**Assignment\_1**

1. Explain One-Hot Encoding

**Ans: In one hot encoding, every word (even symbols) which are part of the given text data are written in the form of vectors, constituting only of 1 and 0 . So one hot vector is a vector whose elements are only 1 and 0.**

1. Explain 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.**

1. Explain Bag of N-Grams

**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.**

1. 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 (words, phrases, lemmas, etc) in a document amongst a collection of documents .**

1. 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.**

1. What are word embeddings?

**Ans: In natural language processing (NLP), word embedding is a term used for the representation of words for text analysis, typically in the form of a real-valued vector that encodes the meaning of the word such that the words that are closer in the vector space are expected to be similar in meaning**.

1. Explain Continuous bag of words (CBOW)

Ans: **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

**Ans: 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. It's reverse of CBOW algorithm. Here, target word is input while context words are output.**

1. Explain Glove Embeddings.

Ans: 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 matrix from a corpus**. The resulting embeddings show interesting linear substructures of the word in vector space