1.Explain one hot encoding.

Ans : One hot encoding is one method of converting data to prepare it for an algorithm and get a better prediction. With one-hot, we convert each categorical value into a new categorical column and assign a binary value of 1 or 0 to those columns. Each integer value is represented as a binary vector.

2.Explain the bag of words.

Ans : The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.

3.Explain the bag of N grams.

Ans : N-grams are continuous sequences of words or symbols or tokens in a document.It's a probabilistic model that's trained on a corpus of text. Such a model is useful in many NLP applications including speech recognition, machine translation and predictive text input.

4.Explain TF-IDF

Ans : TF-IDF stands for term frequency-inverse document frequency and it is a measure, used in the fields of information retrieval (IR) and machine learning, that can quantify the importance or relevance of string representations.

which mathematically eliminates naturally occurring words in the English language, and selects words that are more .

5.what is OOV problem.

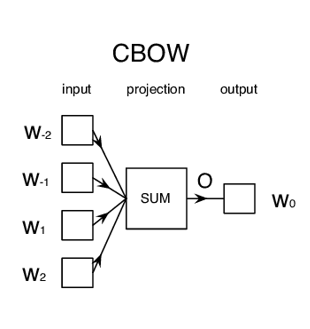
Ans : Out-of-vocabulary (OOV) are terms that are not part of the normal lexicon found in a natural language processing environment. In speech recognition, it's the audio signal that contains these terms. Word vectors are the mathematical equivalent of word meaning.

6.what are the word embeddings ?

Ans :A word embedding is a learned representation for text where words that have the same meaning have a similar representation. It is this approach to representing words and documents that may be considered one of the key breakthroughs of deep learning on challenging natural language processing problems.

7.Explain the continuous bag of words(CBOW)

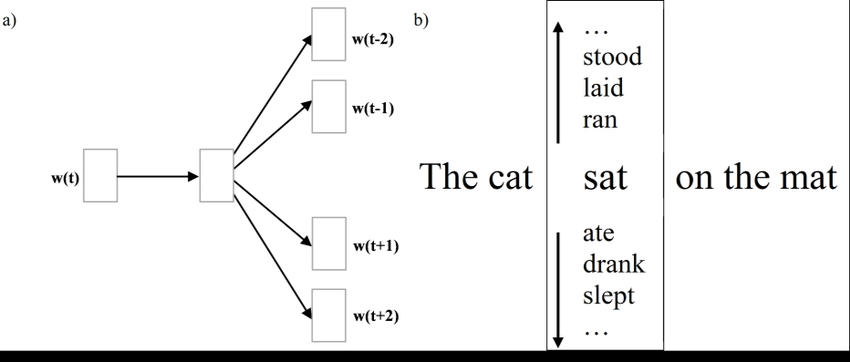
Ans : the distributed representations of context or surrounding words are combined to predict the word in the middle.



The model tries to predict the target word by trying to understand the context of the surrounding words. Consider the same sentence as above, 'It is a pleasant day'. The model converts this sentence into word pairs in the form.

8.Explain skipgram.

Ans : The main idea behind the Skip-Gram model is this it takes every word in a large corpora and also takes one-by-one the words that surround it within a defined 'window' to then feed a neural network that after training will predict the probability for each word to actually appear.



Skip-gram is used to predict the context word for a given target word. It's reverse of CBOW algorithm. Here, target word is input while context words are output.

9.Explain Glove embeddings.

Ans.GloVe stands for Global Vectors for word representation. It is an unsupervised learning algorithm developed by researchers at Stanford University aiming to generate word embeddings by aggregating global word co-occurrence matrices from a given corpus.