1. Explain One-Hot Encoding

One-hot encoding is a technique of representing categorical variables as binary vectors. It is usually done as a preprocessing technique for data before input into a machine learning model.

1. Explain Bag of Words

Bag of Words is a representation of text as the number of occurrences of words in a document.

1. Explain Bag of N-Grams

Bag of N-Grams describe the number of occurrences of each N-Gram in each document in a collection.

1. Explain TF-IDF

Also known as Term Frequency – Inverse Document Frequency. It a statistic that measures a relevant a word is to a document in a collection of documents. It works by increasingly proportionally the number of times a word appear in a document offset by the appearance of the word in other documents in the same collection.

1. What is OOV problem?

Out of Vocabulary words are words that rarely appear in documents or are not contained in the pre-trained word representation file that you are using. If they appear frequently enough in a corpus it could throw off the accuracy of the machine learning model.

1. What are word embeddings?

In NLP, word embeddings are typically representations of words for text analysis. Word embeddings are usually in the form of vectors where similar words are represented with vectors which are close together while words with different meanings have vector representations that are far apart.

1. Explain Continuous bag of words (CBOW)

CBOW is a shallow (2-layer) neural network based models that are used to produce word embeddings. It predicts the current word from a window of surrounding context words. The order of the context words does not influence the prediction hence BOW.

1. Explain SkipGram

SkipGram is similar to CBOW but it uses current word to predict surrounding window of context words.

1. Explain Glove Embeddings.

Glove embeddings is model that makes use of word co-occurences in a corpus to derive word embeddings for words.