Distributional Similarity:

This is the idea that the meaning of a word can be understood from the context in which the word appears for example “NLP Rocks” the literal meaning of the word “rocks” is stones but from the context it’s used to refer to something good.

الفكرة دى مبنية على ان معنى الكلمة يمكن فهمه من سياق الجملة التي ظهرت فيه هذه الكلمة على سبيل المثال صخور معالجة اللغة الطبيعية هذا هو المعنى الحرفي لكن المعنى من السياق انها تشير الى شيء جيد

Distributional hypothesis:

The hypothesize that words that occurs in similar context have similar meaning for example in English words dog and cat occur in similar contexts so the similar representation vectors must also be close to each other.

الافتراضية المبنية على ان الكلمات المكررة في سياقات متشابهة يكون لها معانى متشابهة على سبيل المثال القطط والكلام فكلاهما من الحيوانات ولهم خصائص مشتركة فهم في الاغلب يشتركون في سياقات متشابهة ولذلك يجب ان يكون تمثيل المتجهات لكلاهما قريب

Distributional representation:

Refers to representing word as high dimensional vector based on their distribution in the context in which they appear this achieved using co-occurrence matrix that captures the occurrence of words and context and size of matrix equal the size of vocabulary of the ex: one-hot ,bag of words,n-grams, and tf-idf

يشير الى تمثيل الكلمات في صورة متجهات ذات ابعاد عالية معتمد على توزيعهم في السياق الذي ظهروا فيه ويتم تحقيق ذلك من خلال مصفوفة التكرار وحجم المصفوفة يكون بنفس عدد الكلمات الموجودة ومثال على ذلك...

Distributed Representation:

Is a concept based on the Distributional hypothesis which aims to compress the high dimensional and sparse vectors in distributional representation into low-dimensional and dense this makes them more efficient computationally and easier to learn

هو مبنية على فرضية التوزيع والتي تهدف الى ضغط المتجهات ذات الابعاد الكبيرة ومتفرقة اى تحتوى على اصفار كثيرة وتحويلها الى متجهات لها ابعاد قليلة وكثيفة وهذا يجعلها تكتسب كفاءة وسهله في التعلم

Embedding:

Is mapping between vector space coming from distributional representation to vector space coming from distributed representation.

هي عملية قائمة على التحويل من التمثيل التوزيعي للمتجهات التي كانت ابعادها كبيرة ومتفرقة وتحتوى على العديد من الاصفار الى التمثيل الموزع الذي يحتوى على على متجهات ذات ابعاد صغيرة وكثيفة

Vector semantics:

This refers to the set of NLP methods that aim to learn word representations based on distributional properties of words in large corpus

يقصد بها مجموعة الطرق التي تهدف الى تعلم تمثيل الكلمات بناء على خصائصها التوزيعة في مجموعة النصوص الكبيرة

Word embedding

is a technique used in natural language processing (NLP) to represent words as vectors of numerical values in a high-dimensional space. The goal of word embeddings is to capture the semantic and syntactic relationships between words based on their usage patterns in a large corpus of text. These vectors are learned automatically from the training data using machine learning algorithms such as Word2vec and GloVe.

The key idea behind word embeddings is to capture the distributional similarities between words, which means that words that often appear in similar contexts are likely to have similar meanings. For example, the words "beautiful" and "gorgeous" are likely to have similar meanings because they often appear in similar contexts in text. Word embeddings can capture these subtle differences in meaning and help improve the performance of NLP models on various tasks such as language translation, sentiment analysis, and text classification.

CBOW. In CBOW, the primary task is to build a language model that correctly predicts the center word given the context words in which the center word appears. What is a language model? It is a (statistical) *model* that tries to give a probability distribution over sequences of words. Given a sentence of, say, *m* words, it assigns a probability $Pr(w_1, w_2, \ldots, w_n)$ to the whole sentence. The objective of a language

SkipGram. SkipGram is very similar to CBOW, with some minor changes. In SkipGram, the task is to predict the context words from the center word. For our toy corpus with context size 2, using the center word "jumps," we try to predict every word in context—"brown," "fox," "over," "the"—as shown in Figure 3-10. This constitutes