Assignment 03

Soham Giri

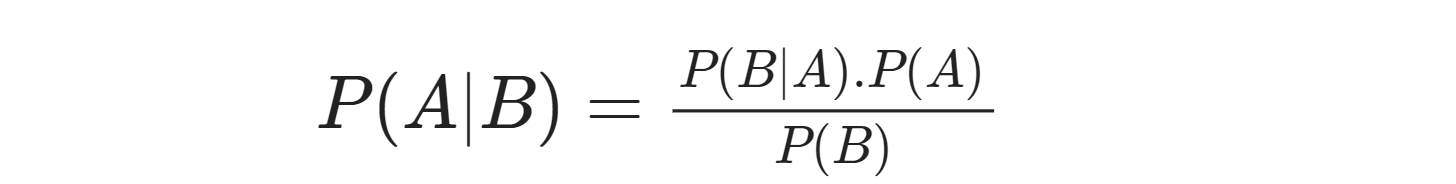
21BDA26

**Write the difference between the following:**

1. **Gaussian Naive Bayes**

Naive Bayes is a basic but effective probabilistic classification model in machine learning that draws influence from Bayes Theorem.

Bayes theorem is a formula that offers a conditional probability of an event A taking happening given another event B has previously happened. Its mathematical formula is as follows: –



Where

* A and B are two events
* P(A|B) is the probability of event A provided event B has already happened.
* P(B|A) is the probability of event B provided event A has already happened.
* P(A) is the independent probability of A
* P(B) is the independent probability of B

<https://www.analyticsvidhya.com/blog/2021/11/implementation-of-gaussian-naive-bayes-in-python-sklearn/>

1. **Multinomial Naive Bayes**

The Multinomial Naive Bayes algorithm is a Bayesian learning approach popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as an email or a newspaper story, using the Bayes theorem. It calculates each tag's likelihood for a given sample and outputs the tag with the greatest chance.

<https://www.upgrad.com/blog/multinomial-naive-bayes-explained/#:~:text=The%20Multinomial%20Naive%20Bayes%20algorithm%20is%20a%20Bayesian%20learning%20approach,tag%20with%20the%20greatest%20chance>.

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.

The Complement Naive Bayes classifier was designed to correct the “severe assumptions” made by the standard Multinomial Naive Bayes classifier. It is particularly suited for imbalanced data sets. New in version 0.20. Additive (Laplace/Lidstone) smoothing parameter (0 for no smoothing).

<https://it-qa.com/what-is-complement-naive-bayes/>

1. **Bernoulli Naive Bayes**

Bernoulli Naive Bayes is one of the variants of the Naive Bayes algorithm in machine learning. It is very useful to be used when the dataset is in a binary distribution where the output label is present or absent. The main advantage of this algorithm is that it only accepts features in the form of binary values such as:

1. True or False
2. Spam or Ham
3. Yes or No
4. 0 or 1

<https://thecleverprogrammer.com/2021/07/27/bernoulli-naive-bayes-in-machine-learning/>

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.

<https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.CategoricalNB.html#:~:text=The%20categorical%20Naive%20Bayes%20classifier,drawn%20from%20a%20categorical%20distribution>.

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

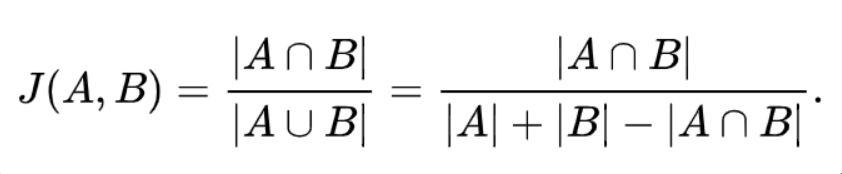
Naive Bayes models can be used to tackle large scale classification problems for which the full training set might not fit in memory. To handle this, MultinomialNB, BernoulliNB and GaussianNB expose a partial\_fit method that can be used incrementally as done with other classifiers. All naive bayes classifiers support sample weighting.

[1.9. Naive Bayes — scikit-learn 1.0.2 documentation](https://scikit-learn.org/stable/modules/naive_bayes.html)

**What is Jaccard and Cosine Similarity?**

It is the method of measuring similarity between two sets, which means that when you apply two sets as documents, the size of the intersection between two documents is divided by the size of the union. In other words, the similarity measure for two datasets in the range 0 or 1, the closer you are to 1, the more similar the two populations are.

Formula:



**Cosine Similarity:**

Cosine similarity refers to the similarity that can be obtained by vectorizing a document and using the cosine angle between two vectors. If the two vectors have exactly the same orientation, they will have a value of 1, and if they have an angle of 90 degree, they will have a value of -1 if they have opposite directions of 0 or 180 degrees. If the value is closer to 1, the higher the similarity between the two documents.

Formula:

