bayes model fitting:** It is used to handle cases of large-scale classification problems for which the complete training dataset might not fit in the memory.

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

Jaccard Similarity: Jaccard Similarity is a common proximity measurement used to compute the similarity between two objects, such as two text documents. Jaccard similarity can be used to find the similarity between two asymmetric binary vectors or to find the similarity between two sets. In literature, Jaccard similarity, symbolized by J, can also be referred to as Jaccard Index, Jaccard Coefficient, Jaccard Dissimilarity, and Jaccard Distance.

Cosine Similarity: Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.