Assignment – 03

21BDA35 – Nakul Ramesh Varma

Question 1

A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Write the difference between the following:

1) **Gaussian Naive Bayes**

When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a gaussian distribution.

2) **Multinomial Naive Bayes**

This is mostly used for document classification problem, i.e., whether a document belongs to the category of sports, politics, technology etc. The features/predictors used by the classifier are the frequency of the words present in the document.

3) **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.

4) **Bernoulli Naive Bayes**

This is similar to the multinomial naive bayes but the predictors are Boolean variables. The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

5) **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.

6) **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

Question 1 Reference:

Towards DataScience

Scikit Learn

Question 2

What is Jaccard and Cosine Similarity?

Jaccard Similarity or intersection over union is defined as size of intersection divided by size of union of two sets.

Cosine Similarity calculates similarity by measuring the cosine of angle between two vectors

Differences between Jaccard Similarity and Cosine Similarity:

1. Jaccard similarity takes only unique set of words for each sentence / document while cosine similarity takes total length of the vectors. (These vectors could be made from bag of words term frequency or tf-idf)
2. This means that if you repeat the word “friend” in Sentence 1 several times, cosine similarity changes but Jaccard similarity does not. For ex, if the word “friend” is repeated in the first sentence 50 times, cosine similarity drops to 0.4 but Jaccard similarity remains at 0.5.
3. Jaccard similarity is good for cases where duplication does not matter, cosine similarity is good for cases where duplication matters while analyzing text similarity. For two product descriptions, it will be better to use Jaccard similarity as repetition of a word does not reduce their similarity.

Question 2 Reference : Towards DataScience