ML-Assignment 03

Sarvesh S

21BDA05

**1.a.**

**1.Gaussian Naive Bayes**

This type of Naive Bayes is used when variables are continuous in nature. It assumes that all the variables have a normal distribution. Because of that we can’t use probability while doing math in spite we do likelihood here

**2.Multinomial Naive Bayes:**

This basically works on frequency. Here we are going to collect each unique words. And we can make histogram from this. And for text classification multinomial is more used

**3.Complement Naive Bayes:**

It is generally more efficient on imbalance data (i.e if we have very few spam data). In this technique we calculate probabilities of an item belonging to all other class( i.e opposite of doing in naïve bayes)

Why we use this means some class has very few occurance of a particular word and while using the complement naïve bayes we came to know that all other class has low frequency of this prominent word or not

**4.Bernoli Naïve Bayes:**

It is like a bernoli distribution, the distribution has only two outcome i.e success or failure same like here features has two outcome 0 and 1 whether the word is present or not in the document .here the predictor are Boolean value

**5. Categorical Naive Bayes:**

It is suitable for classification with discrete features which assumes categorically distribution for each feature

**6. Out-of-core naive Bayes model fitting:**

This classifier is used to handle cases of large scale classification problems for which the complete training dataset might not fit in the memory**.**

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

Jaccard similarity takes only unique set of words for each sentence document while cosine similarity takes total length of the vectors