Word2vec

Efficient estimation of words representation in vector space

Kazi Wasif Amin (1705079)
Md. Mehedi Hasan (1705082)

Level 3, Term 1
Department of CSE
Bangladesh University of Engineering and Technology

July 13, 2021

Table of Contents

Table of Contents

ML algorithms work on numerical values. To work with words, they should be converted to numbers.

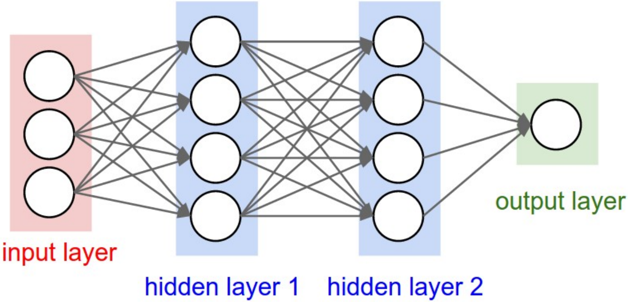


Figure: Neural Network

Table of Contents

One hot encoding

Creates new (binary) columns, indicating the presence of each possible value from the original data.

Word
Apple
Mango
Cow
Cat

Apple	Mango	Cow	Cat
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Figure: One hot encoding

Problem 1

As the number of words increase, the numbers of columns also increases in the table which costs more computation.

Word
Apple
Mango
Cow
Cat
- - -
- - -
Rose

Apple	Mango	Cow	Cat	- - -	- - -	Rose
1	0	0	0	- - -	- - -	0
0	1	0	0	- - -	- - -	0
0	0	1	0	- - -	- - -	0
0	0	0	1	- - -	- - -	0
- - -	- - -	- - -	- - -	- - -	- - -	- - -
- - -	- - -	- - -	- - -	- - -	- - -	- - -
0	0	0	0	- - -	- - -	1

Figure: One hot encoding

One hot encoding

Problem 2

No way to find out the semantic similarities between words.

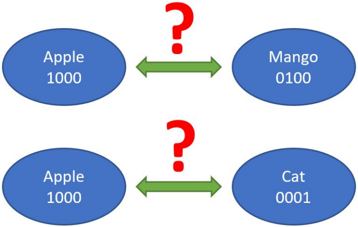


Figure: One hot encoding

Table of Contents

Word2vec motivation

- Relationship among the words are preserved.
- Makes the best of addition of new words in the dictionary.
- Enhanced outcomes in lots of deep learning applications.

Word2vec motivation

- Relationship among the words are preserved.
- Makes the best of addition of new words in the dictionary.
- Enhanced outcomes in lots of deep learning applications.

Word2vec motivation

- Relationship among the words are preserved.
- Makes the best of addition of new words in the dictionary.
- Enhanced outcomes in lots of deep learning applications.

Functioning of Word2vec

The main goal is, the words that occur in similar context should have similar embeddings.

Functioning of Word2vec

The main goal is, the words that occur in similar context should have similar embeddings.

Example

- Mango is a seasonal fruit.
- Orange is a seasonal fruit.

The word **Mango** and **Orange** will have similar word vector

Table of Contents

Skip Gram

- Predict the context words from target

Skip Gram

- Predict the context words from target

$\boxed{\text{student}}$

Skip Gram

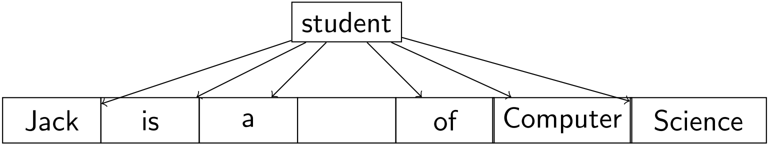
- Predict the context words from target

student

Jack	is	a		of	Computer	Science

Skip Gram

- Predict the context words from target



Continuous Bag of Words (CBOW)

- Predict the target word from the context.

Continuous Bag of Words (CBOW)

- Predict the target word from the context.

Jack	is	a		of	Computer	Science

Continuous Bag of Words (CBOW)

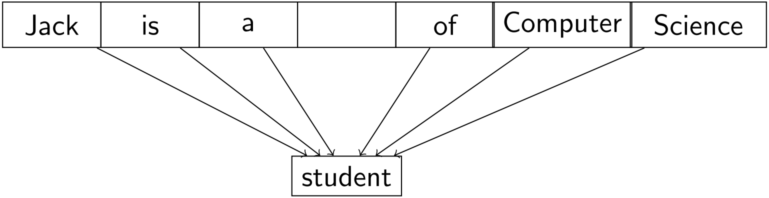
- Predict the target word from the context.

Jack	is	a		of	Computer	Science

student

Continuous Bag of Words (CBOW)

- Predict the target word from the context.



Working flow - Skip Gram

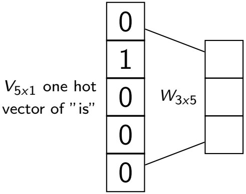
$\boxed{\text{Jack is a}}$ university student

Working flow - Skip Gram

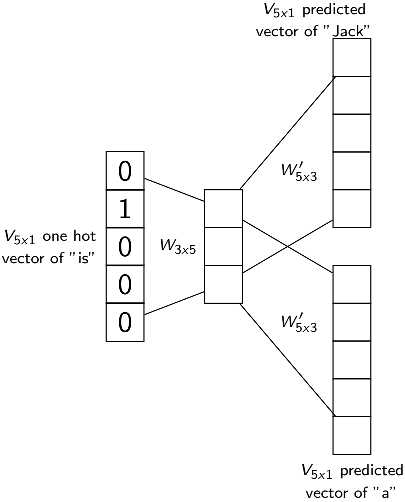
$\boxed{\text{Jack is a}}$ university student

V_{5x1} one hot vector of "is" $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

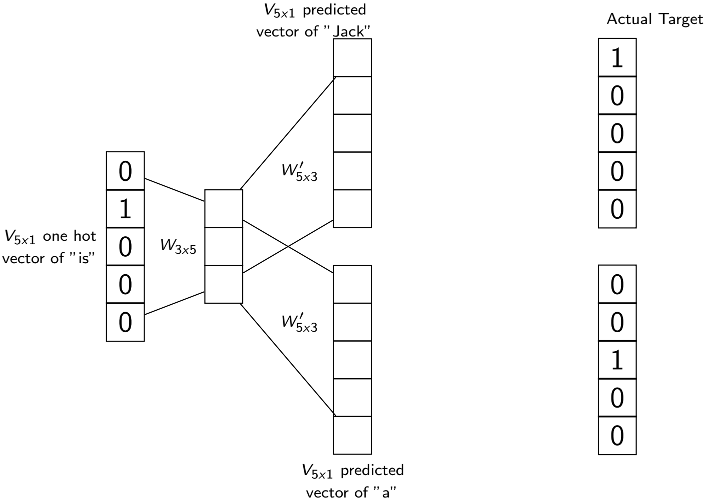
$\boxed{\text{Jack is a}}$ university student

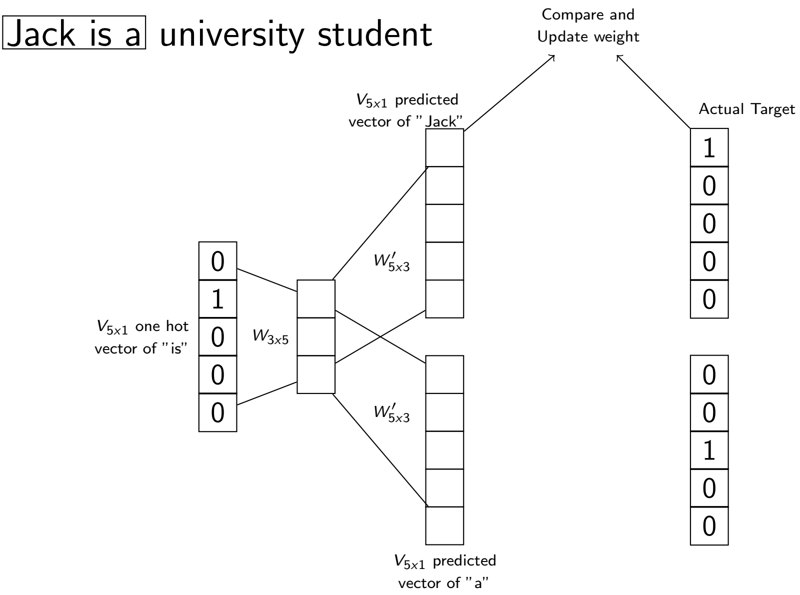


$\boxed{\text{Jack is a}}$ university student



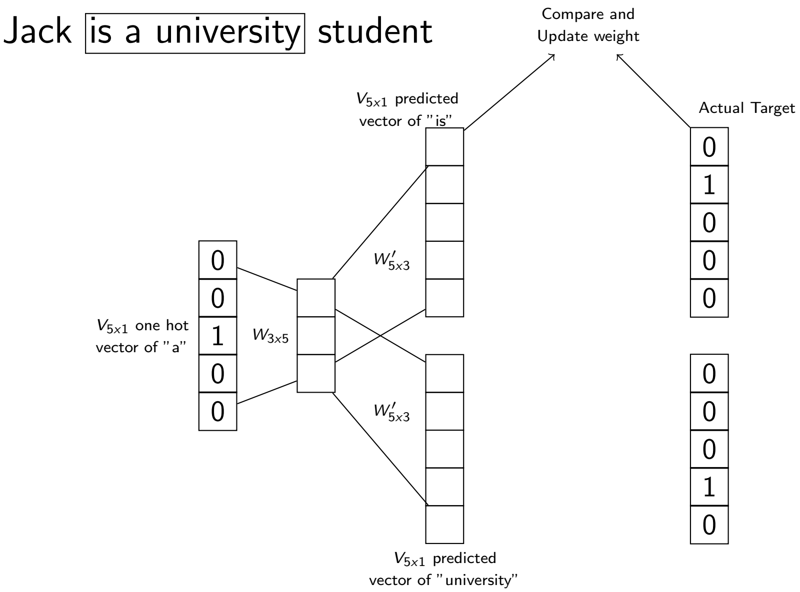
$\boxed{\text{Jack is a}}$ university student





Jack is a university student
Compare and Update weight
V_{5x1} predicted vector of "Jack"
Actual Target
1
0
0
0
0
W'_{5x3}
0
1
0
0
0
V_{5x1} one hot vector of "is"
W_{3x5}
W'_{5x3}
0
0
1
0
0
V_{5x1} predicted vector of "a"

Working flow - Skip Gram



Working flow - Skip Gram

Jack is $\boxed{\text{a university student}}$

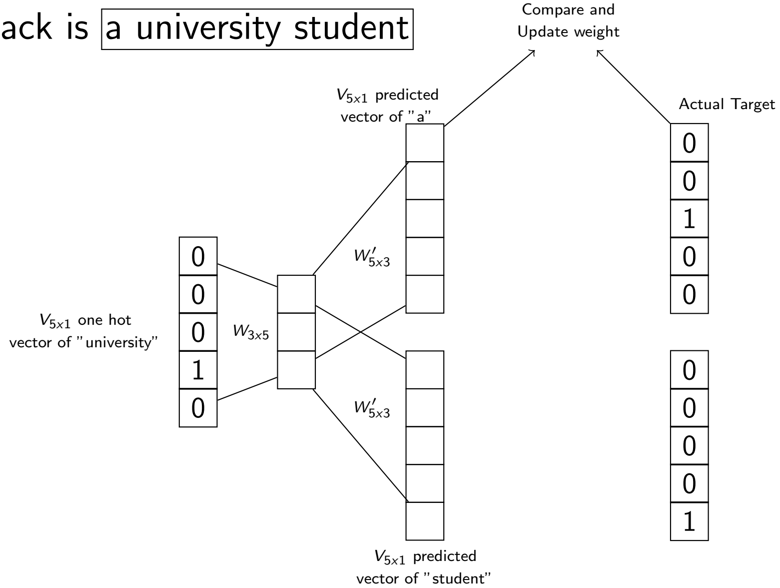


Table of Contents

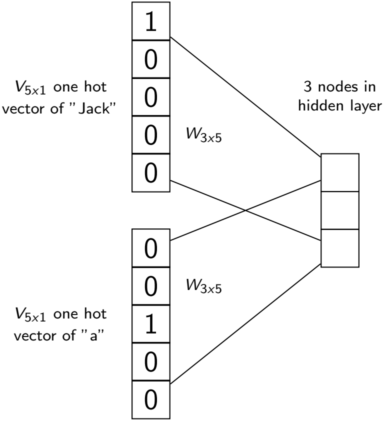
$\boxed{\text{Jack is a}}$ university student

$\boxed{\text{Jack is a}}$ university student

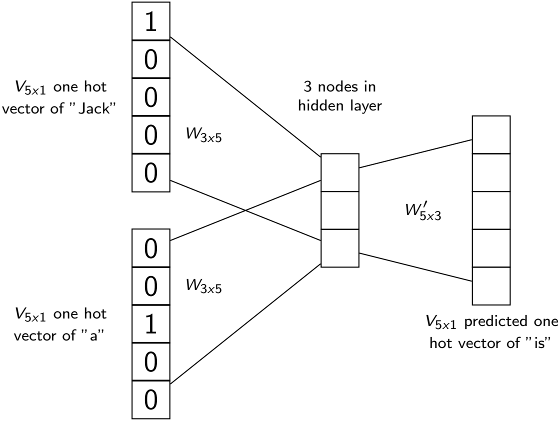
V_{5x1} one hot vector of "Jack" $\begin{array}{|c|}\hline 1 \\ \hline 0 \\ \hline 0 \\ \hline 0 \\ \hline 0 \\ \hline \end{array}$

V_{5x1} one hot vector of "a" $\begin{array}{|c|}\hline 0 \\ \hline 0 \\ \hline 1 \\ \hline 0 \\ \hline 0 \\ \hline \end{array}$

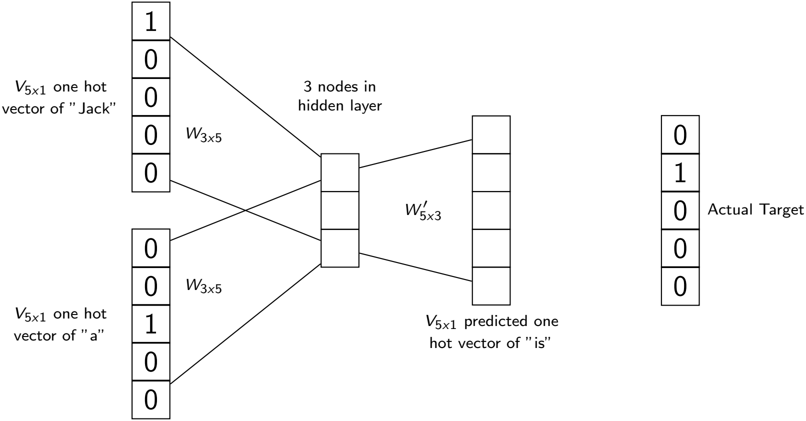
$\boxed{\text{Jack is a}}$ university student



$\boxed{\text{Jack is a}}$ university student

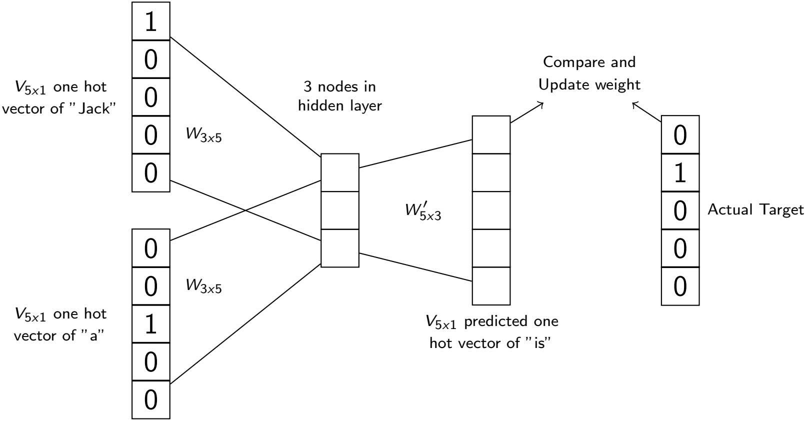


$\boxed{\text{Jack is a}}$ university student



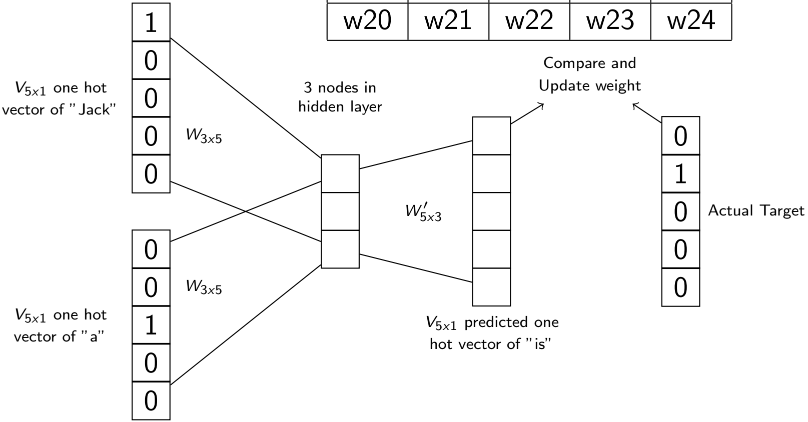
Working flow - CBOW

Jack is a university student



$\boxed{\text{Jack is a}}$ university student

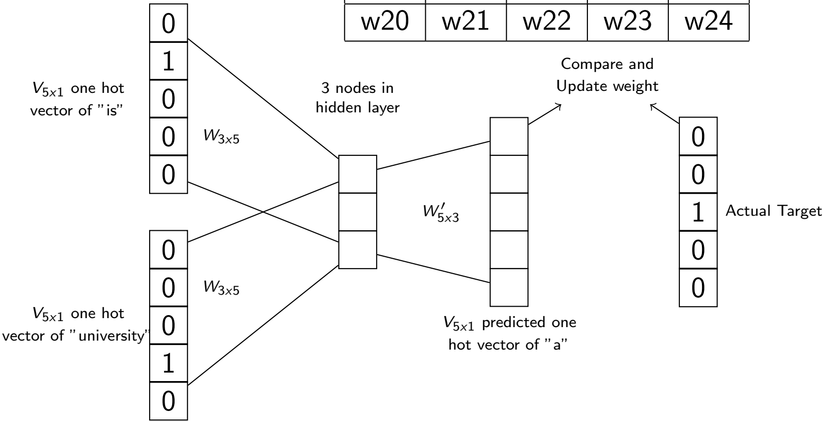
w00	w01	w02	w03	w04
w10	w11	w12	w13	w14
w20	w21	w22	w23	w24



Working flow - CBOW

Jack $\boxed{\text{is a university}}$ student

w00	w01	w02	w03	w04
w10	w11	w12	w13	w14
w20	w21	w22	w23	w24



Jack is $\boxed{\text{a university student}}$

w00	w01	w02	w03	w04
w10	w11	w12	w13	w14
w20	w21	w22	w23	w24

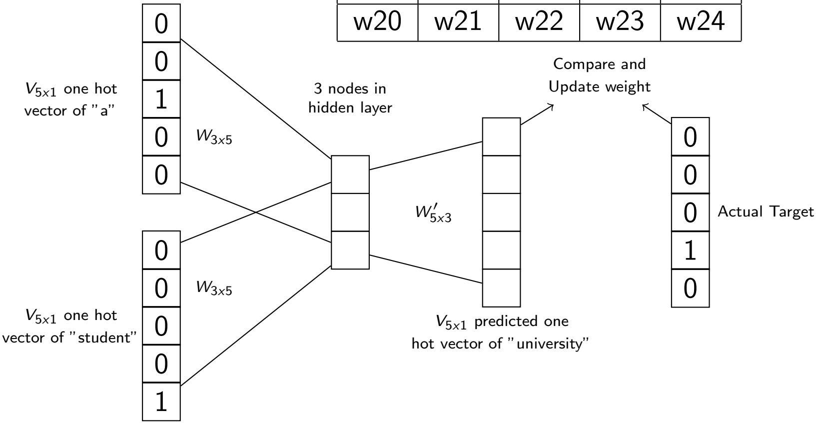


Table of Contents

Generating Word Vectors

Weights after training $W_{3\times5}$

w00	w01	w02	w03	w04
w10	w11	w12	w13	w14
w20	w21	w22	w23	w24

One hot vector of words, $V_{5\times1}$

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
Jack	is	a	university	student

Word vector for “Jack” = $W_{3\times5} \times V_{5\times1}$

$$\begin{bmatrix} w00 & w01 & w02 & w03 & w04 \\ w10 & w11 & w12 & w13 & w14 \\ w20 & w21 & w22 & w23 & w24 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} w00 \\ w10 \\ w20 \end{bmatrix}}_{V_{3\times1}}$$

Word Vector for “Jack”

Table of Contents

Precision Improvement

Precision Improvement

- Selection of framework architecture
  - Large corpus, higher dimensions - **Skip gram** (slower)
  - Small corpus, lower dimensions - **CBOW** (faster)

Precision Improvement

- Selection of framework architecture
  - Large corpus, higher dimensions - **Skip gram** (slower)
  - Small corpus, lower dimensions - **CBOW** (faster)
- Training data set improvement

- Selection of framework architecture
  - Large corpus, higher dimensions - **Skip gram** (slower)
  - Small corpus, lower dimensions - **CBOW** (faster)
- Training data set improvement
- Expansion of vector dimensions

Precision Improvement

- Selection of framework architecture
  - Large corpus, higher dimensions - **Skip gram** (slower)
  - Small corpus, lower dimensions - **CBOW** (faster)
- Training data set improvement
- Expansion of vector dimensions
- Combining word2vec with **Pos2vec** or **Glove**

Table of Contents

Use Case

Use Case

- Natural Language Processing

Use Case

- Natural Language Processing
- Relationship between a country name and its capital name

Use Case

- Natural Language Processing
- Relationship between a country name and its capital name
- Analyze news headline to predict article success


THANK YOU