CLASS 20 17-06-2021

**QUESTIONS**

👉 What is naive Bayes classifier in machine learning?  
👉 What are the different types of naive Bayes classifier?  
👉 Which type of naive Bayes classifier is best suited for document classification problem?  
👉 How do you use naive Bayes?  
👉 Why do we use naive Bayes classifier?  
👉 What is CountVectorizer in machine learning?  
👉 What is Tfidftransformer used for?

**ANSWERS**

1. **Naïve Bayes Classifier** is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

2. There are three **types of Naïve Bayes model** under the scikit-learn library:

1. Gaussian: It is used in classification and it assumes that features follow a normal distribution.
2. Multinomial: It is used for discrete counts.
3. Bernoulli: The binomial model is useful if your feature vectors are binary (i.e. zeros and ones).

3**. 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.

4. **Naïve Bayes** is a kind of classifier which uses the Bayes Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class.

5. **Naïve Bayes** is the most straightforward and fast classification algorithm, which is suitable for a large chunk of data. Naive Bayes classifier is successfully used in various applications such as spam filtering, text classification, sentiment analysis, and recommender systems.

6. The **CountVectorizer** provides a simple way to both tokenize a collection of text documents and build a vocabulary of known words, but also to encode new documents using that vocabulary.

7. With **Tfidftransformer** you will systematically compute word counts using CountVectorizer and then compute the Inverse Document Frequency (IDF) values and only then compute the Tf-idf scores. Under the hood, it computes the word counts, IDF values, and Tf-idf scores all using the same dataset.