1. One hot encoding can be defined as **the essential process of converting the categorical data variables to be provided to machine and deep learning algorithms which in turn improve predictions as well as classification accuracy of a model**.
2. 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. A bag-of-n -grams model is **a way to represent a document, similar to a [bag-of-words][/terms/bag-of-words/] model**. A bag-of-n -grams model represents a text document as an unordered collection of its n -grams. For example, let's use the following phrase and divide it into bi-grams (n=2 ).
4. TF-IDF which means **Term Frequency and Inverse Document Frequency**, is a scoring measure widely used in information retrieval (IR) or summarization. TF-IDF is intended to reflect how relevant a term is in a given document.
5. 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.
6. 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.
7. The Continuous Bag of Words (CBOW) **The CBOW model architecture tries to predict the current target word (the center word) based on the source context words (surrounding words)**.
8. 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.
9. GloVe is **an unsupervised learning algorithm for obtaining vector representations for words**. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.