1. **Explain One-Hot Encoding**

One Hot Encoding is a technique used for converting a categorical variable into multiple columns for each class in the categorical variable. It helps with variables with ordinal values.

1. **Explain Bag of Words.**

Bag of Words model is used to preprocess the text by converting it into a *bag of words*, which keeps a count of the total occurrences of most frequently used words.

1. **Explain Bag of N-Grams.**

Bag of N-Grams model is used to preprocess the text by converting it into a *bag of N-Grams*, which keeps a count of the total occurrences of most frequently used N-Grams.

1. **Explain TF-IDF?**

TF-IDF stands for “Term Frequency — Inverse Document Frequency”. This is a technique to quantify words in a set of documents. We generally compute a score for each word to signify its importance in the document and corpus. This method is a widely used technique in Information Retrieval and Text Mining.

1. **What is the OOV problem?**

Out-of-vocabulary (OOV) are terms that are not part of the normal lexicon found in a natural language processing environment.

1. **What are word embeddings?**

It is an approach for representing words and documents. Word Embedding or Word Vector is a numeric vector input that represents a word in a lower-dimensional space. It allows words with similar meaning to have a similar representation. They can also approximate meaning. A word vector with 50 values can represent 50 unique features.

1. **Explain Continuous bag of words (CBOW).**

The CBOW model architecture tries to predict the current target word (the center word) based on the source context words (surrounding words).

1. **Explain SkipGram.**

Skip-gram is one of the unsupervised learning techniques used to find the most related words for a given word. Skip-gram is used to predict the context word for a given target word.

1. **Explain Glove Embeddings.**

GloVe stands for global vectors for word representation. It is an unsupervised learning algorithm developed by Stanford for generating word embeddings by aggregating global word-word co-occurrence matrices from a corpus. The resulting embeddings show interesting linear substructures of the word in vector space.