Cognate/Professional Electives

Semantics & Sentiment Analysis

Renato R. Maaliw III, *DIT*
College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines

- **Word2vec** is a two-layer neural net that processes text

- Its input is a text corpus and its output is a set of vectors: feature vectors for words in that corpus

- The purpose and usefulness of **Word2vec** is to group the vectors of similar words together in vector space

- That is, it detects similarities mathematically

- **Word2vec** creates vectors that are distributed numerical representations of word feature, features such as the context of individual words

- It does so without human intervention

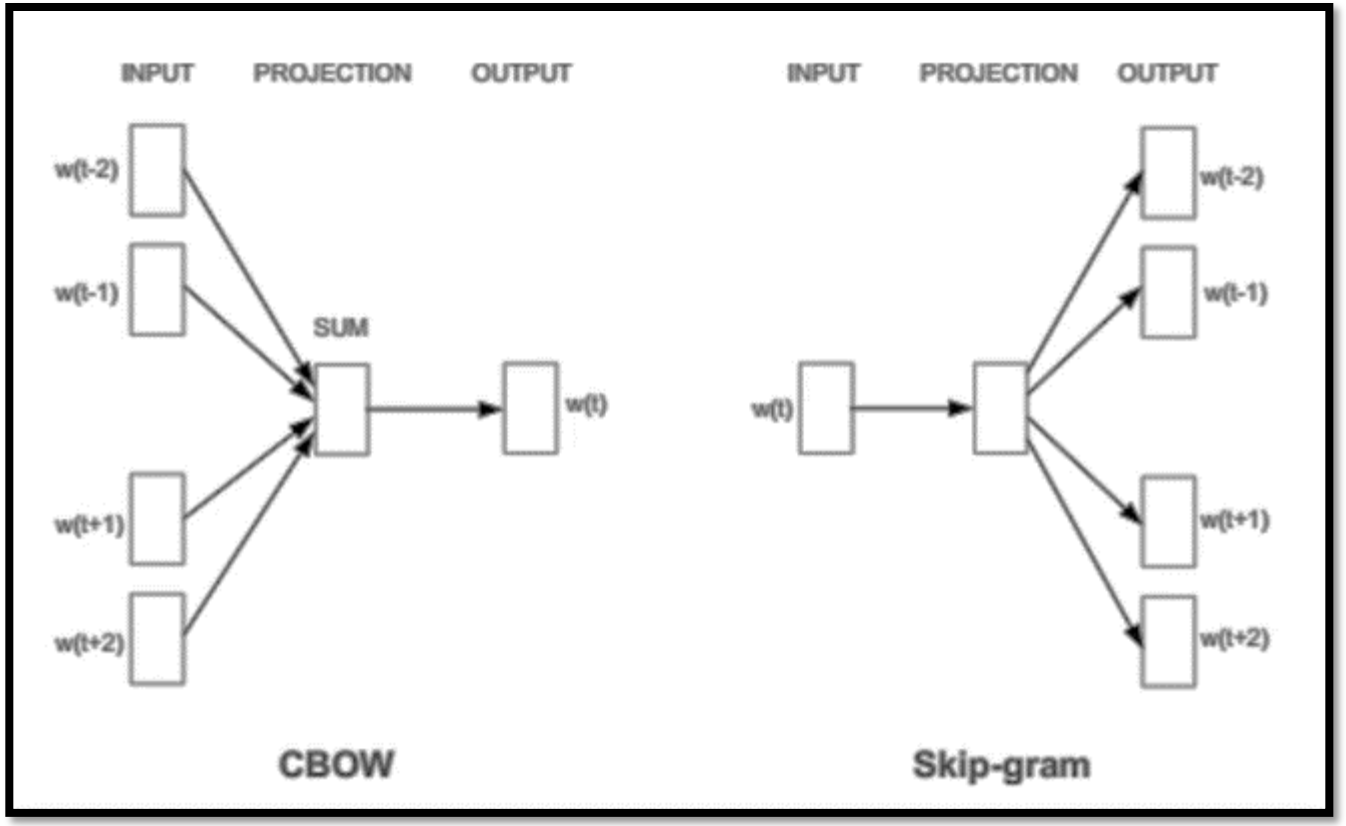
- Given enough data, usage and contexts, Word2vec can make highly accurate guesses about a word's meaning based on past appearances

- Example: *"man" is to "boy"* what *"woman"* is to *"girl"*

- Word2vec trains words against other words that neighbor them in the input corpus

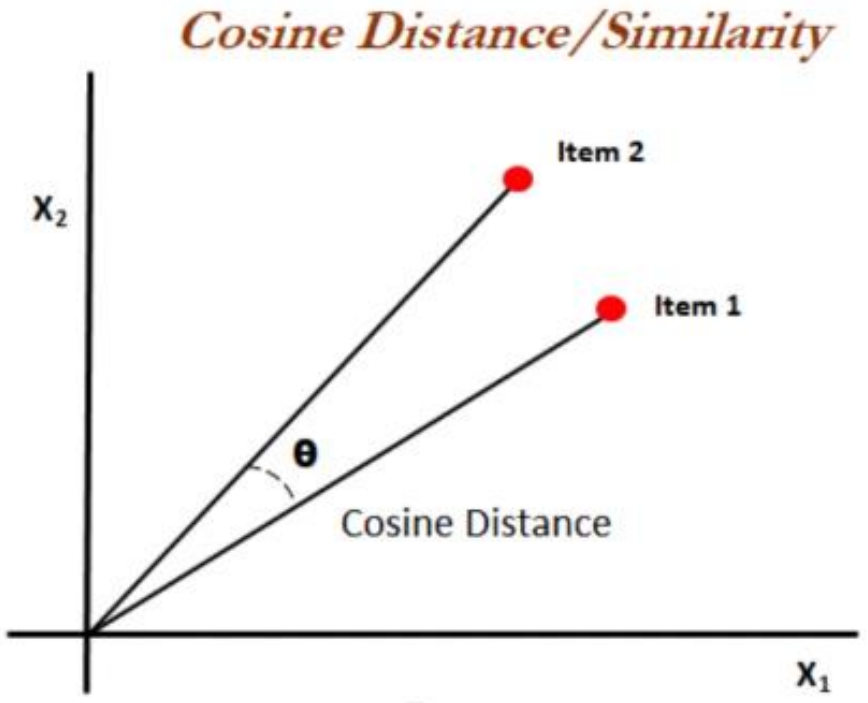
- Either using context to predict a target word (continuous bagging of words - CBOW) or using a word to predict a target context (skip-gram)

Two possible approaches

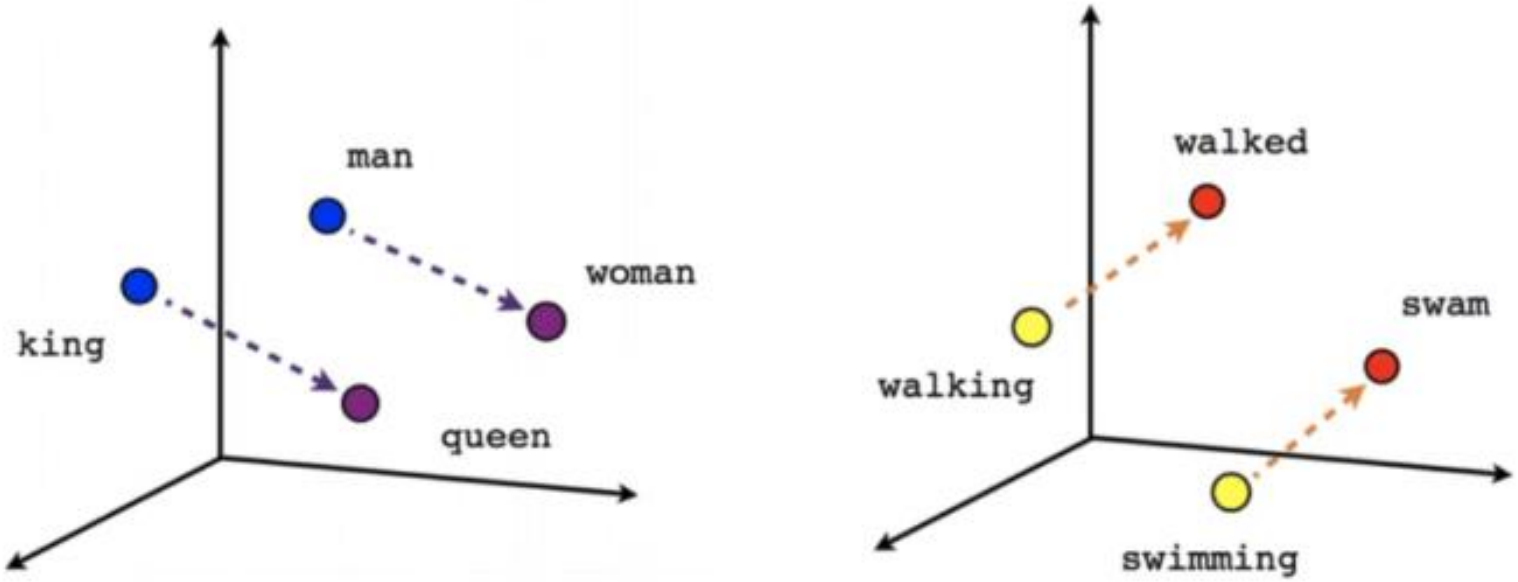


- Each word is now represented by a vector
- In **SpaCy**, vectors has 300 dimensions

- This means we can use Cosine Similarity to measure how similar word vectors are to each other



- Interesting relationships can be established between the word vectors



[CODE DEMO]

Thank you very much for listening.