**LSA(Latent Semantic Analysis)**

LSA is quite effective when working in low-dimensional space for the analysis of incoming documents from the same domain zone as the ones, which are already processed and put into the term-document matrix

Advantages:

Disadvantages:

**Count Vectorization (based on One-hot encoding ):**

Counts occurrence of each word in each document.

The idea is to collect a set of documents (they can be words, sentences, paragraphs or even articles) and count the occurrence of every word in them. Strictly speaking, the columns of the resulting matrix are words and the rows are documents.

Advantages:

Disadvantages:

**TF-IDF transforming:**

The idea behind this approach is term weighting by exploitation of useful statistical measure called tf-idf.

Advantages:

Disadvantages:

**Word2Vec parameter learning:**

Doc2Vec is for paragraphs representation.

*One-word context:*

The intuition behind it is the fact that we’re considering one word per one context (we’re predicting one word given only one word); this approach is often referred to as CBOW model.

*Multi-word context:*

*Skip-gram model:*

situation opposite to CBOW multi-word model: we’d like to predict c context words having one target word on the input.

Advantages:

Disadvantages:

**Glove: Global vectors for word representation**

The approach of global word representation is used to capture the meaning of one word embedding with the structure of the whole observed corpus; word frequency and co-occurence counts are the main measures on which the majority of unsupervised algorithms are based on. GloVe model trains on global co-occurrence counts of words and makes a sufficient use of statistics by minimizing least-squares error and, as result, producing a word vector space with meaningful substructure. Such an outline sufficiently preserves words similarities with vector distance.

To store this information we use co-occurrence matrix X, each entry of which corresponds to the number of times word j occurs in the context of word i.

Advantages:

Disadvantages:

**FastText (Enriching Word Vectors with Subword Information):**

Takes into account morphology of words.

The FastText model takes into account internal structure of words by splitting them into a bag of character n-grams and adding to them a whole word as a final feature

Advantages:

Disadvantages: