

Word embeddings

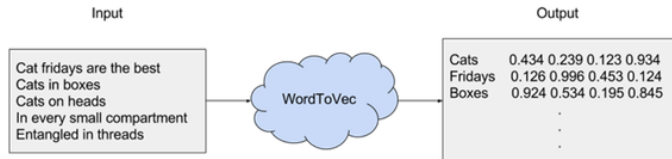
Word2vec Doc2vec

«Высшая школа экономики»

15.10

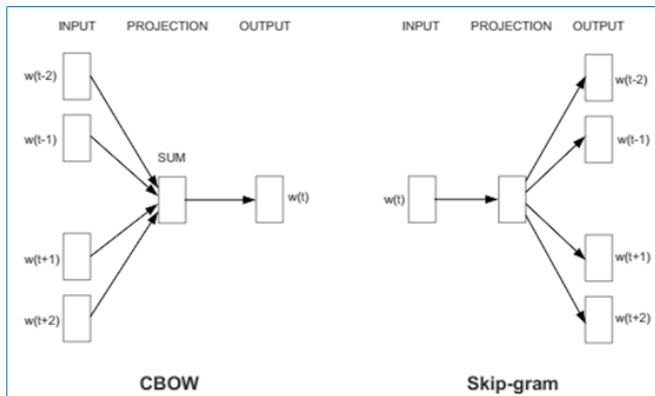
Word2vec

Word2vec is a combination of models used to represent distributed representations of words in a corpus.



Word2vec

- CBOW
- Skipgram



Word2vec

Example:

«The quick brown fox jumps over the lazy dog.»

Word2vec

Source Text

The quick brown fox jumps over the lazy dog. →

The quick brown fox jumps over the lazy dog. →

The quick brown fox jumps over the lazy dog. →

The quick brown fox jumps over the lazy dog. →

Training Samples

(the, quick)
(the, brown)

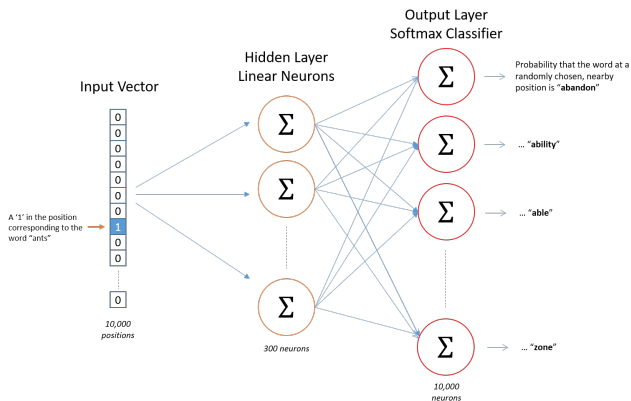
(quick, the)
(quick, brown)
(quick, fox)

(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)

(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

Word2vec

Architecture.



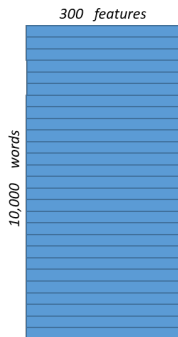
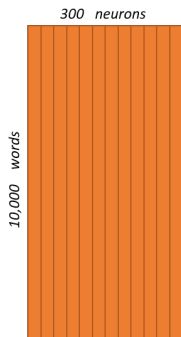
Word2vec

Output.

Hidden Layer
Weight Matrix



*Word Vector
Lookup Table!*



Doc2vec

