1. Explain One-Hot Encoding

**One-Hot Encoding is a binary representation of categorical variables (words or tokens in NLP). Each unique word in a vocabulary is assigned a unique binary index or "one-hot" vector. This vector has a length equal to the vocabulary size, with all values set to zero except for the index corresponding to the word, which is set to one. One-Hot Encoding is simple but results in high-dimensional, sparse vectors.**

1. Explain Bag of Words

**The Bag of Words model is a simple text representation technique in NLP. It involves counting the frequency of each word in a document and representing the document as a vector where each element corresponds to a word in the vocabulary, and the value is the word's frequency in the document. The order of words is ignored in BoW, making it computationally efficient but losing sequence information.**

1. Explain Bag of N-Grams

**Bag of N-Grams extends the BoW model by considering sequences of words (N-grams) instead of individual words. For example, in the case of 2-grams (bigrams), the model considers pairs of consecutive words in the text. This approach captures some degree of local word order information.**

1. Explain TF-IDF

**TF-IDF is a numerical statistic used in NLP to evaluate the importance of a word in a document relative to a collection of documents (corpus). It combines two components: Term Frequency (TF), which measures how often a word appears in a document, and Inverse Document Frequency (IDF), which measures how unique or rare a word is across the corpus. TF-IDF helps identify important words while reducing the impact of common words.**

1. What is OOV problem?

**The OOV problem occurs when a word that is not present in the vocabulary of a language model or dataset is encountered during text processing. Handling OOV words is crucial in NLP tasks to prevent errors. Techniques like subword tokenization and character-level embeddings can help address the OOV problem.**

1. What are word embeddings?

**Word embeddings are dense vector representations of words in a continuous vector space. These representations capture semantic relationships between words. Popular word embedding techniques include Word2Vec, GloVe, and FastText. Word embeddings are learned from large text corpora and can be used in various NLP tasks.**

1. Explain Continuous bag of words (CBOW)

**CBOW is a type of word embedding model in which the goal is to predict a target word based on its context (surrounding words). CBOW takes a context window of words as input and learns to predict the center word. It is useful for creating word embeddings that capture word semantics.**

1. Explain SkipGram

**SkipGram is another word embedding model that works in the opposite way of CBOW. It takes a target word as input and learns to predict the words in its context window. SkipGram is known for capturing more detailed word relationships and is often preferred when the focus is on rare words.**

1. Explain Glove Embeddings.

**GloVe is an unsupervised learning algorithm for generating word embeddings. It combines the global statistics of word co-occurrence with matrix factorization techniques to create word vectors that capture both local and global word relationships. GloVe embeddings are pre-trained on large text corpora and can be used for various NLP tasks.**