# One-hot encoding

"how are you ? ok"

Length would be 5 of each word' vector

"how" -> [1,0,0,0,0]
"are" -> [0,1,0,0,0]
"you" -> [0,0,1,0,0]
"?" -> [0,0,0,1,0]
"ok" -> [0,0,0,0,1]

1 in unique spot and all 0

On-hot Encoding

| How | Are | You | ? | Ok |
|---|---|---|---|---|
| 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 |

# Bag of words (BOW)

1 . "I will go to Mumbai then go to Pune"
2. "Yesterday I went to Mumbai"

| | |
|---|---|
| I | 1 |
| Will | 1 |
| Go | 2 |
| To | 2 |
| Mumbai | 1 |
| Then | 1 |
| Pune | 1 |

| | |
|---|---|
| Yesterday | 1 |
| I | 1 |
| Went | 1 |
| To | 1 |
| Mumbai | 1 |

| | |
|---|---|
| I | 2 |
| Will | 1 |
| Go | 2 |
| To | 3 |
| Mumbai | 2 |
| Then | 1 |
| Pune | 1 |
| Yesterday | 1 |
| went | 1 |

**corpus** : the tokens we have from all the documents, considering for the bag of words creation

BOW

| word / document | I | Will | Go | To | Mumbai | Then | Pune | Yesterday | Went |
|---|---|---|---|---|---|---|---|---|---|
| doc1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| doc2 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

# Bag of words (BOW)

| word | I | Will | Go | To | Mumbai | Then | Pune | Yesterday | Went |
|---|---|---|---|---|---|---|---|---|---|
| document | | | | | | | | | |
| doc1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| doc2 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

[1 1 1 1 1 1 1 0 0],
[1 0 0 1 1 0 0 1 1]

# Bag of words (BOW)

## Limitations

- Importance are same for all words
- No semantic information

1. “I will **go** to Mumbai then go to Pune”
2. “I will be **going** to Mumbai”

## Solution??

# TF-IDF

1 . “I will go to Mumbai then go to Pune”
2. “Yesterday I went to Mumbai”

| | |
|---|---|
| I | 1/7 |
| Will | 1/7 |
| Go | 2/7 |
| To | 2/7 |
| Mumbai | 1/7 |
| Then | 1/7 |
| Pune | 1/7 |

| | |
|---|---|
| Yesterday | 1/5 |
| I | 1/5 |
| Went | 1/5 |
| To | 1/5 |
| Mumbai | 1/5 |

World frequency in all documents

| | |
|---|---|
| I | 2 |
| Will | 1 |
| Go | 2 |
| To | 3 |
| Mumbai | 2 |
| Then | 1 |
| Pune | 1 |
| Yesterday | 1 |
| went | 1 |

## TF

| word<br>document | I | Will | Go | To | Mumbai | Then | Pune | Yesterday | Went |
|---|---|---|---|---|---|---|---|---|---|
| TF1 | 0.14 | 0.14 | 0.28 | 0.28 | 0.14 | 0.14 | 0.14 | 0 | 0 |
| TF2 | 0.25 | 0 | 0 | 0.25 | 0.25 | 0 | 0 | 0.25 | 0.25 |

# TF-IDF

World frequency in all documents

1 . "I will go to Mumbai then go to Pune"
2. "Yesterday I went to Mumbai"

| | |
|---|---|
| I | 1/7 |
| Will | 1/7 |
| Go | 2/7 |
| To | 2/7 |
| Mumbai | 1/7 |
| Then | 1/7 |
| Pune | 1/7 |

| | |
|---|---|
| Yesterday | 1/5 |
| I | 1/5 |
| Went | 1/5 |
| To | 1/5 |
| Mumbai | 1/5 |

IDF – inverse document frequency

$$\text{IDF} = \text{LOG}\left( \frac{\textit{No of documents}}{\text{(No of documents that contains a word)}} \right)$$

IDF is used to calculate weight of words.
Rare word has high Weight

| words | Word frequency in all documents | IDF | IDF |
|---|---|---|---|
| I | 2 | Log(2/2) | 0 |
| Will | 1 | Log(2/1) | 0.69 |
| Go | 2 | Log(2/1) | 0.69 |
| To | 3 | Log(2/2) | 0 |
| Mumbai | 2 | Log(2/2) | 0 |
| Then | 1 | Log(2/1) | 0.69 |
| Pune | 1 | Log(2/1) | 0.69 |
| Yesterday | 1 | Log(2/1) | 0.69 |
| went | 1 | Log(2/1) | 0.69 |

# TF-IDF

| | | | | | | TF-IDF | TF-IDF |
|---|---|---|---|---|---|---|---|
| words | Word frequency in all documents | TF1 | TF2 | IDF | IDF | IDF*TF1 | IDF*TF2 |
| I | 2 | 0.14 | 0.25 | Log(2/2) | 0 | 0 | 0 |
| Will | 1 | 0.14 | 0 | Log(2/1) | 0.69 | 0.1 | 0 |
| Go | 2 | 0.28 | 0 | Log(2/1) | 0.69 | 0.19 | 0 |
| To | 3 | 0.28 | 0.25 | Log(2/2) | 0 | 0 | 0 |
| Mumbai | 2 | 0.14 | 0.25 | Log(2/2) | 0 | 0 | 0 |
| Then | 1 | 0.14 | 0 | Log(2/1) | 0.69 | 0.1 | 0 |
| Pune | 1 | 0.14 | 0 | Log(2/1) | 0.69 | 0.1 | 0 |
| Yesterday | 1 | 0 | 0.25 | Log(2/1) | 0.69 | 0 | 0.17 |
| went | 1 | 0 | 0.25 | Log(2/1) | 0.69 | 0 | 0.17 |

# Word2Vec

King = Man

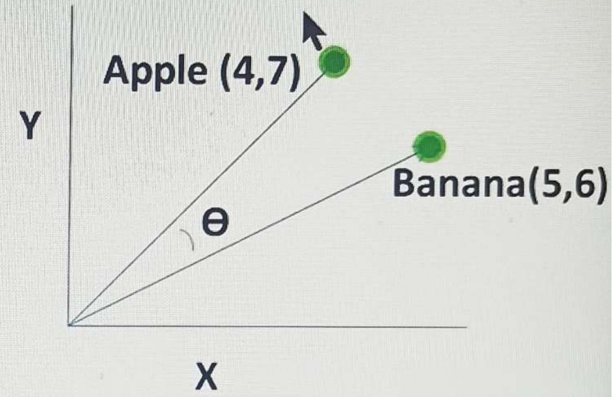
Queen = Woman

How to get relation ??

# Word2Vec

- Cosine of the angle should between such vectors should be close to 1

$$\mathrm{Cos}(\Theta) = \frac{A.B}{\sqrt{\sum_{i=1}^{n} A_i^2}\ \sqrt{\sum_{i=1}^{n} B_i^2}} = \frac{A.B}{||A||.||B||}$$




It can have n – dimension, showing for you in 2-D

$$A.B = (4x5) + (7X6) = 62$$

$$||A|| = \sqrt{4^2 + 7^2} = 8.06$$

$$||B|| = \sqrt{5^2 + 6^2} = 7.81$$

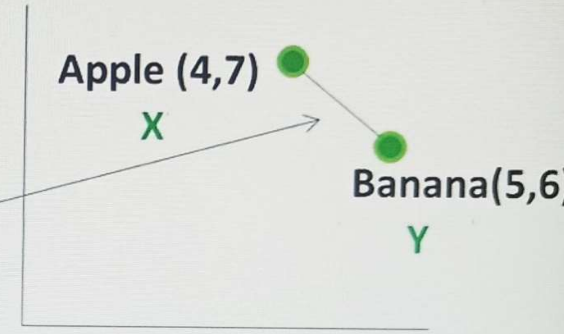
Similarity = 62/(8.06*7.81) = 0.961

# Word2Vec

***Euclidean distance***

$$d^E(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

*Euclidean distance* $= \sqrt{(4\text{-}5)^2 + (7\text{-}6)^2} = 1.41$




Word2Vec
Apple
Mango
Banana
Room
Hotel
Potato
Kitchen
tomato
Cook
Restaurant
Customer
City
Word relations

## Word2Vec

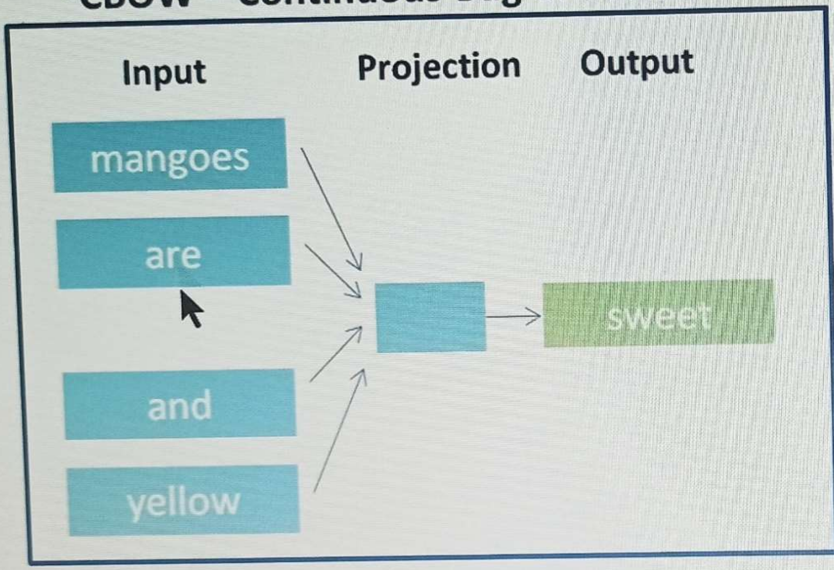
Word2Vec can be obtained using two methods

- CBOW – Continuous Bag of Words
- Skip Gram

Mangoes are ······ and yellow

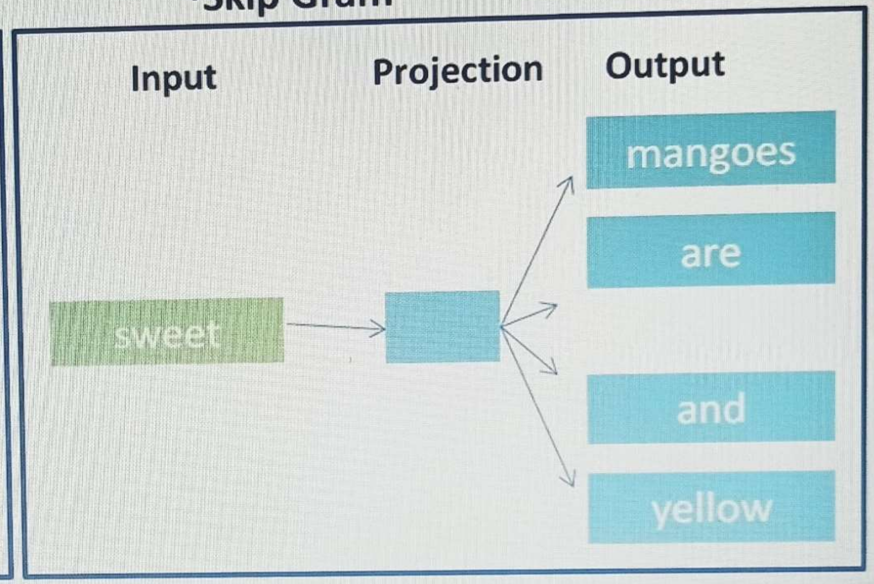
# Word2Vec

## •CBOW – Continuous Bag of Words

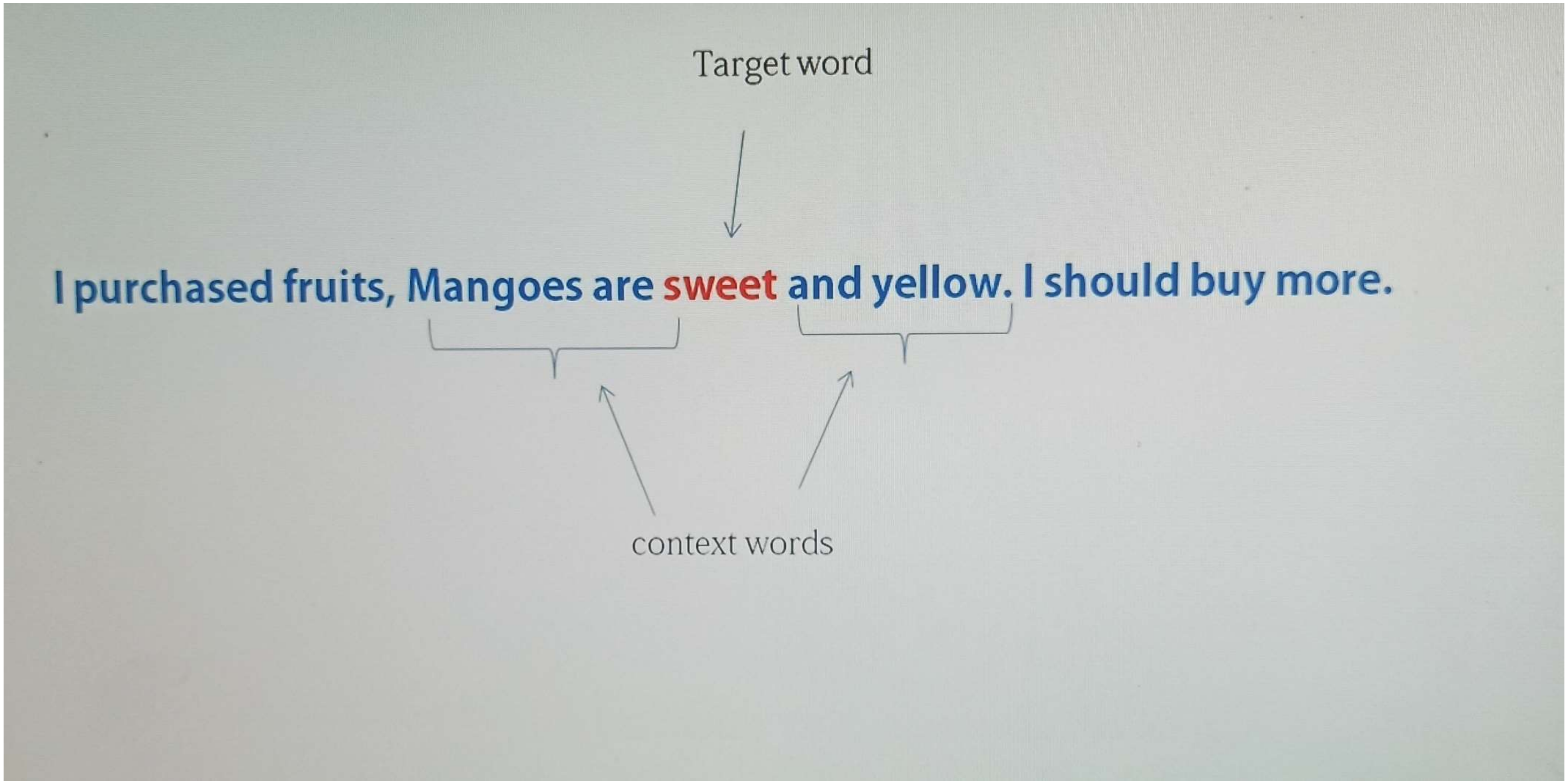


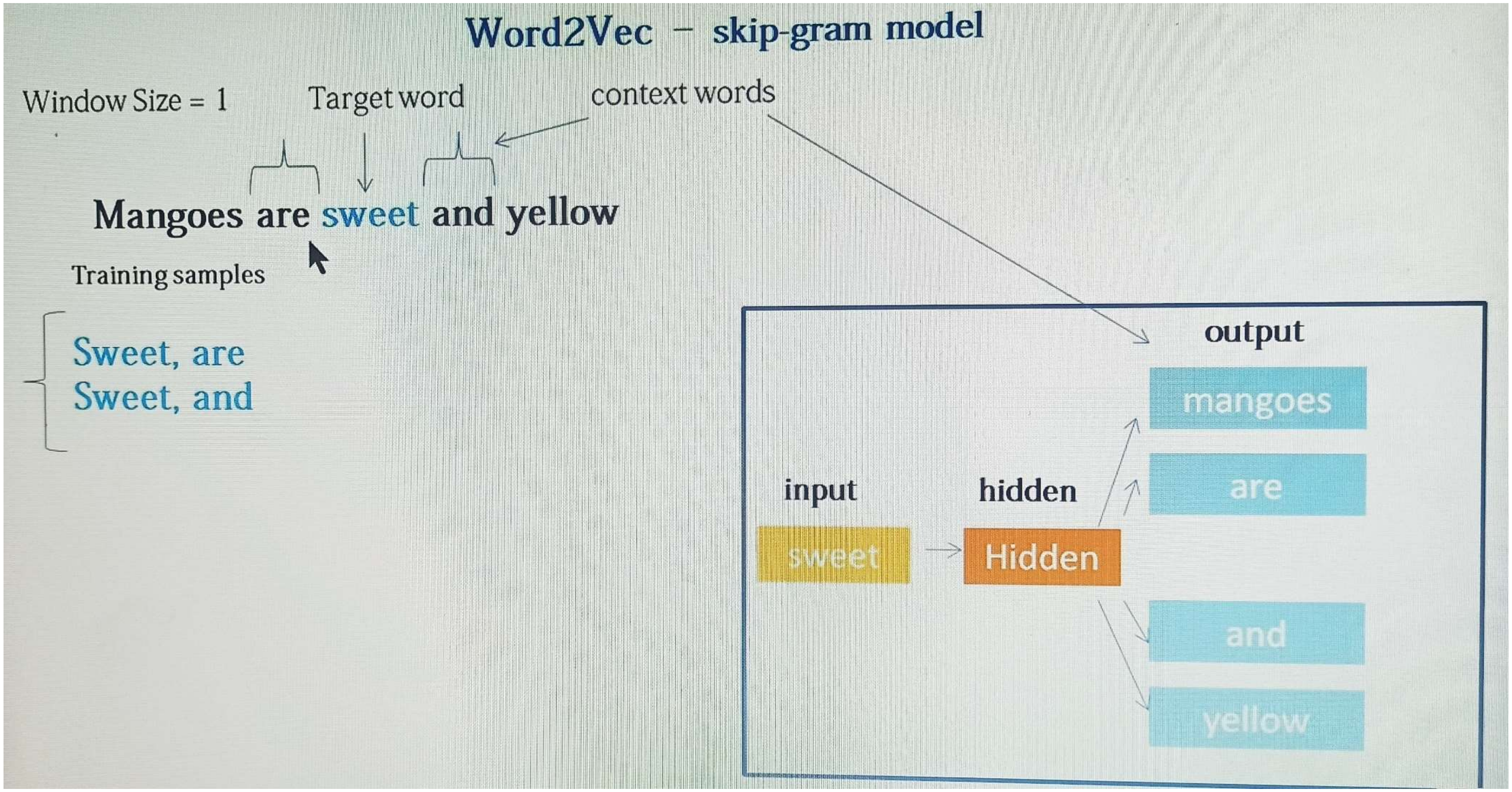

Neighbor words are input and output is the center word.
Works good with large datasets.
Better representation for frequent words than rarer.

## •Skip Gram




Input is centre word and output is context words (neighbour words).
Works well with small datasets.
Skip-gram identifies rarer words better.


Target word
I purchased fruits, Mangoes are sweet and yellow. I should buy more.
context words


Word2Vec – skip-gram model
Window Size = 1
Target word
context words
Mangoes are sweet and yellow
Training samples
Sweet, are
Sweet, and
input
hidden
output
sweet
Hidden
mangoes
are
and
yellow

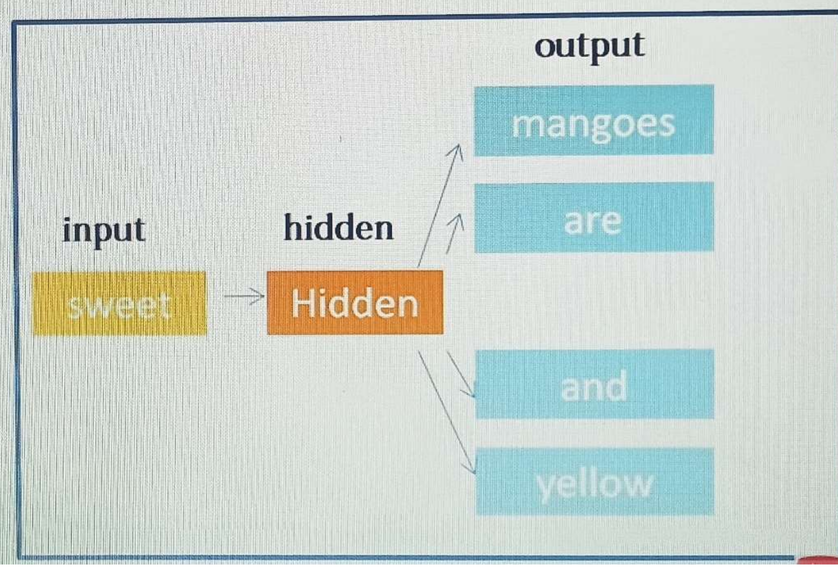
# Word2Vec – skip-gram model

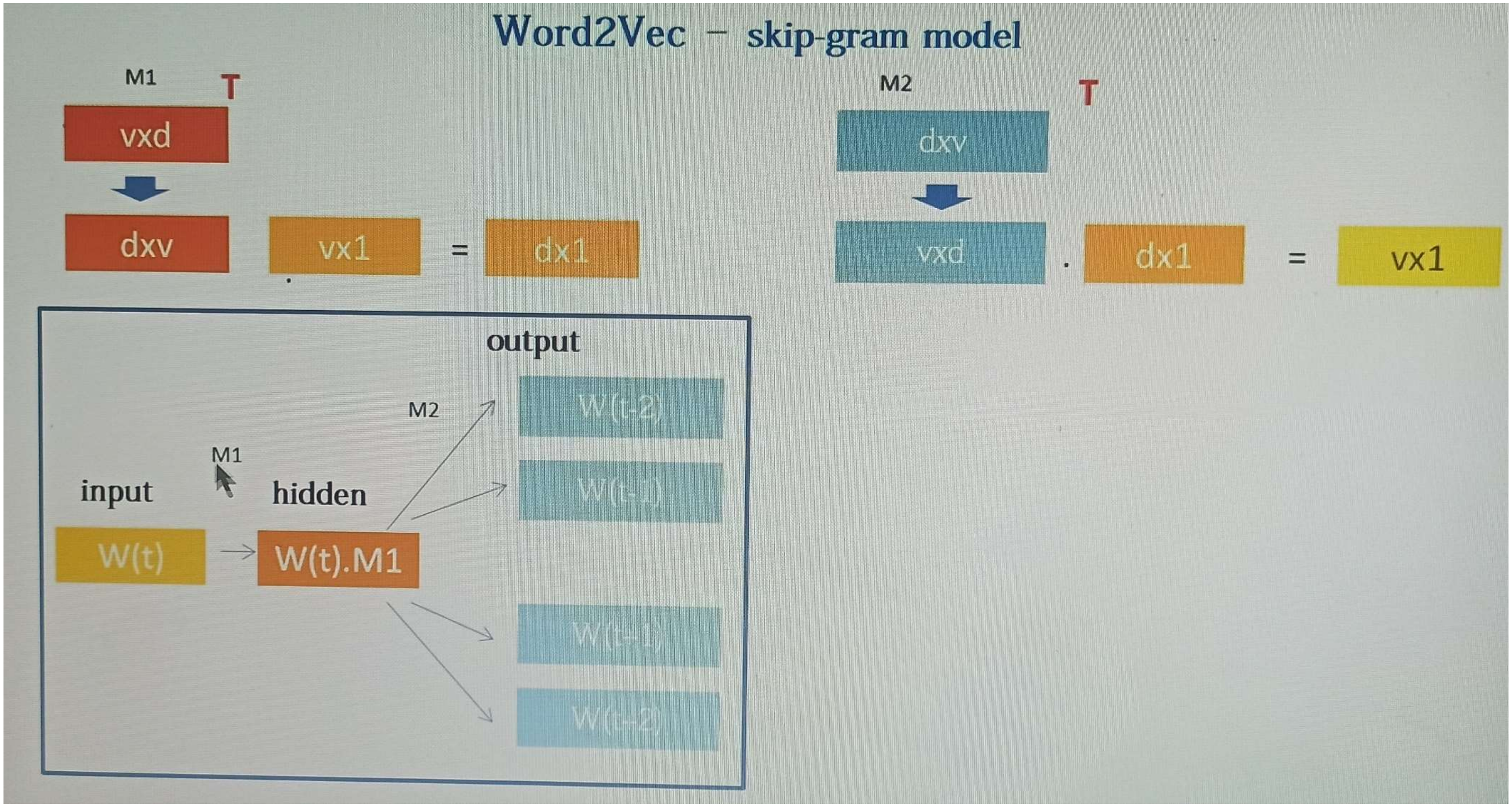
Window Size = 2    Target word    context words

**Mangoes are sweet and yellow**

Training samples

Sweet, mangoes
Sweet, are
Sweet, and
Sweet, yellow

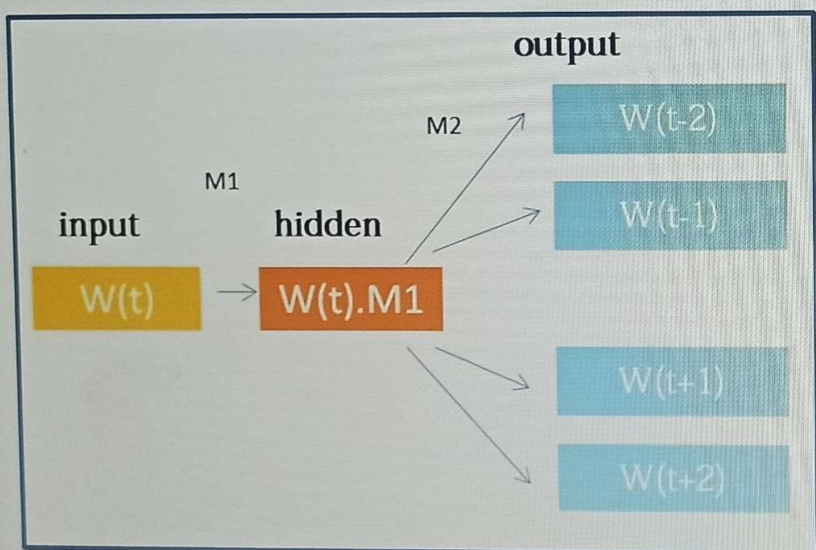


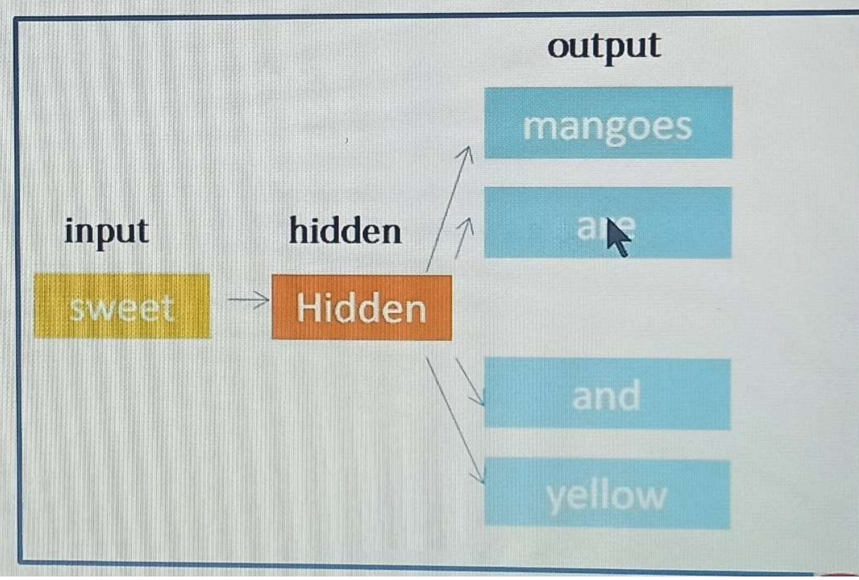

Word2Vec – skip-gram model
M1
T
vxd
dxv
vx1
=
dx1
M2
T
dxv
vxd
.
dx1
=
vx1
output
M2
W(t-2)
W(t-1)
M1
input
hidden
W(t)
W(t).M1
W(t+1)
W(t+2)

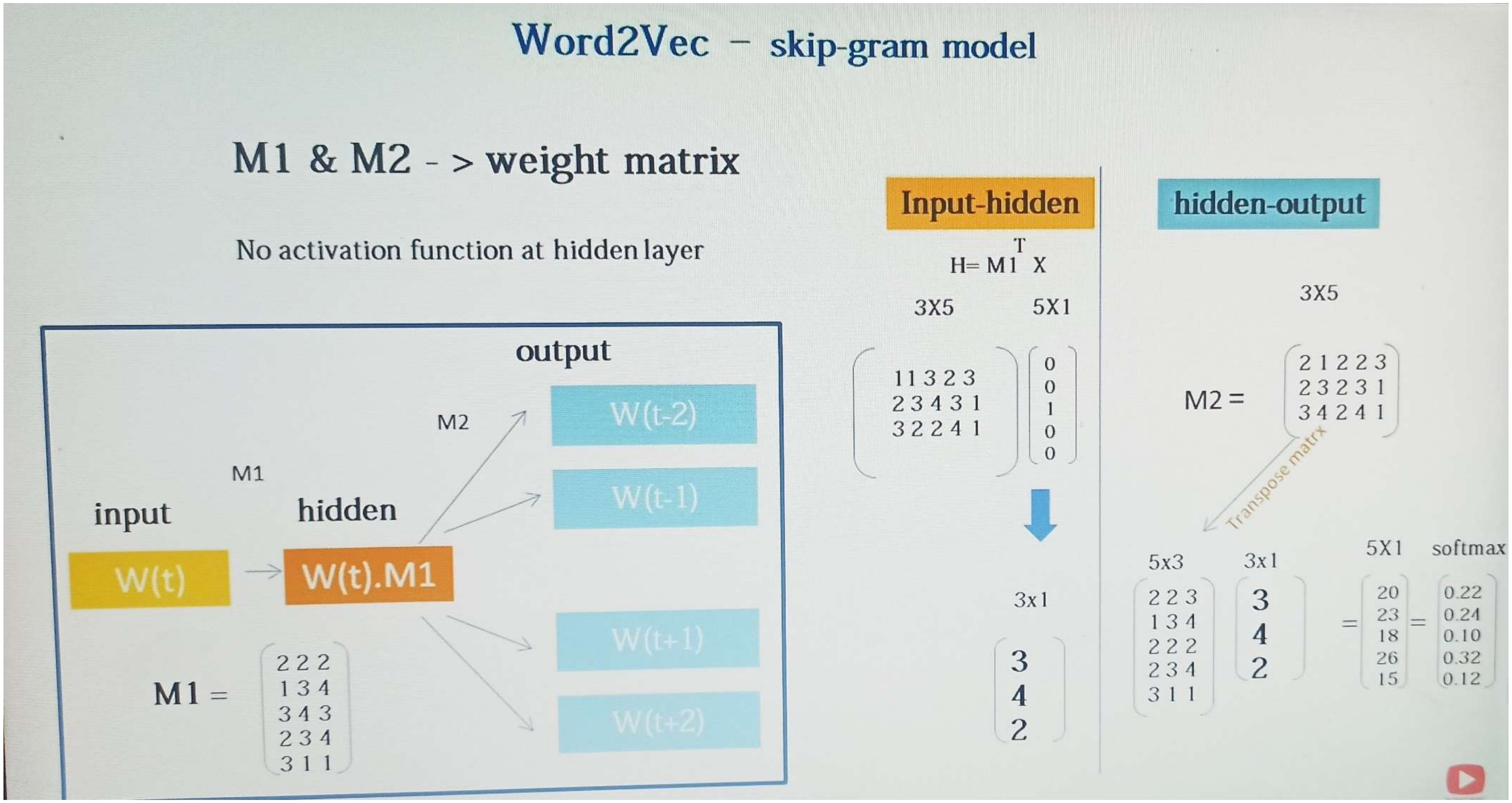
# Word2Vec – skip-gram model

## M1 & M2 - > weight matrix

- Imp: two weight matrices.
- No activation function at hidden layer
- Softmax fn at output layer



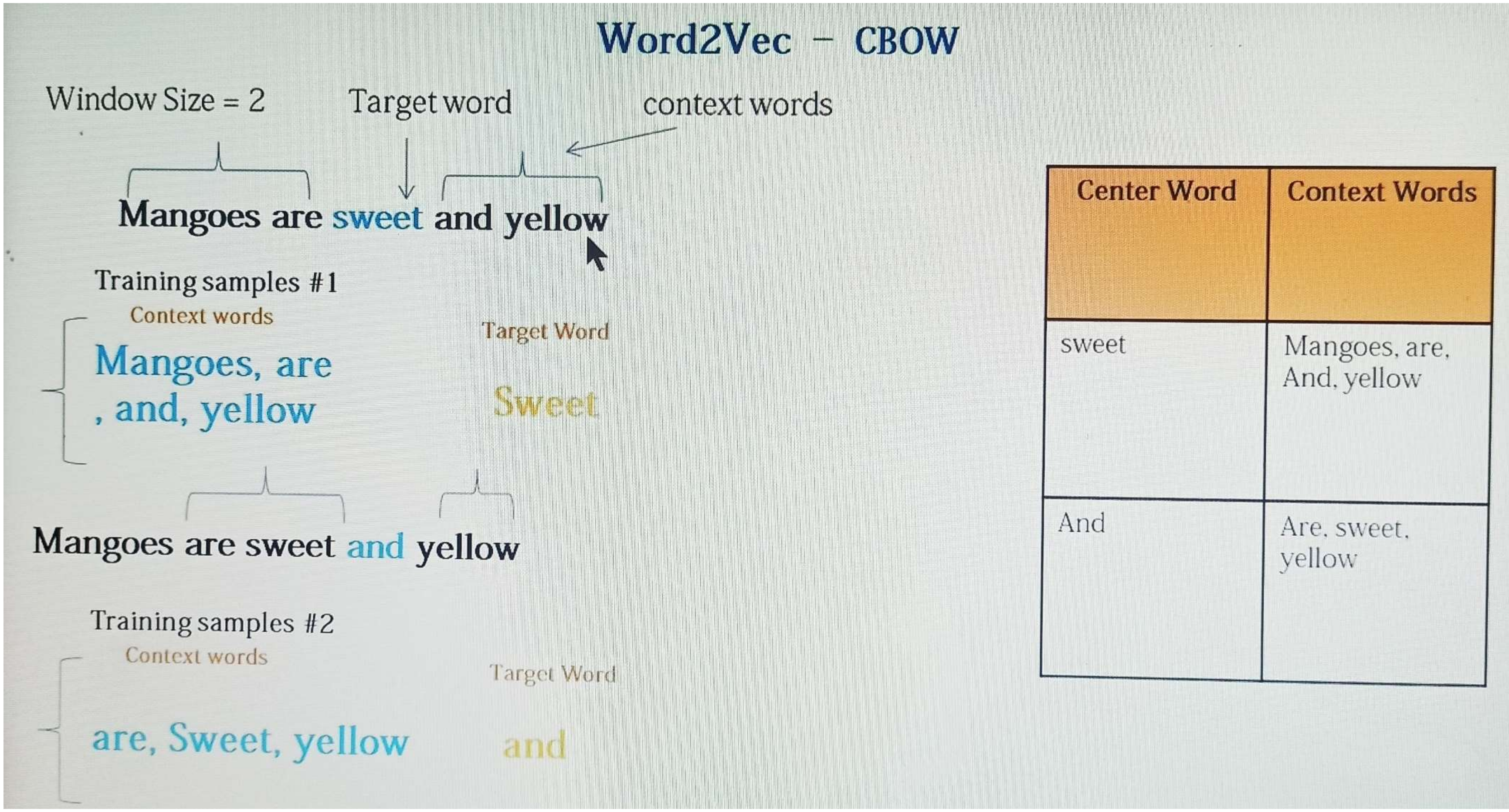





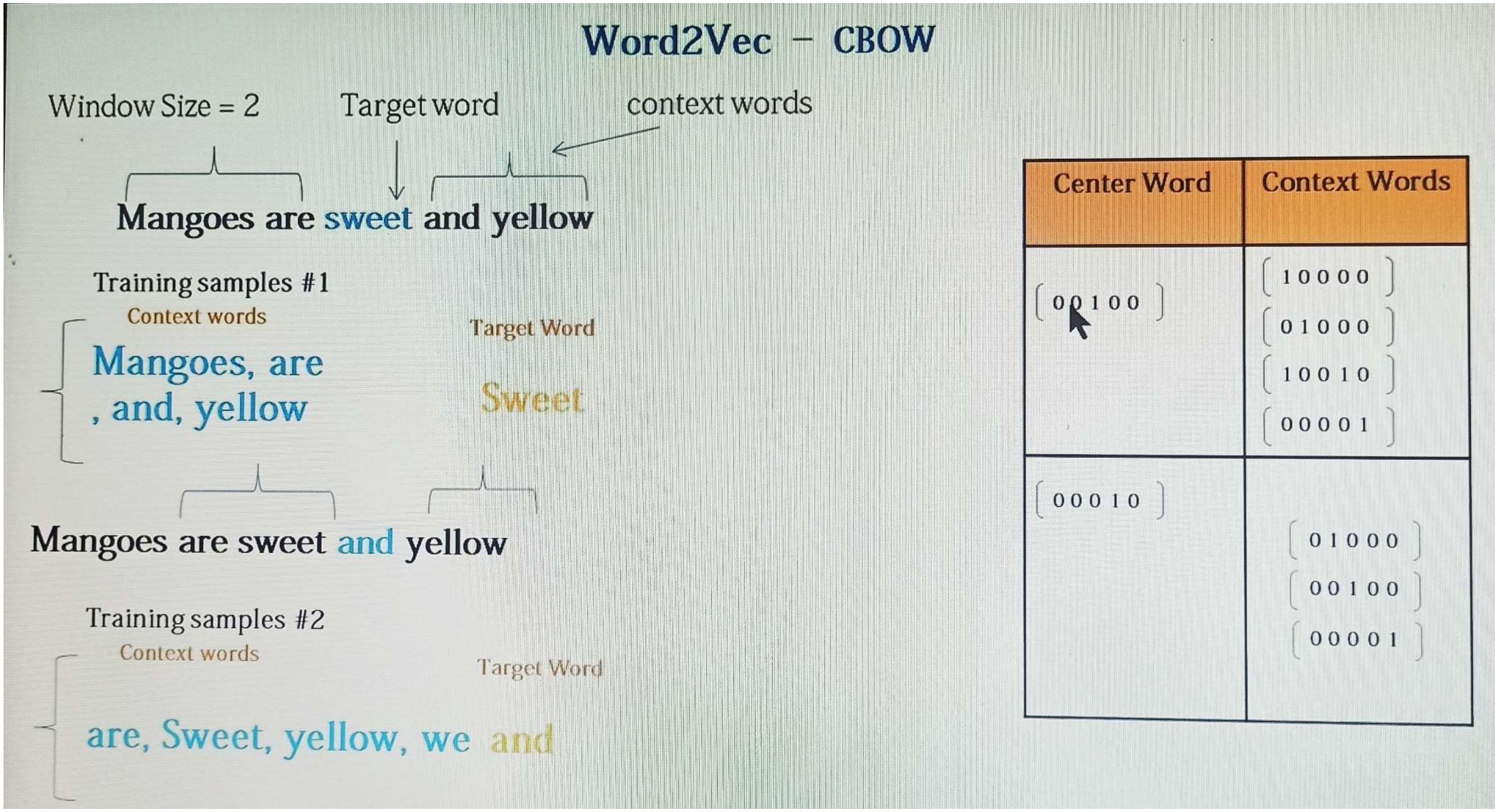
Word2Vec – skip-gram model
M1 & M2 - > weight matrix
No activation function at hidden layer
output
M2
W(t-2)
W(t-1)
M1
input
hidden
W(t)
W(t).M1
W(t+1)
W(t+2)
M1 =
2 2 2
1 3 4
3 4 3
2 3 4
3 1 1
Input-hidden
H= M1 T X
3X5
5X1
1 1 3 2 3
2 3 4 3 1
3 2 2 4 1
0
0
1
0
0
3x1
3
4
2
hidden-output
3X5
M2 =
2 1 2 2 3
2 3 2 3 1
3 4 2 4 1
Transpose matrx
5x3
3x1
5X1
softmax
2 2 3
1 3 4
2 2 2
2 3 4
3 1 1
3
4
2
=
20
23
18
26
15
=
0.22
0.24
0.10
0.32
0.12

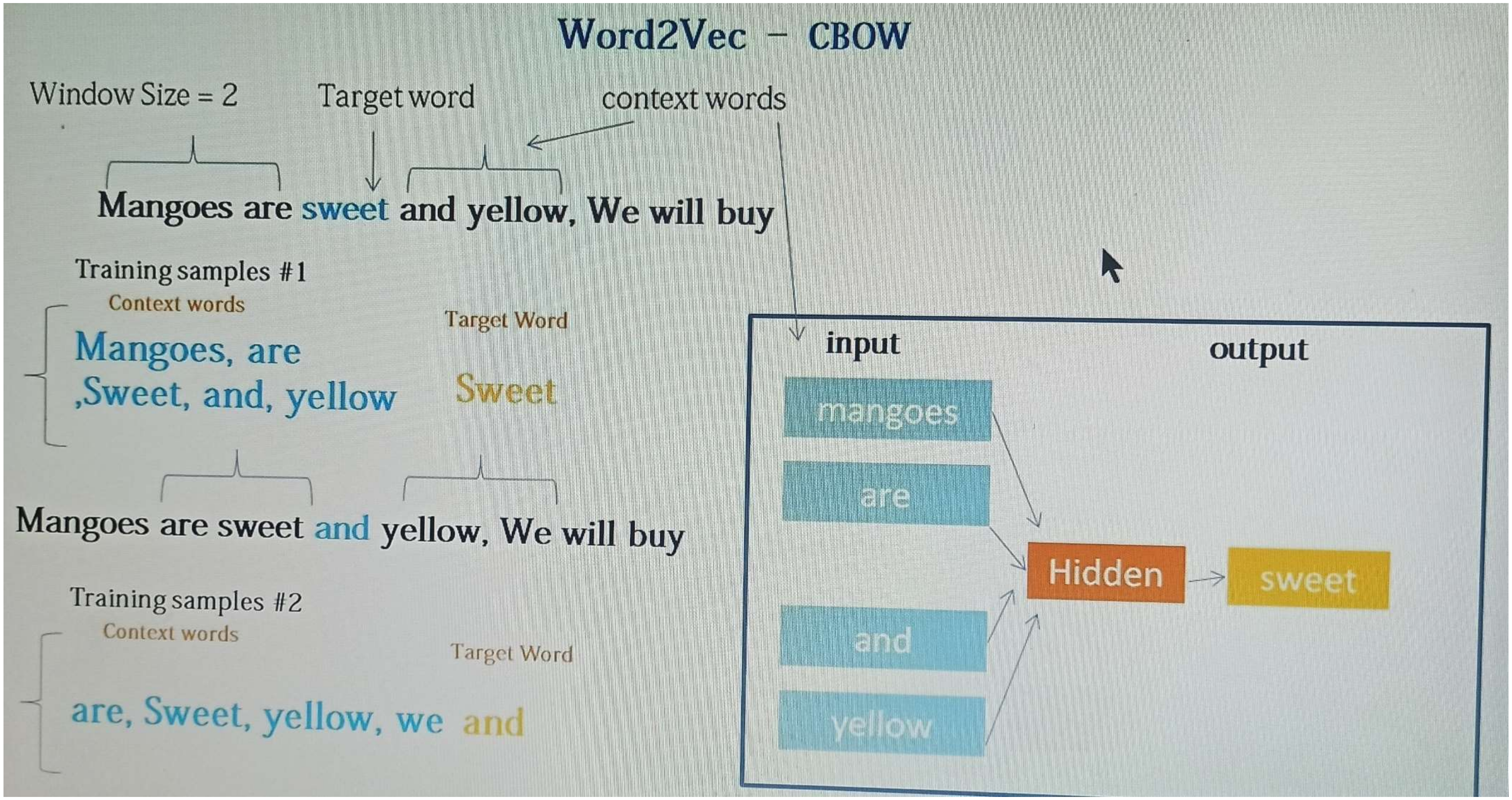
# NLP

- **Word embedding techniques**
  - Word2vec
    - CBOW - Continuous Bag of Words




| Center Word | Context Words |
|---|---|
| sweet | Mangoes, are, And, yellow |
| And | Are, sweet, yellow |


Word2Vec – CBOW
Window Size = 2
Target word
context words
Mangoes are sweet and yellow
Training samples #1
Context words
Mangoes, are
, and, yellow
Target Word
Sweet
Mangoes are sweet and yellow
Training samples #2
Context words
are, Sweet, yellow, we
Target Word
and
Center Word
Context Words
0 0 1 0 0
1 0 0 0 0
0 1 0 0 0
1 0 0 1 0
0 0 0 0 1
0 0 0 1 0
0 1 0 0 0
0 0 1 0 0
0 0 0 0 1


Word2Vec - CBOW
Window Size = 2
Target word
context words
Mangoes are sweet and yellow, We will buy
Training samples #1
Context words
Mangoes, are
,Sweet, and, yellow
Target Word
Sweet
Mangoes are sweet and yellow, We will buy
Training samples #2
Context words
are, Sweet, yellow, we
Target Word
and
input
output
mangoes
are
and
yellow
Hidden
sweet

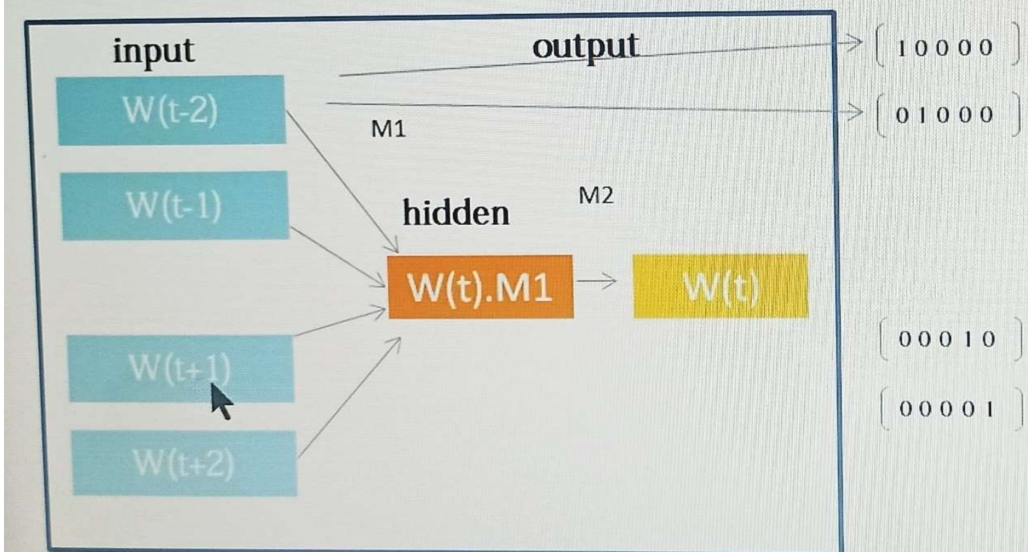
# Word2Vec – CBOW model

## M1 & M2 - > weight matrix

No activation function at hidden layer

**Input-hidden**

Input

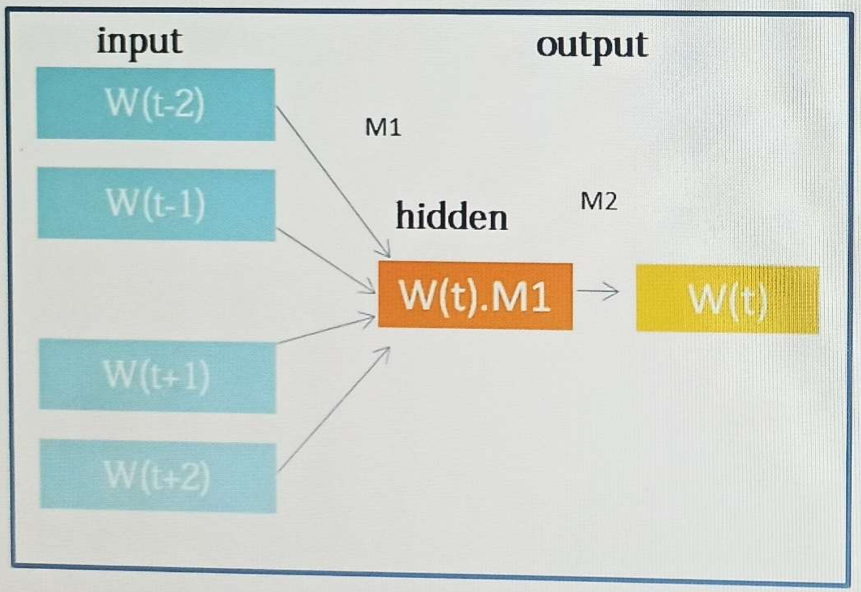



$$M1\,(\text{weight}) = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 3 & 4 & 2 \\ 2 & 3 & 4 \\ 3 & 1 & 1 \end{pmatrix}$$

# Word2Vec – CBOW model

## M1 & M2 - > weight matrix

No activation function at hidden layer




**Input-hidden**

Input

$$H = M1^{T} x$$

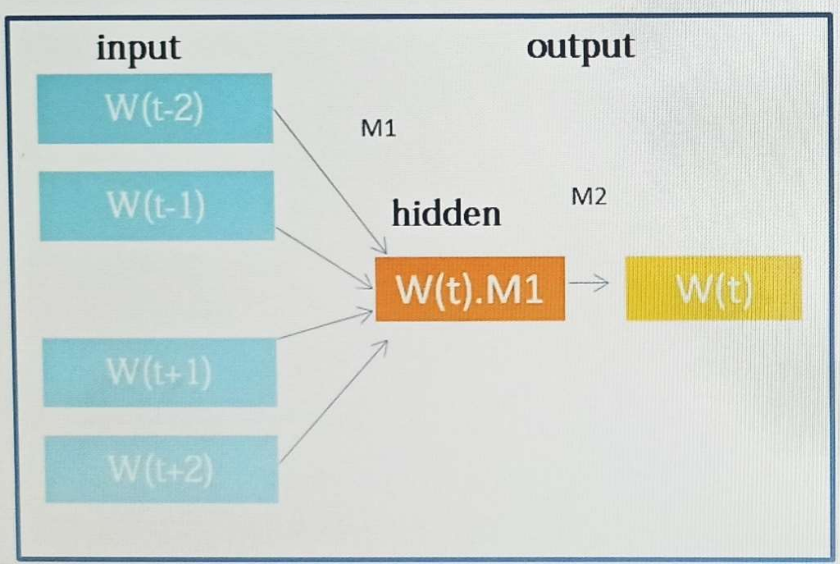
Hidden O/p

$$\begin{pmatrix} 1 & 1 & 3 & 2 & 3 \\ 2 & 3 & 4 & 3 & 1 \\ 3 & 2 & 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} 0.25 & 0.25 & 0 & 0.25 & 0.25 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 2.7 \\ 2.5 \end{pmatrix}$$

# Word2Vec – CBOW model

## M1 & M2 - > weight matrix

No activation function at hidden layer



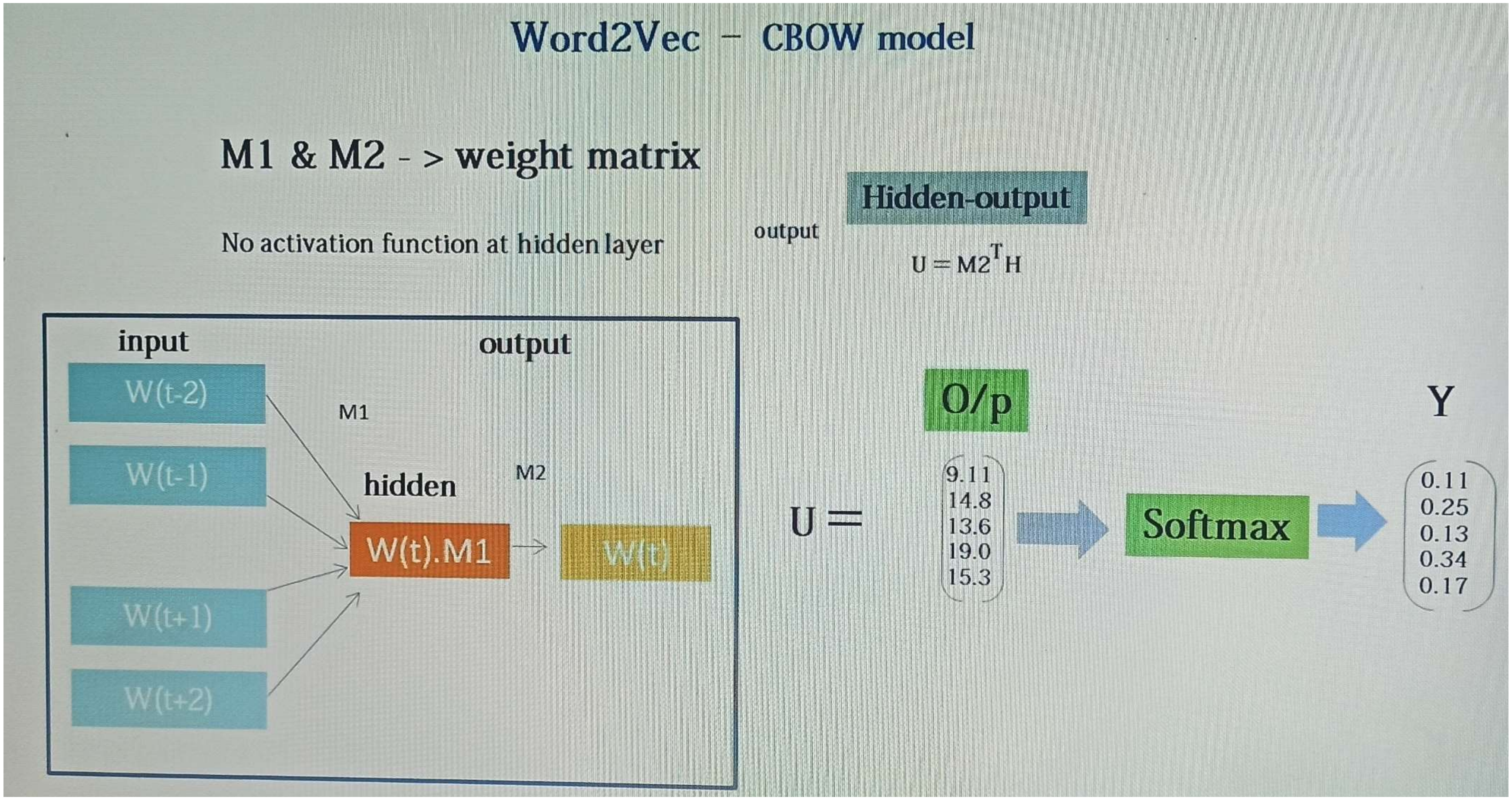

output

**Hidden-output**

$$U = M2^T H$$

**O/p - U**

$$\underset{M2^T}{\begin{pmatrix} 2 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 2 \\ 3 & 1 & 2 \end{pmatrix}} \underset{\text{Hidden O/p}}{\begin{pmatrix} 1.7 \\ 2.7 \\ 2.5 \end{pmatrix}} = \begin{pmatrix} 2x1.7 + 1x2.7 + 2X2.5 \\ 1x1.7 + 3x2.7 + 2X2.5 \\ 2x1.7 + 1x2.7 + 3X2.5 \\ 2x1.7 + 3x2.7 + 2X2.5 \\ 3x1.7 + 1x2.7 + 2X2.5 \end{pmatrix} = \begin{pmatrix} 9.11 \\ 14.8 \\ 13.6 \\ 19.0 \\ 15.3 \end{pmatrix}$$



Word2Vec – CBOW model
M1 & M2 - > weight matrix
No activation function at hidden layer
output
Hidden-output
U = M2^T H
input
output
W(t-2)
W(t-1)
W(t+1)
W(t+2)
M1
M2
hidden
W(t).M1
W(t)
O/p
U =
9.11
14.8
13.6
19.0
15.3
Softmax
Y
0.11
0.25
0.13
0.34
0.17