ASSIGNMENT-03

Hitesh S

21BDA47

B.

a. Write the difference between the following:

i. Gaussian Naive Bayes:

The assumption of a normal probability distribution underpins this type of Naive Bayes. A normal distribution is sometimes known as a Gaussian distribution. It produces a bell-shaped curve that is symmetric about the mean of the feature values when plotted.

ii. Multinomial Naive Bayes:

For discrete data such as word counts, multinomial classification is ideal. The absence of the features is ignored. As a result, if the frequency of a feature is zero, the likelihood of it occurring is zero, hence multinomial naive Bayes ignores it.

iii. Complement Naive Bayes:

This method is nearly identical to the Multinomial, with the exception that we now count the occurrences of a word in the complement of the class. With unstable data, Multinomial Naive Bayes do not perform well.

iv. Bernoulli Naive Bayes:

When characteristics are binary, the Bernoulli formula is similar to the multinomial one. If you have discrete features in 1s and 0s that signify the existence or absence of a feature, you can use them instead of the frequency of the word. In that instance, the features will be binary, and Bernoulli Naive Bayes will be used.

v. Categorical Naive Bayes:

The categorical Naive Bayes classifier is well suited to classification with categorically distributed discrete features. Each feature's categories are chosen from a categorical distribution.

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

Various nave Bayes methods can be used to solve large-scale classification issues where the entire training data is too vast to fit in memory. As a result, most types supply a partial fit technique for this problem, which may be used progressively like other classifiers, as seen in Out-of-core classification of text documents.

b. What is Jaccard and Cosine Similarity?

A. The cosine of the angle formed by two vectors in their dot product space is used to calculate cosine similarity between them. Their similarity is one if the angle is zero; the larger the angle, the less similar they are.By dividing the intersection of both nominal qualities by their union, Jaccard similarity determines how similar two nominal attributes are.

B. When calculating the similarity between two titles, the cosine similarity technique is used. When calculating the similarity between two titles, the Jaccard similarity technique is used.

C. In comparison to Jaccard similarity and the joint of Cosine and Jaccard similarity, the results of cosine similarity have the highest value.

D. When analyzing text similarity, Jaccard similarity is good for circumstances when duplication is unimportant, whereas cosine similarity is good for cases where duplication is important.