Word Embedding

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

How to Represent Texts?

- Consider the below sentences
  - Today is a beautiful day.
  - Tomorrow will be a better day.
- How can we represent the word tomorrow and what should the representation indicate?
- Some classic methods
  - Assigning an id to each word
  - One-Hot encoding
  - Co-Occurrence matrix

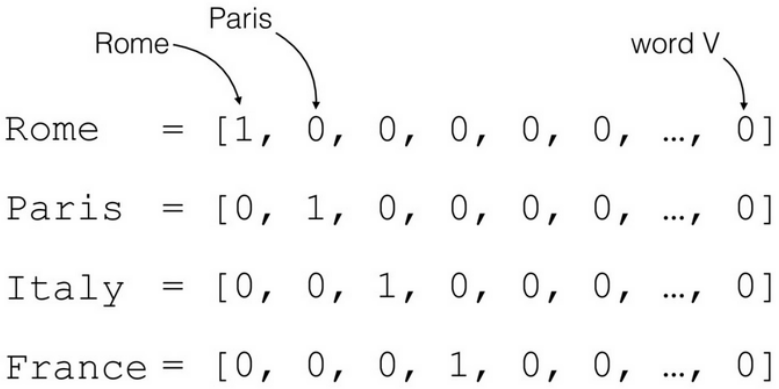

Figure: One-Hot Encoding, Source

Co-Occurrence Matrix

- Example
  - I like deep learning.
  - I like NLP.
  - I enjoy flying.

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

Figure: Co-Occurrence Matrix, Source

Bag of Words (BoW)

- It is an approach used widely in information retrieval.
- BoW is based on counting occurrence of words in each text.
- Each word is represented by the documents it occurs in.
- It is called Bag of Words because it does not consider the order of words.
- But does it convey a meaning?

	the	red	dog	cat	eats	food
1. the red dog →	1	1	1	0	0	0
2. cat eats dog →	0	0	1	1	1	0
3. dog eats food →	0	0	1	0	1	1
4. red cat eats →	0	1	0	1	1	0

Figure: Example of BoW method, Source

Continuous Bag of Words (CBoW)

- To obtain a meaning for each word, a fake task should be defined.
- Consider the following incomplete sentence

  S := I prefer to travel by ... rather than cars.
- By using which one of the words flowers, airplanes, or lions should we fill in the above sentence? The most probable one!

$$\underset{w \in \{\text{flowers, airplanes, lions}\}}{argmax} \quad P(w_i = w | S)$$

- What if a sentence is too long? How should we deal with alternative length of sentences? Use a window . (It comes from an assumption that to guess a word, its neighbourhood should be enough.)

$$\underset{w \in \{\text{flower, airplane, lion}\}}{argmax} \quad P(w_i = w | w_{i-l}, w_{i-l+1}, ..., w_{i-1}, w_{i+1}, ..., w_{i+l-1}, w_{i+l})$$

- How to calculate the probabilities given a corpus ?

Continuous Bag of Words (CBoW)

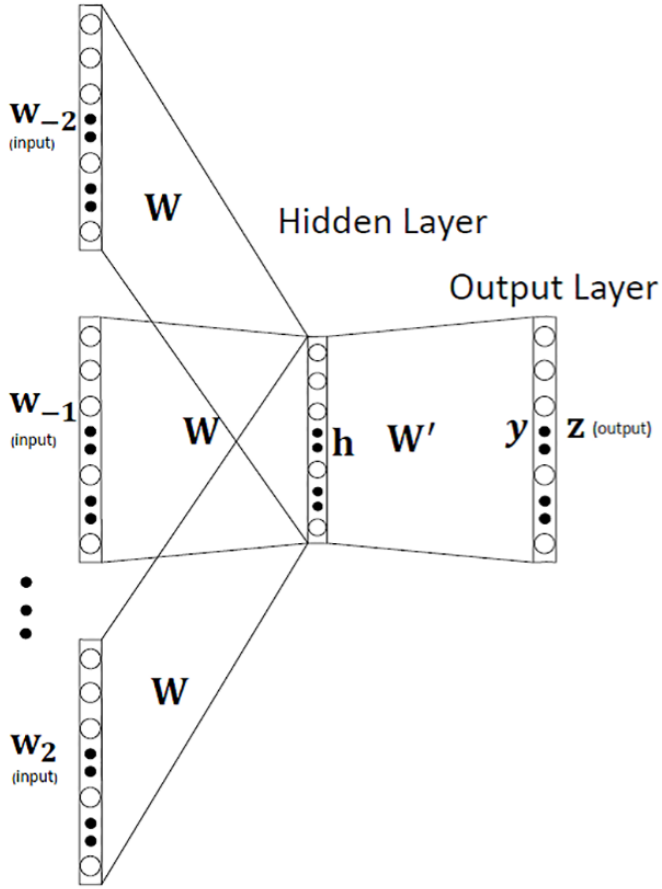

Figure: CBoW model, Source

- Note that matrix W is shared . (The order is not important)

Continuous Bag of Words (CBoW)

- Considering the fake task defined earlier, how can we represent a word in a meaningful way? Hidden layer matrix. (Matrix W)
- By determining the number of the hidden layer neurons (as hyper parameter m), each word is represented by a m-dimension vector.

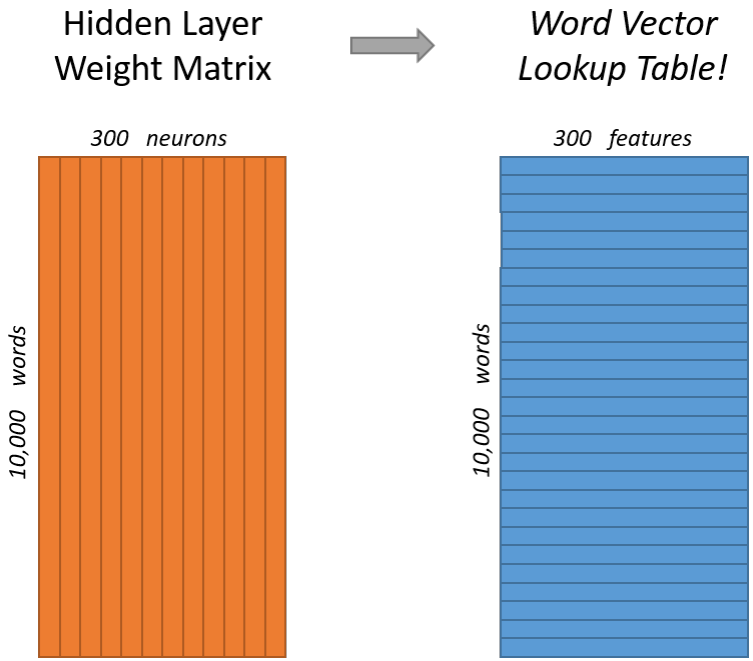

Figure: Obtained Feature Vector, Source

Skip-gram

- It is similar to CBoW and just the fake task is inverted. (Here matrix W is the word embedding matrix)

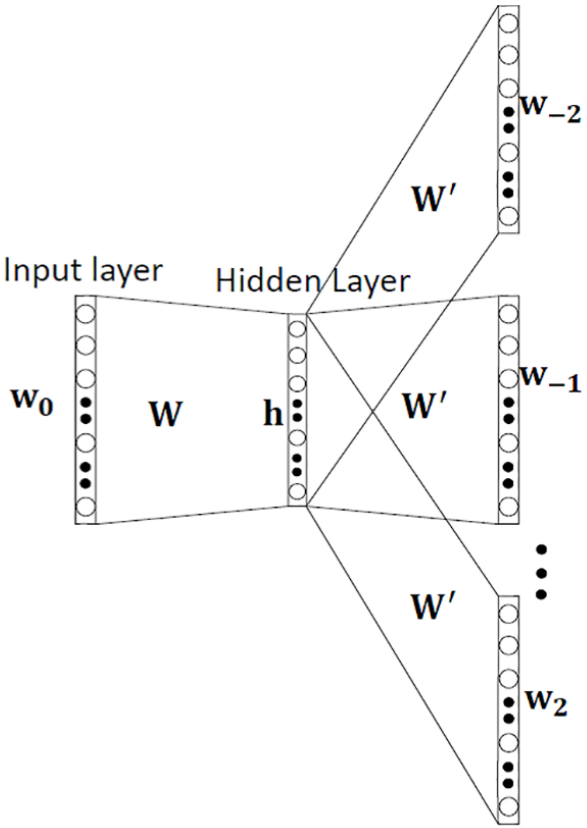

Figure: Skip-gram model, Source

Skip-gram

- **Example** (Source text: The man who passes the sentence should swing the sword.)

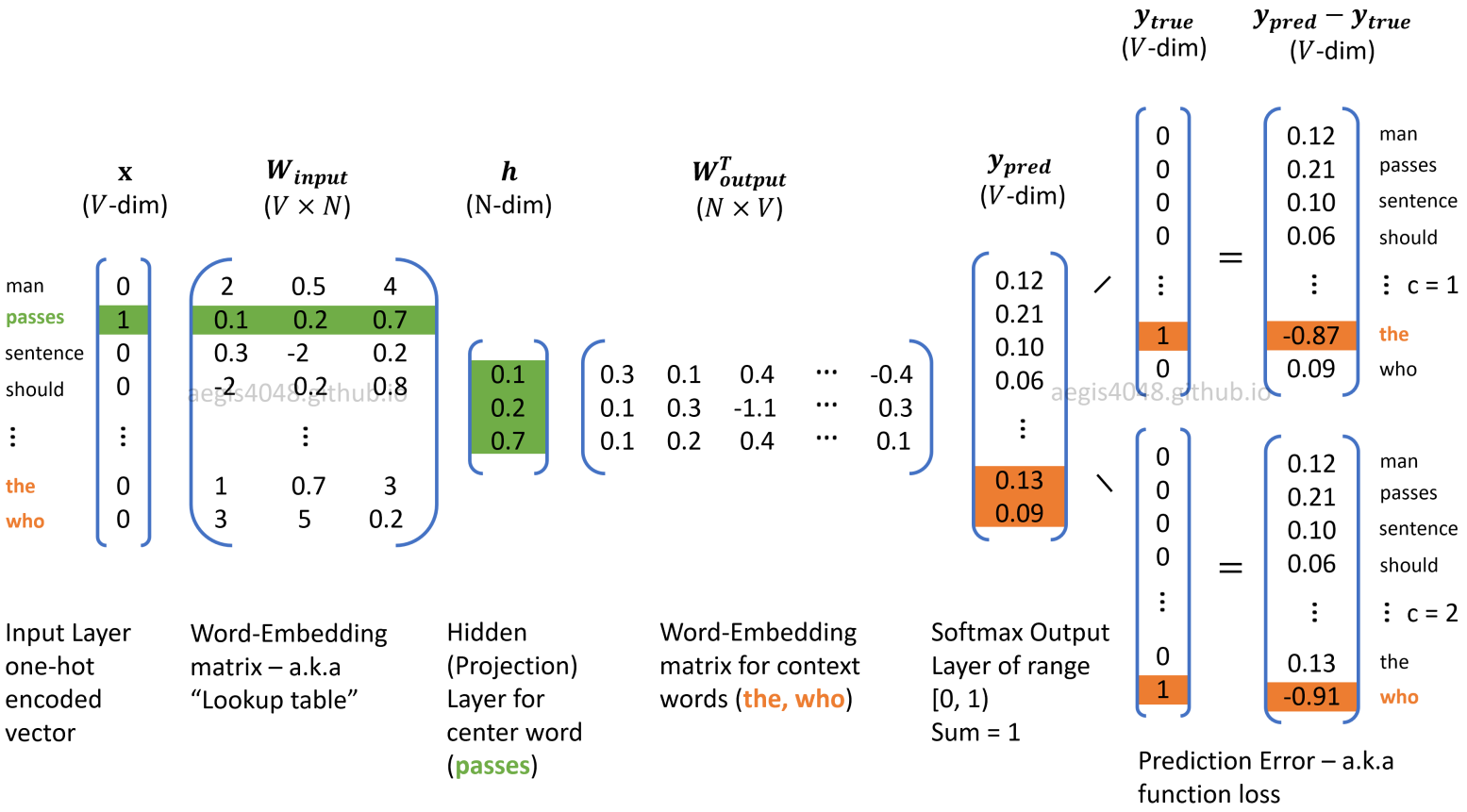

Figure: Skip-gram Method for Window Size 1 Centered Word passes, Source

References

Deep learning for natural language processing lecture 2: Word vectors.
https://cs224d.stanford.edu/lectures/CS224d-Lecture2.pdf.
Accessed: 2022-12-03.

Demystifying neural network in skip-gram language modeling.
https://aegis4048.github.io/demystifying_neural_network_in_skip_gram_language_modeling.
Accessed: 2022-12-03.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean.
Efficient estimation of word representations in vector space, 2013.

Thank You!

Any Question?