1. Explain One-Hot Encoding
2. Explain Bag of Words
3. Explain Bag of N-Grams
4. Explain TF-IDF
5. What is OOV problem?
6. What are word embeddings?
7. Explain Continuous bag of words (CBOW)
8. Explain SkipGram
9. Explain Glove Embeddings.

Answer:

1. One-Hot Encoding: One-Hot Encoding is a technique to represent categorical variables as numerical data. It encodes each category as a binary vector where all values are 0 except for the category's corresponding index, which is set to 1. This technique is widely used in machine learning, especially in natural language processing, where it is used to encode words, characters, or parts of speech as numerical data.
2. Bag of Words: Bag of Words is a text representation model that converts a document into a bag of its constituent words, discarding the order and context of the words. It counts the frequency of each word in a document and represents it as a sparse vector. This model is commonly used in natural language processing applications like text classification, sentiment analysis, and document clustering.
3. Bag of N-Grams: Bag of N-Grams is an extension of the Bag of Words model, which captures the context of the words in a document by grouping adjacent words into n-grams. Instead of treating individual words as features, it treats n-grams as features. This model captures more information about the structure of the language, but it requires more memory and processing power than the Bag of Words model.
4. TF-IDF: TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure that evaluates the importance of a word in a document collection. It combines two measures: the term frequency (TF), which measures the frequency of a word in a document, and the inverse document frequency (IDF), which measures how much information a word provides across the entire collection of documents. This technique is commonly used in information retrieval, text mining, and content-based recommendation systems.
5. OOV Problem: OOV (Out-of-Vocabulary) problem refers to the challenge of dealing with words that are not present in the vocabulary of a machine learning model. This can happen when a model encounters new words or rare words that are not seen during the training phase. The OOV problem can affect the performance of the model, especially in natural language processing applications, where the vocabulary is large and dynamic.
6. Word Embeddings: Word Embeddings are a family of techniques that represent words as dense vectors in a high-dimensional space, where the distance between the vectors reflects the semantic similarity between the words. Word embeddings capture the meaning and context of the words, making them more suitable for machine learning tasks like text classification, sentiment analysis, and machine translation.
7. Continuous Bag of Words (CBOW): Continuous Bag of Words (CBOW) is a neural network architecture that predicts a target word from its surrounding context words. It takes a sequence of context words as input and predicts the probability distribution of the target word. CBOW is a fast and efficient way to learn word embeddings, especially when the training data is large.
8. SkipGram: SkipGram is a neural network architecture that predicts the surrounding context words from a target word. It takes a target word as input and predicts the probability distribution of the context words. SkipGram is a slower but more accurate way to learn word embeddings than CBOW, especially when the training data is small.
9. Glove Embeddings: Glove (Global Vectors for Word Representation) Embeddings is a word embedding technique that combines the advantages of both CBOW and SkipGram models. It learns word embeddings by factorizing a matrix of word co-occurrence counts, which captures the global context of the words. Glove embeddings are widely used in natural language processing applications, where they have shown to outperform other word embedding techniques in various tasks.