ASSIGNMENT 3

NAME: Yana Kaveramma A A

REGISTER NO: 21BDA32

1. A) Naïve Bayes is based on Bayes’ Theorem and it assumes that all the variables are independent i.e presence of a particular feature in a class is unrelated to the presence of any other.

There are multiple Naïve Bayes:

1. Gaussian Naïve Bayes: 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. Since only mean and standard deviation is used, It is easy to work with.

<https://www.youtube.com/watch?v=H3EjCKtlVog>

1. Multinomial Naive Bayes: Multinomial Naive Bayes classifier is a specific instance of a Naive Bayes classifier which uses a multinomial distribution for each of the features. The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification).

<https://stats.stackexchange.com/questions/33185/difference-between-naive-bayes-multinomial-naive-bayes>

1. Complement Naive Bayes: In complement Naive Bayes, instead of calculating the probability of an item belonging to a certain class, we calculate the probability of the item belonging to all the classes. This is the literal meaning of the word, complement and hence is called Complement Naive Bayes. It overcomes the problem of class imbalance.

<https://www.youtube.com/watch?v=Rhs3RIECfe4>

1. Bernoulli Naive Bayes: BernoulliNB implements the naive Bayes training and classification algorithms for data that is distributed according to multivariate Bernoulli distributions; i.e., there may be multiple features but each one is assumed to be a binary-valued (Bernoulli, boolean) variable.

<https://www.youtube.com/watch?v=ZxbF0qDe-pw>

1. 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. It is an extension of Bernoulli distribution i.e  discrete variable with more than two possible outcomes.

<https://blog.ineuron.ai/Categorical-Naive-Bayes-Classifier-implementation-in-Python-dAVqLWkf7E>

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 a partial\_fit method that can be used incrementally as done with other classifiers as demonstrated in [Out-of-core classification of text documents](https://scikit-learn.org/stable/auto_examples/applications/plot_out_of_core_classification.html#sphx-glr-auto-examples-applications-plot-out-of-core-classification-py).

2. C) 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.