COMP5046
Natural Language Processing

Lecture 4: Word Classification and Machine Learning 2

Dr. Caren Han

Semester 1, 2021
School of Computer Science,
University of Sydney

LECTURE PLAN

Lecture 4: Word Classification and Machine Learning 2

1. **Machine Learning and NLP: Finish**
2. Seq2Seq Learning
3. Seq2Seq Deep Learning
   1. RNN (Recurrent Neural Network)
   2. LSTM (Long Short-Term Memory)
   3. GRU (Gated Recurrent Unit)
4. Data Transformation for Deep Learning NLP
5. Next Week Preview
   - Natural Language Processing Stack

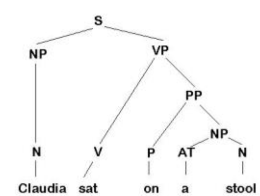
…. And some interesting notice in the end of the lecture!

The purpose of Natural Language Processing: Overview

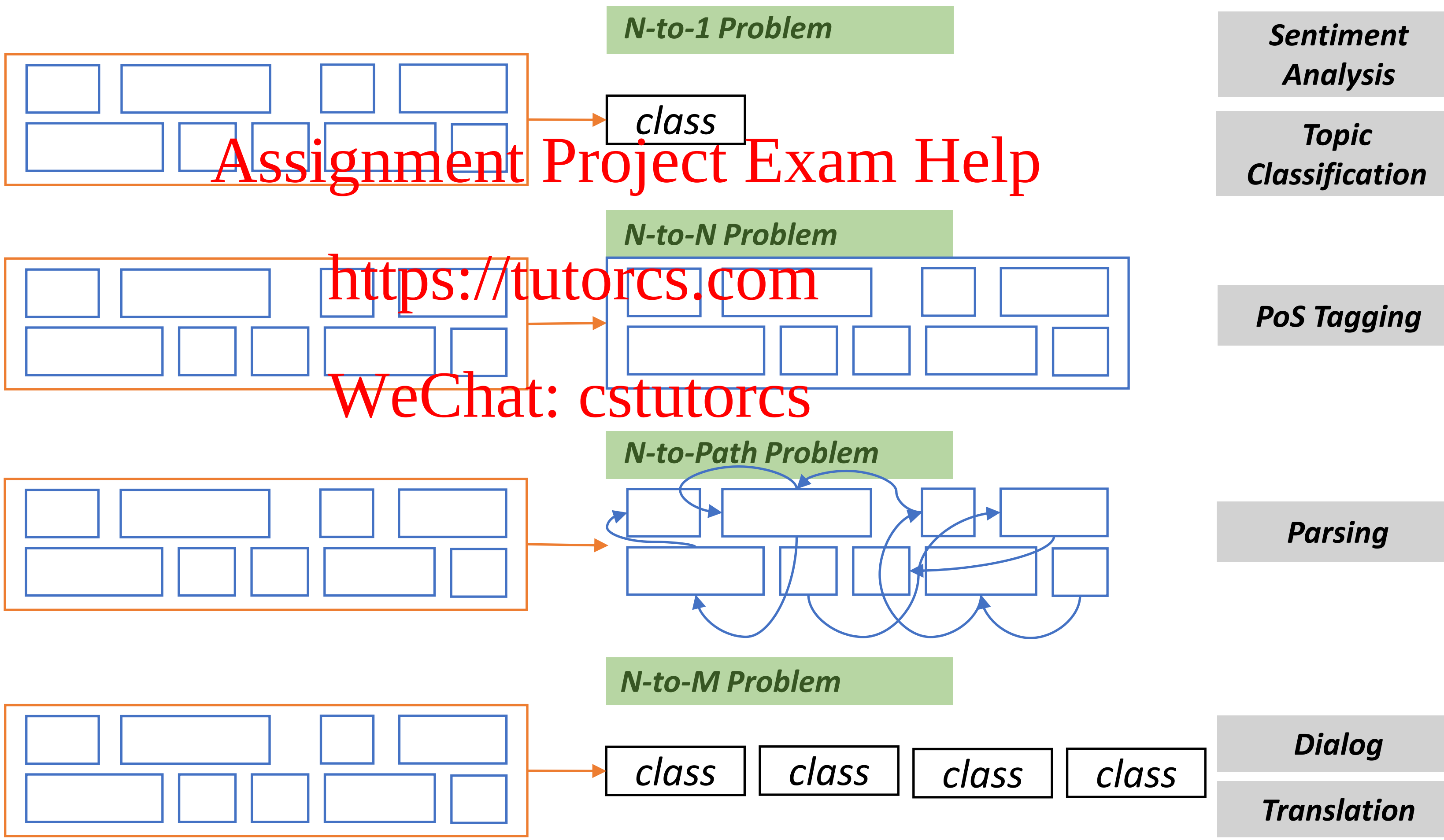
Application

Understanding			Searching
		Dialog	
Sentiment Analysis	*Topic Classification*	*Translation*	*Search*
Topic Modelling	….		….

NLP Stack

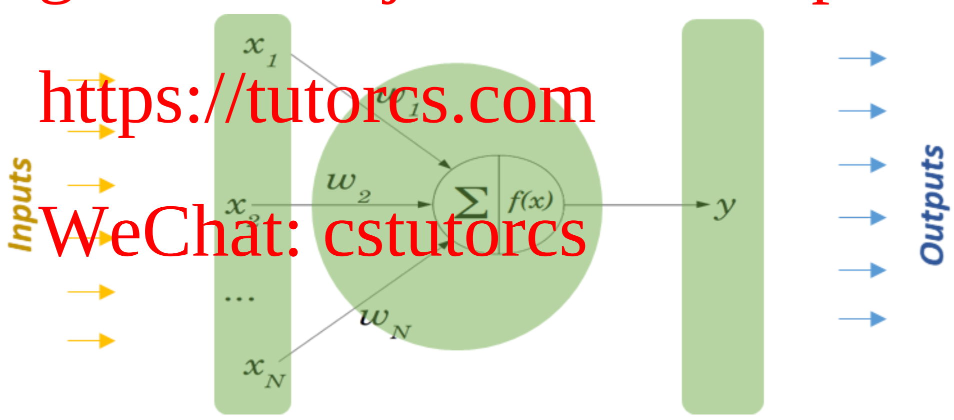
Entity Extraction	When Sebastian Thrun …	When Sebastian Thrun PERSON started at Google ORG in 2007 DATE
Parsing	Claudia sat on a stool	
PoS Tagging	She sells seashells	**[she/PRP] [sells/VBZ] [seashells/NNS]**
Stemming	Drinking, Drank, Drunk	**Drink**
Tokenisation	How is the weather today	**[How] [is] [the] [weather] [today]**

Problem Abstraction



1 Machine Learning and NLP

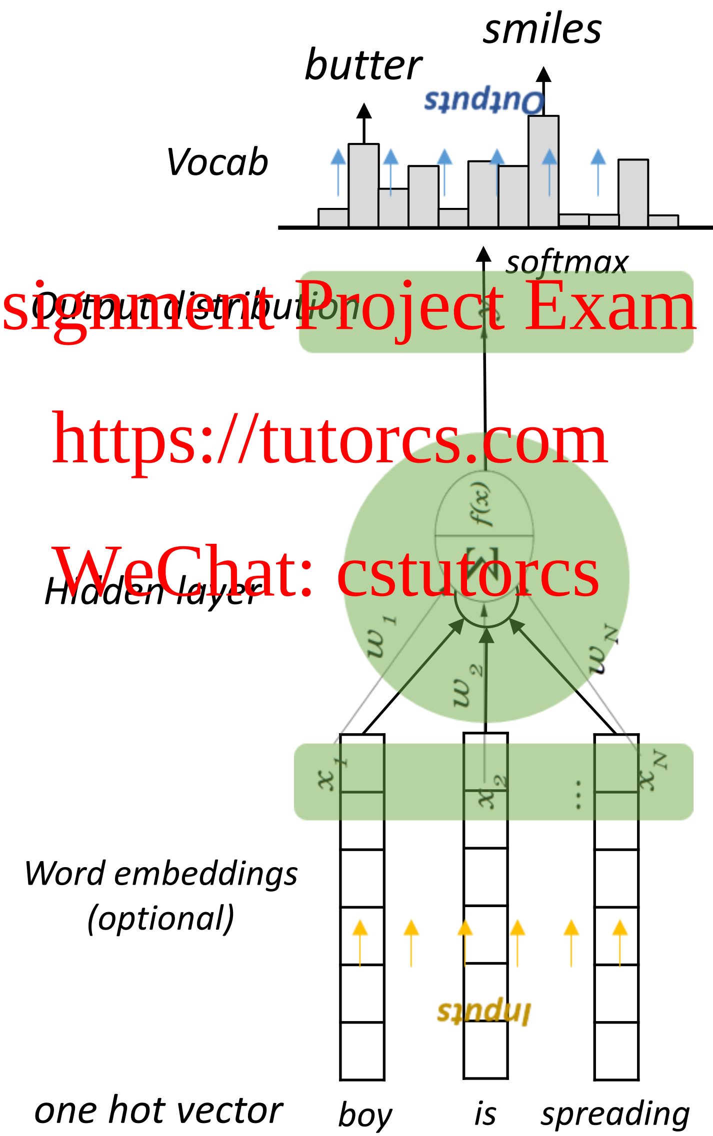
Prediction



xi	Inputs	**Features** words (indices or vectors!), context windows, sentences, documents, etc.
yi	Outputs (labels)	**What we try to predict/classify** • E.g. word meaning, sentiment, name entity

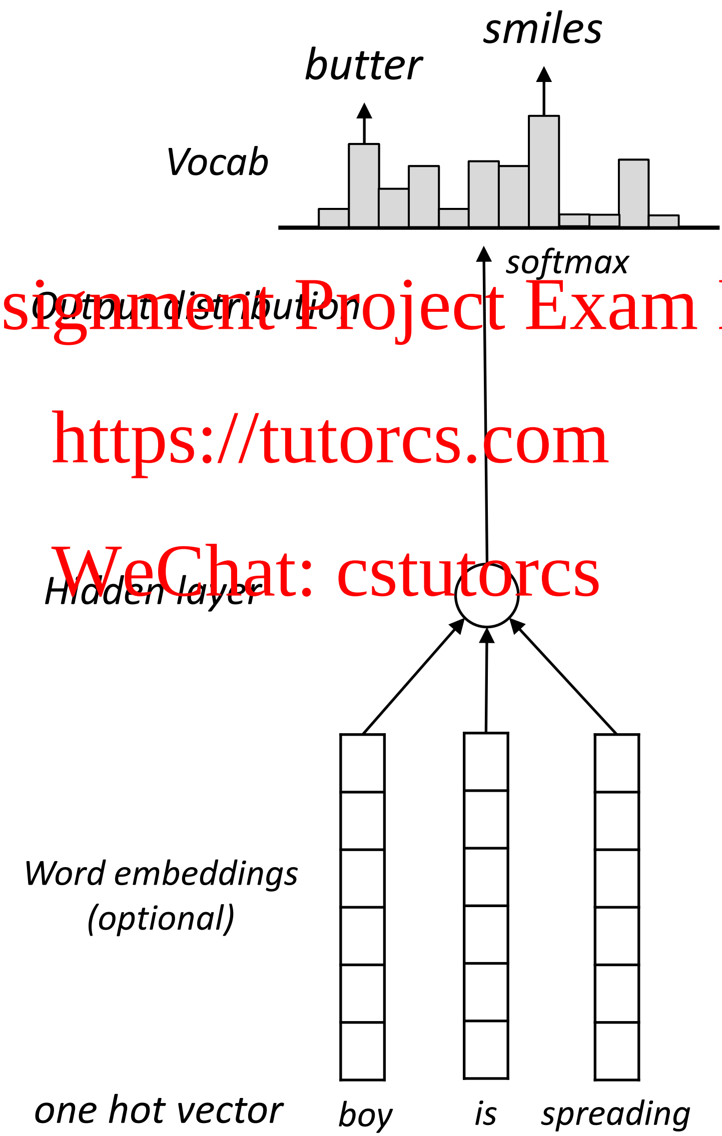
1 Machine Learning and NLP

Prediction



1 Machine Learning and NLP

Prediction



What if we consider this as a __sequential input__?
Let's add the concept 'time'

0

LECTURE PLAN

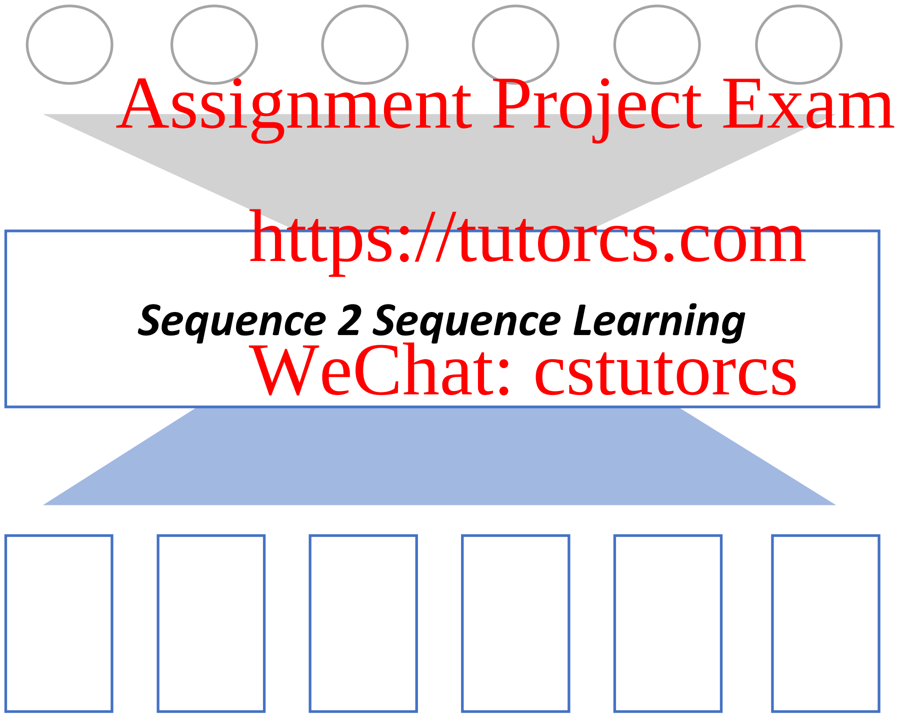
THE UNIVERSITY OF SYDNEY

Lecture 4: Word Classification and Machine Learning 2

1. Machine Learning and NLP: Finish
2. **Seq2Seq Learning**
3. Seq2Seq Deep Learning
   1. RNN (Recurrent Neural Network)
   2. LSTM (Long Short-Term Memory)
   3. GRU (Gated Recurrent Unit)
4. Data Transformation for Deep Learning NLP
5. Next Week Preview
   - Natural Language Processing Stack

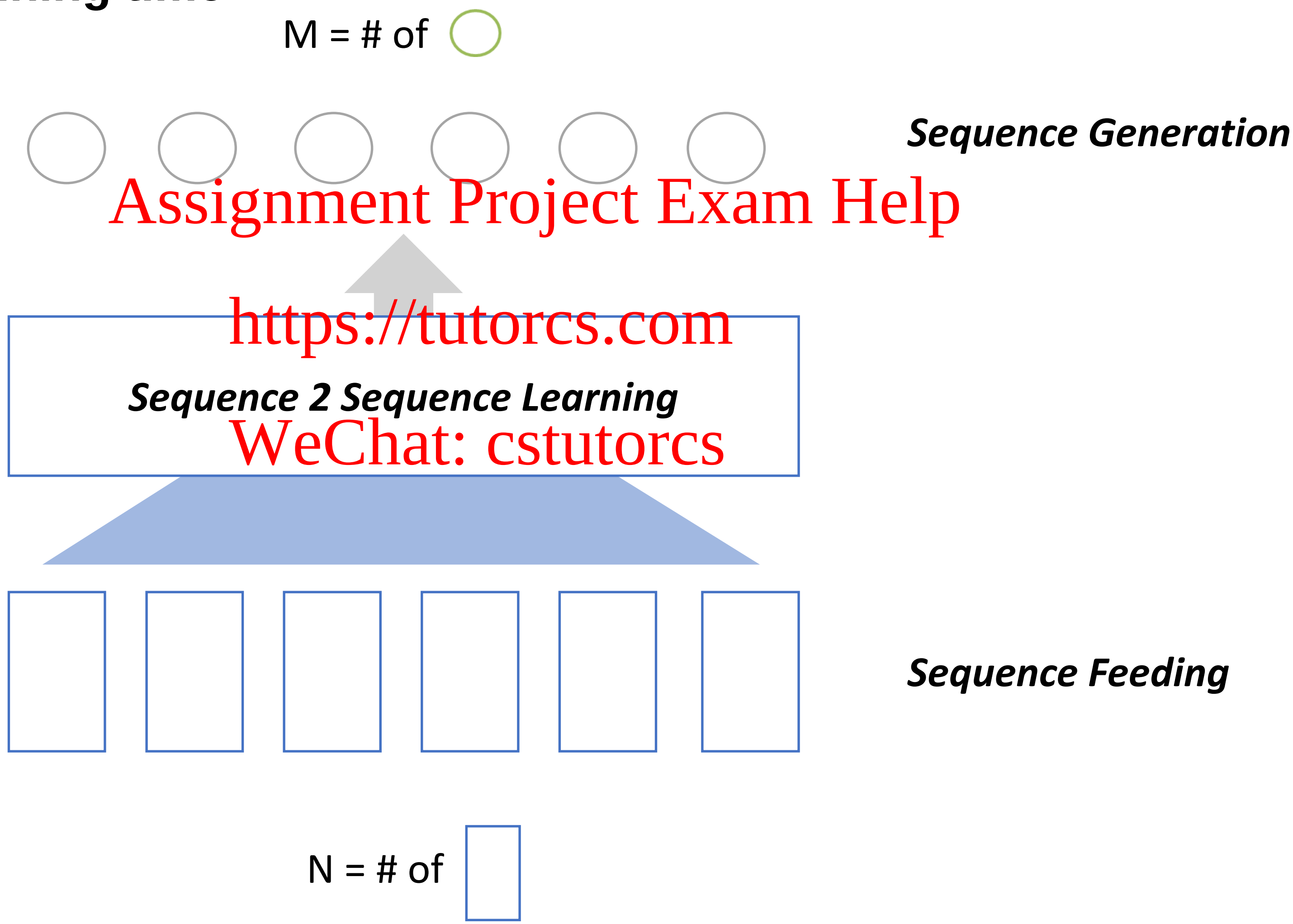
Sequence 2 Sequence Learning

Illustration

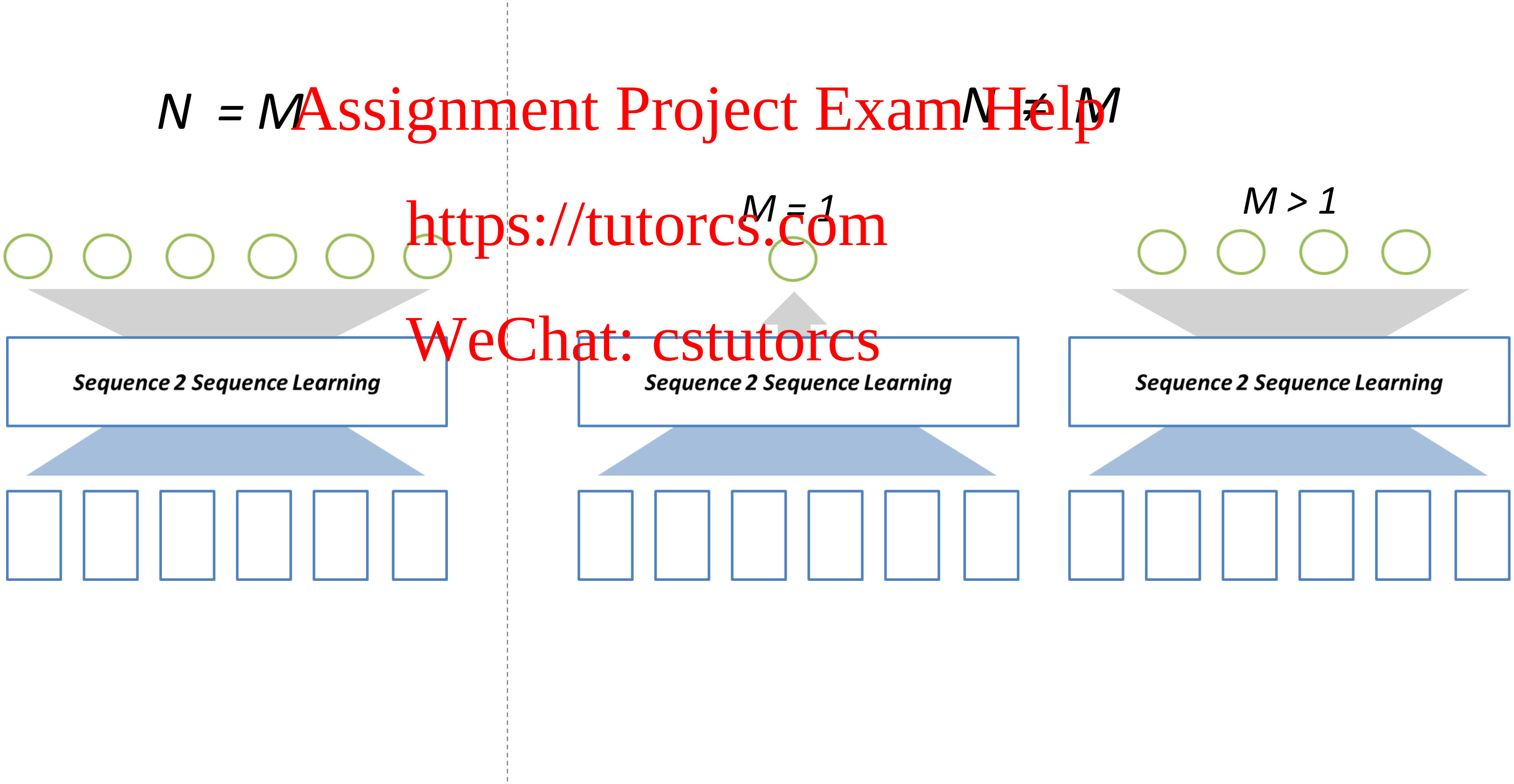


Sequence 2 Sequence Learning

Running time

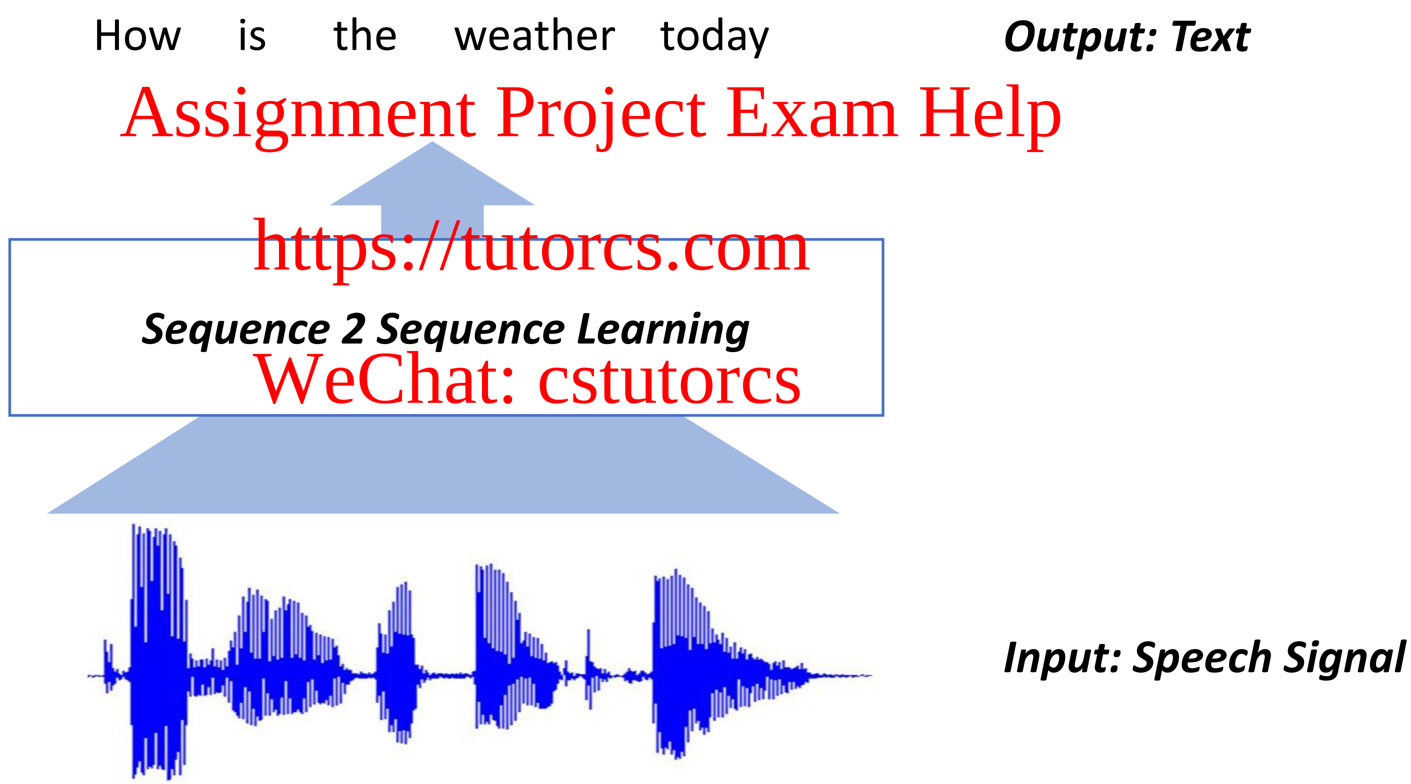


Sequence 2 Sequence Learning



Sequence 2 Sequence Learning

Seq2Seq – Speech Recognition



2 Sequence 2 Sequence Learning

Seq2Seq – Movie Frame Labelling

Swing Swing Hit Bat_Broken

Output: Scene Labels

Sequence 2 Sequence Learning

Input: Video Frame

2 Sequence 2 Sequence Learning

Seq2Seq – PoS Tagging

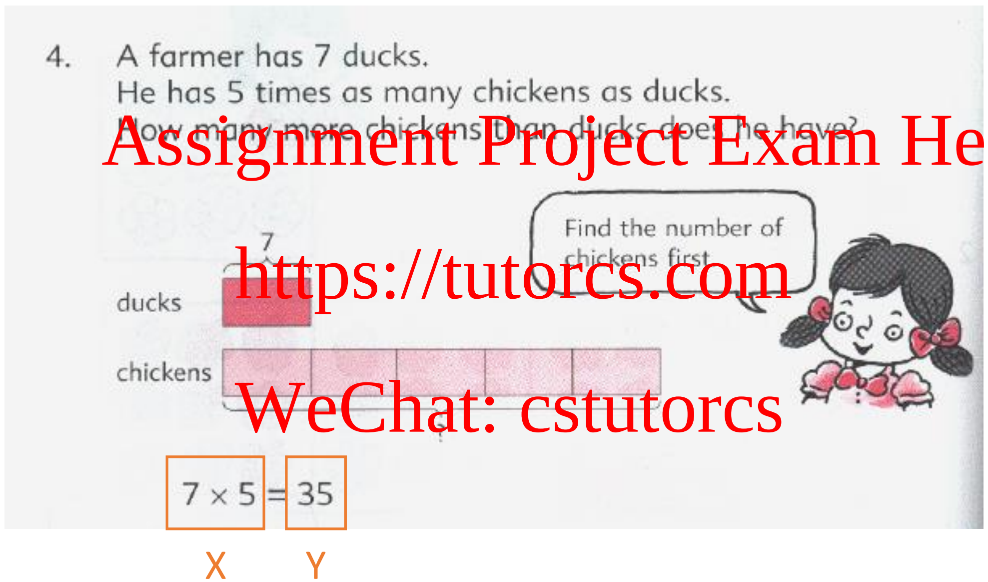
ADV VERB DET NOUN NOUN ***Output: Part of Speech***

Sequence 2 Sequence Learning

How is the weather today ***Input: Text***

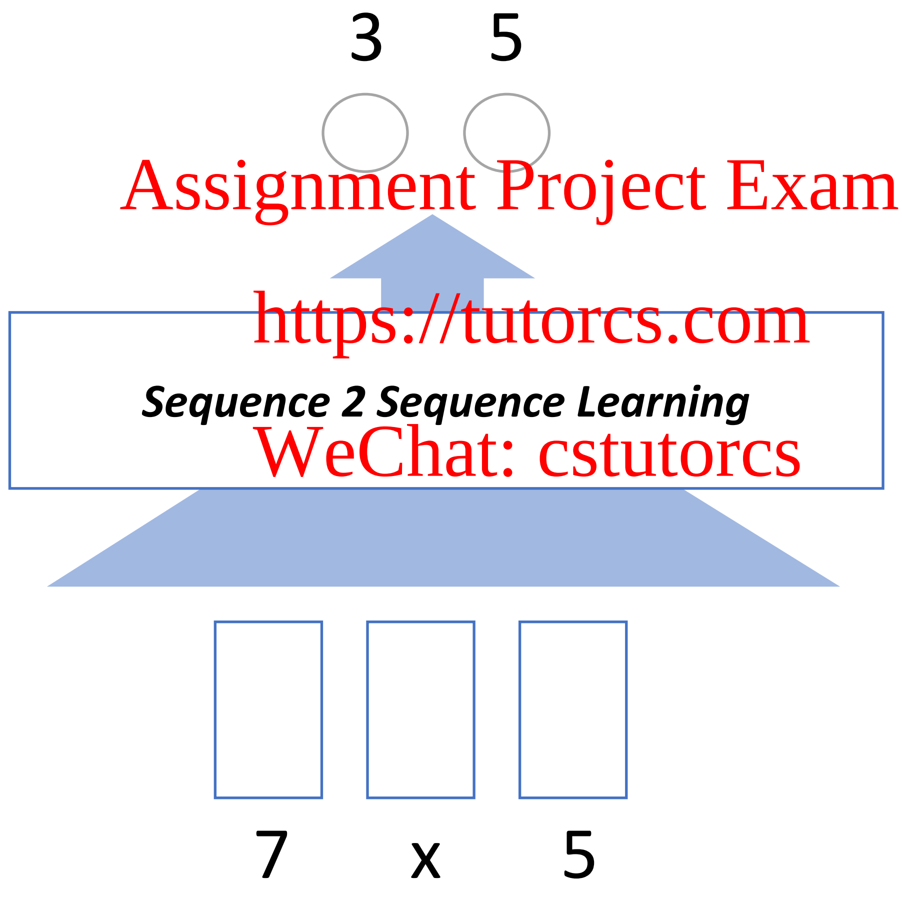
Sequence 2 Sequence Learning

Seq2Seq – Arithmetic Calculation



2 Sequence 2 Sequence Learning

Seq2Seq – Arithmetic Calculation



Output: Numbers

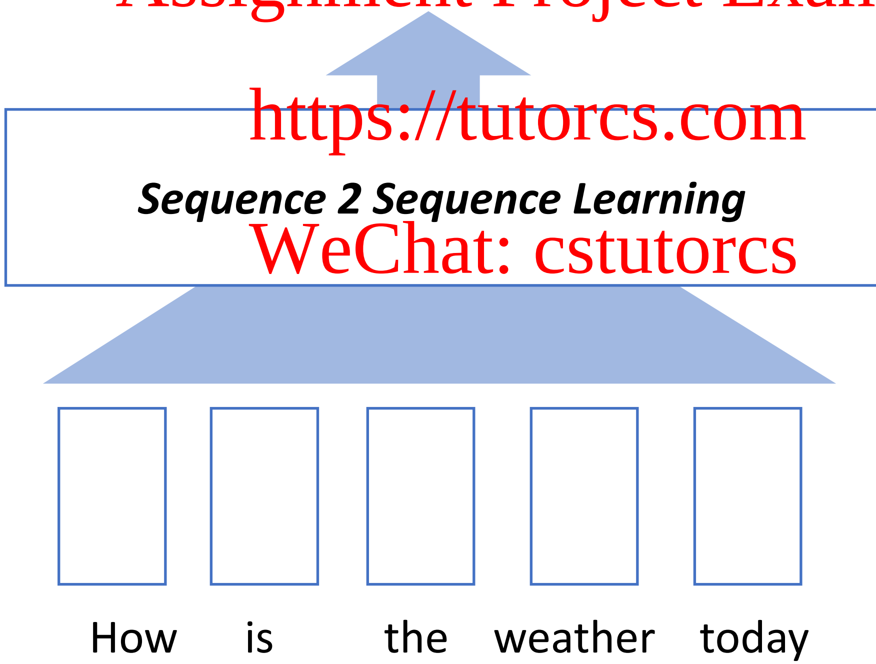
Input: Math Expression

2 Sequence 2 Sequence Learning

Seq2Seq – Machine Translation

今天 天气 怎么 样?

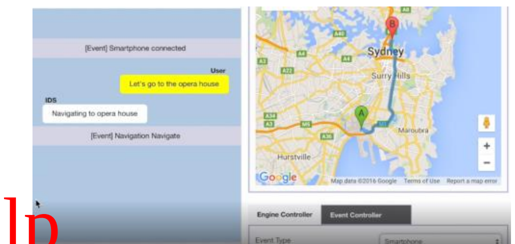
Output: Chinese Text



Input: English Text

2 Sequence 2 Sequence Learning

Seq2Seq – Sentence Completion



How is the weather today?

How long does it take?

Let's go to the opera house

It is quite hot inside

I may need to stop by Darling Harbour

When is the dinner appointment

Change the schedule

Text him that I cannot meet at 6:30pm

I like learning Natural Language Processing

2 Sequence 2 Sequence Learning

Seq2Seq – Sentence Completion

How is the weather today?

How long does it take?

Let's go to the opera house

It is quite hot inside

I may need to stop by Darling Harbour

When is the dinner appointment

Change the schedule

Text him that I cannot meet at 6:30pm

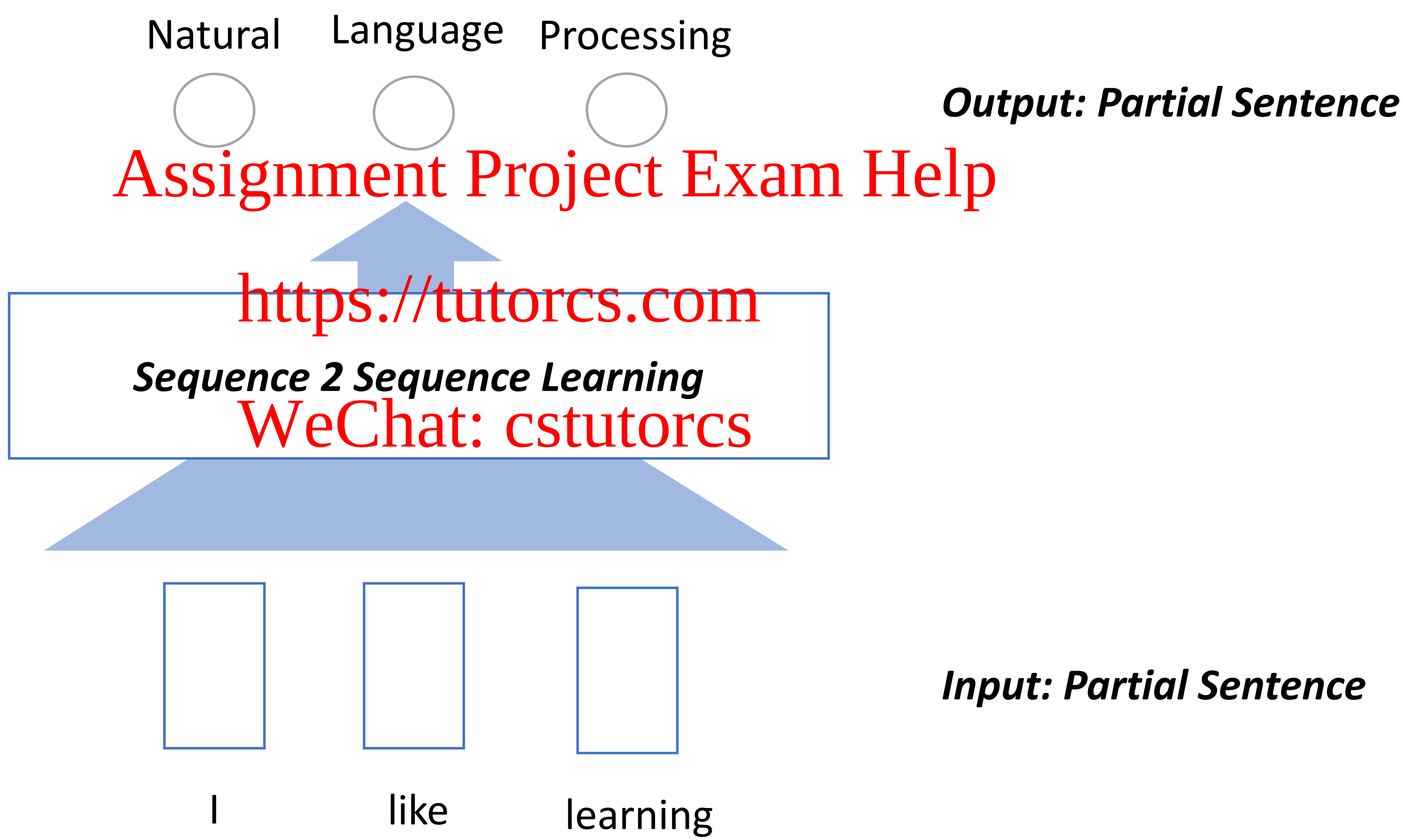
I like learning Natural Language Processing

X *Y*

Sequence 2 Sequence Learning

I like learning Natural Language Processing

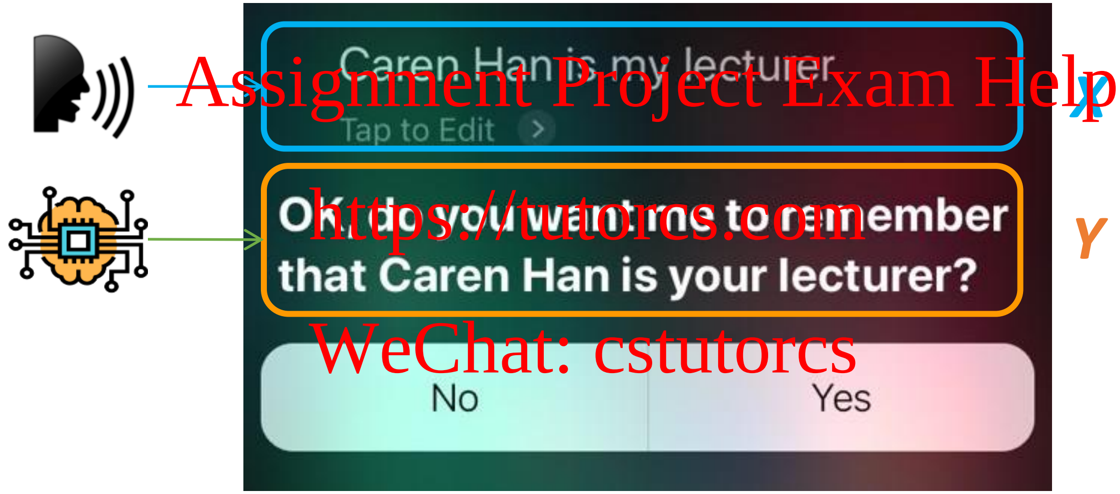
Seq2Seq – Sentence Completion



2 Sequence 2 Sequence Learning

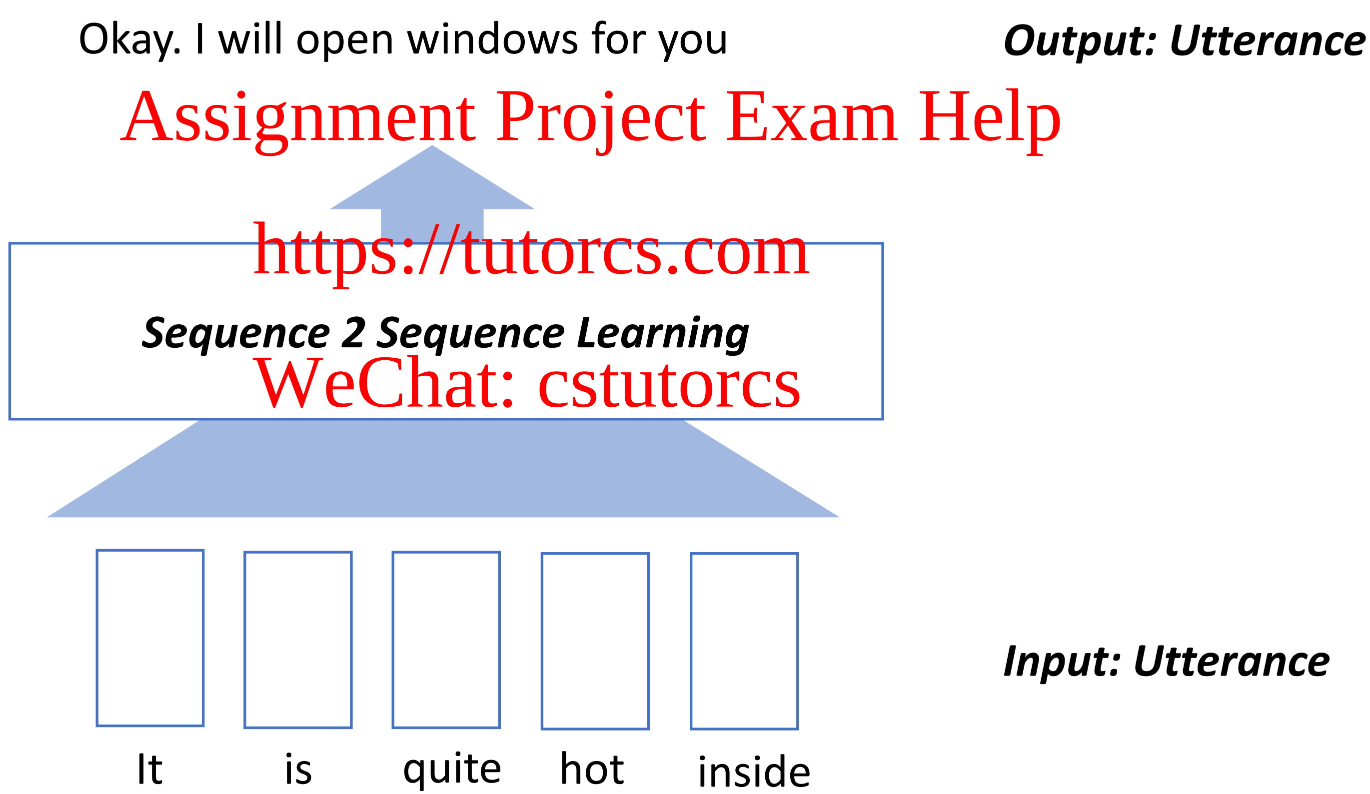
Seq2Seq – Conversation Modelling

Conversation



2 Sequence 2 Sequence Learning

Seq2Seq – Conversation Modelling



LECTURE PLAN

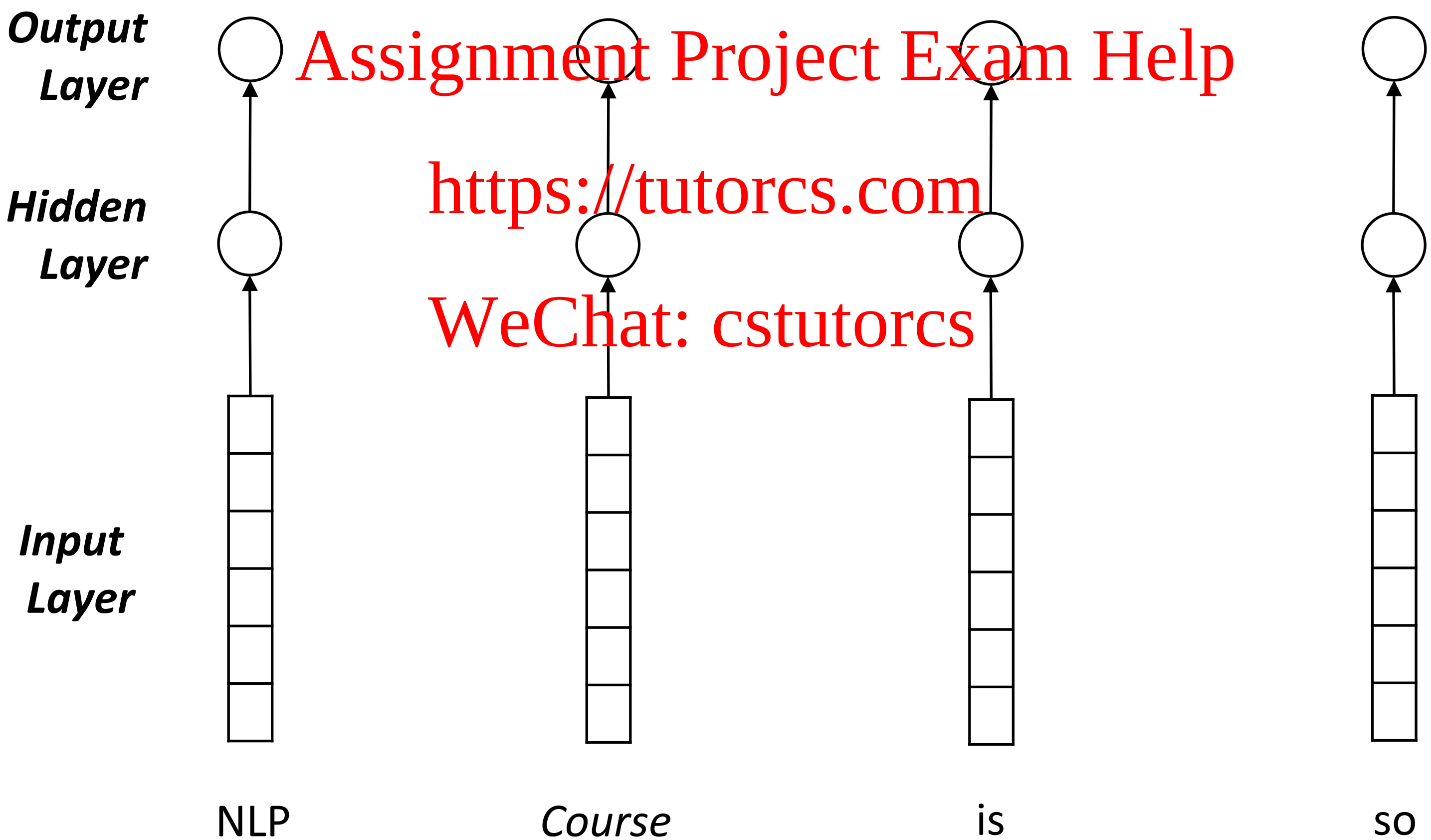
THE UNIVERSITY OF SYDNEY

Lecture 4: Word Classification and Machine Learning 2

1. Machine Learning and NLP: Finish
2. Seq2Seq Learning
3. **Seq2Seq Deep Learning**
   1. RNN (Recurrent Neural Network)
   2. LSTM (Long Short-Term Memory)
   3. GRU (Gated Recurrent Unit)
4. Data Transformation for Deep Learning NLP
5. Next Week Preview
   - Natural Language Processing Stack

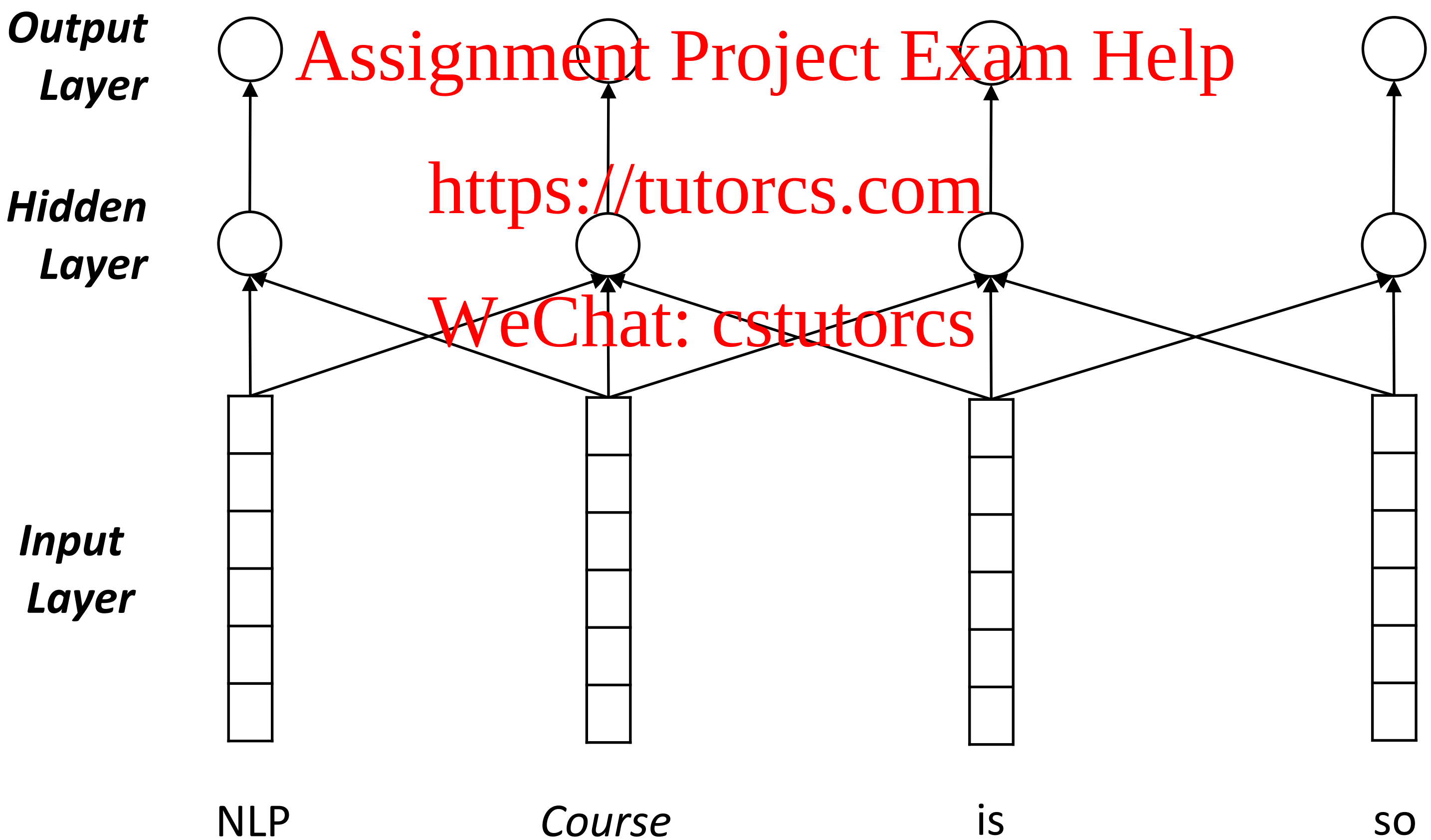
3 Seq2Seq with Deep Learning

Prediction



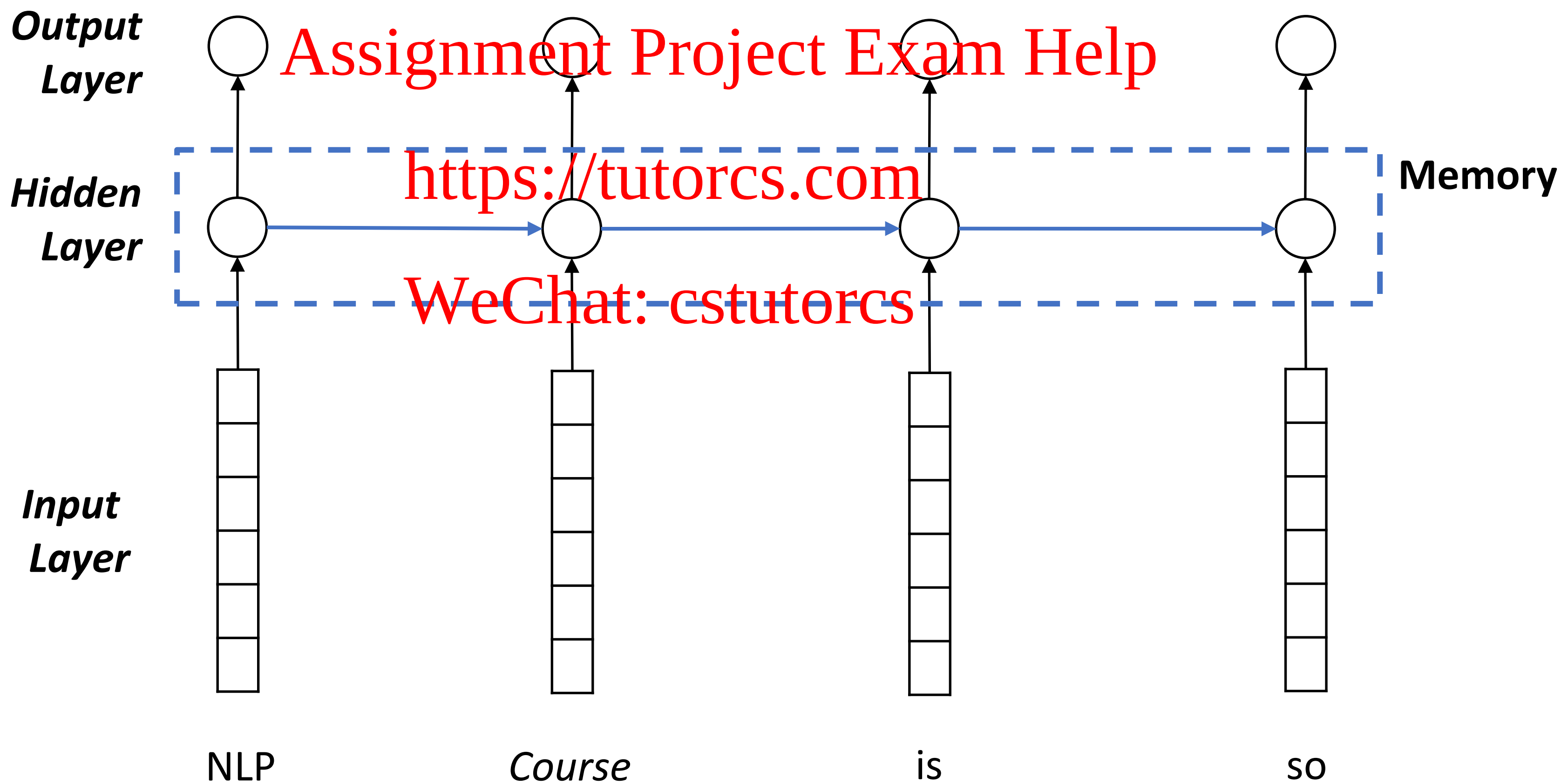
3 Seq2Seq with Deep Learning

Prediction + Convolution Idea



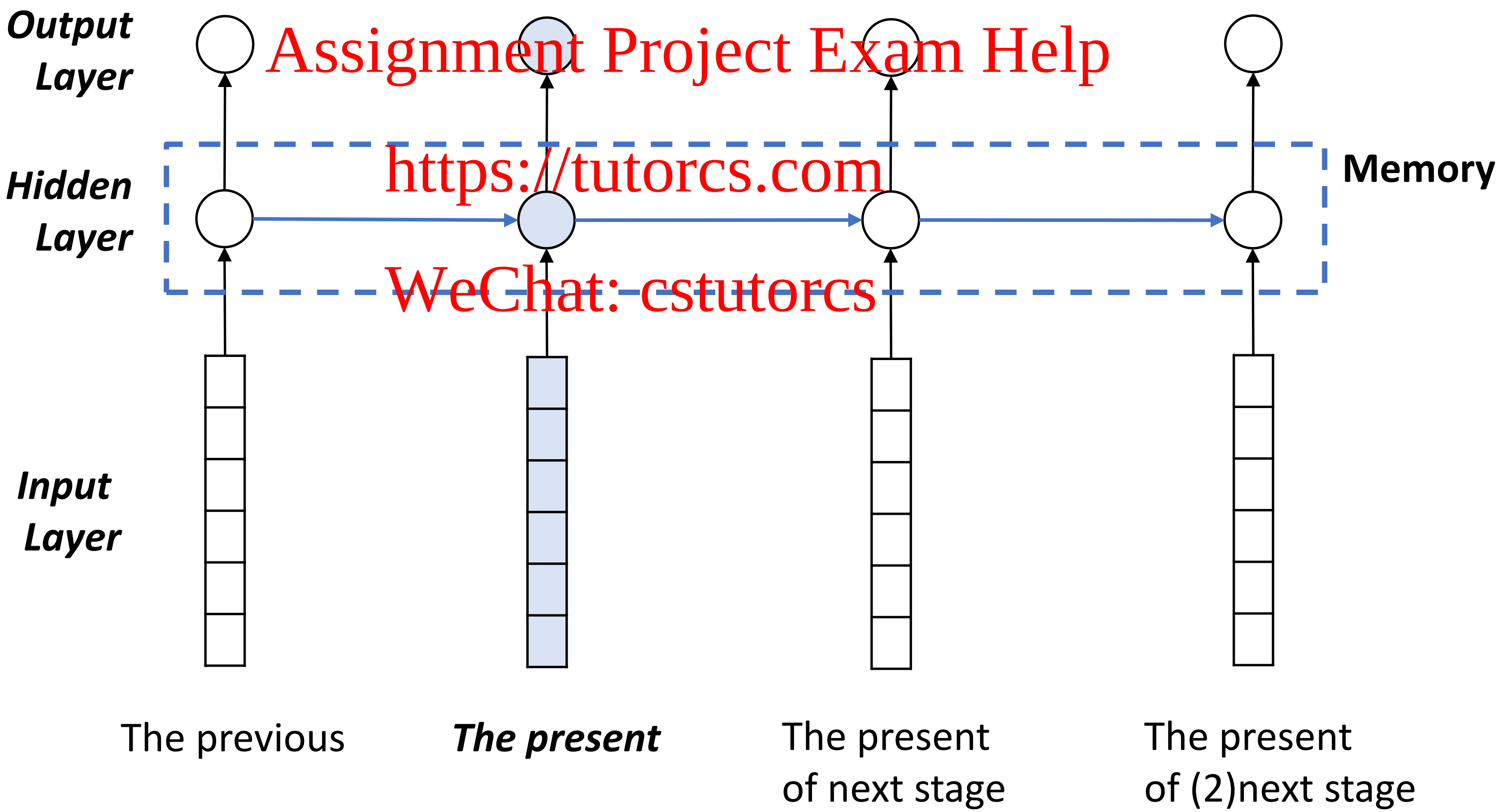
Seq2Seq with Deep Learning

Prediction + Memory = Sequence Modelling



3 Seq2Seq with Deep Learning

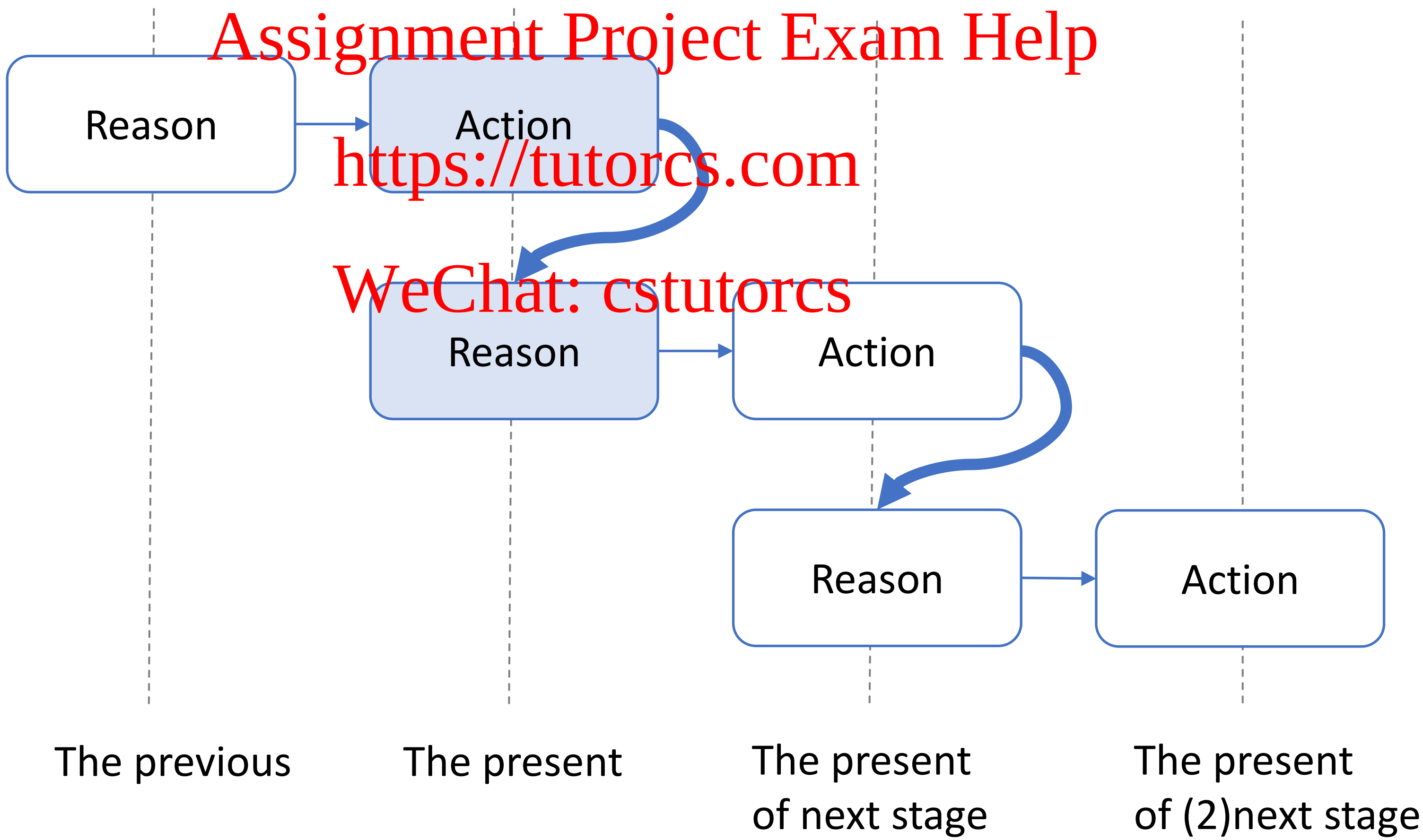
Prediction + Memory = Sequence Modelling



3 Seq2Seq with Deep Learning

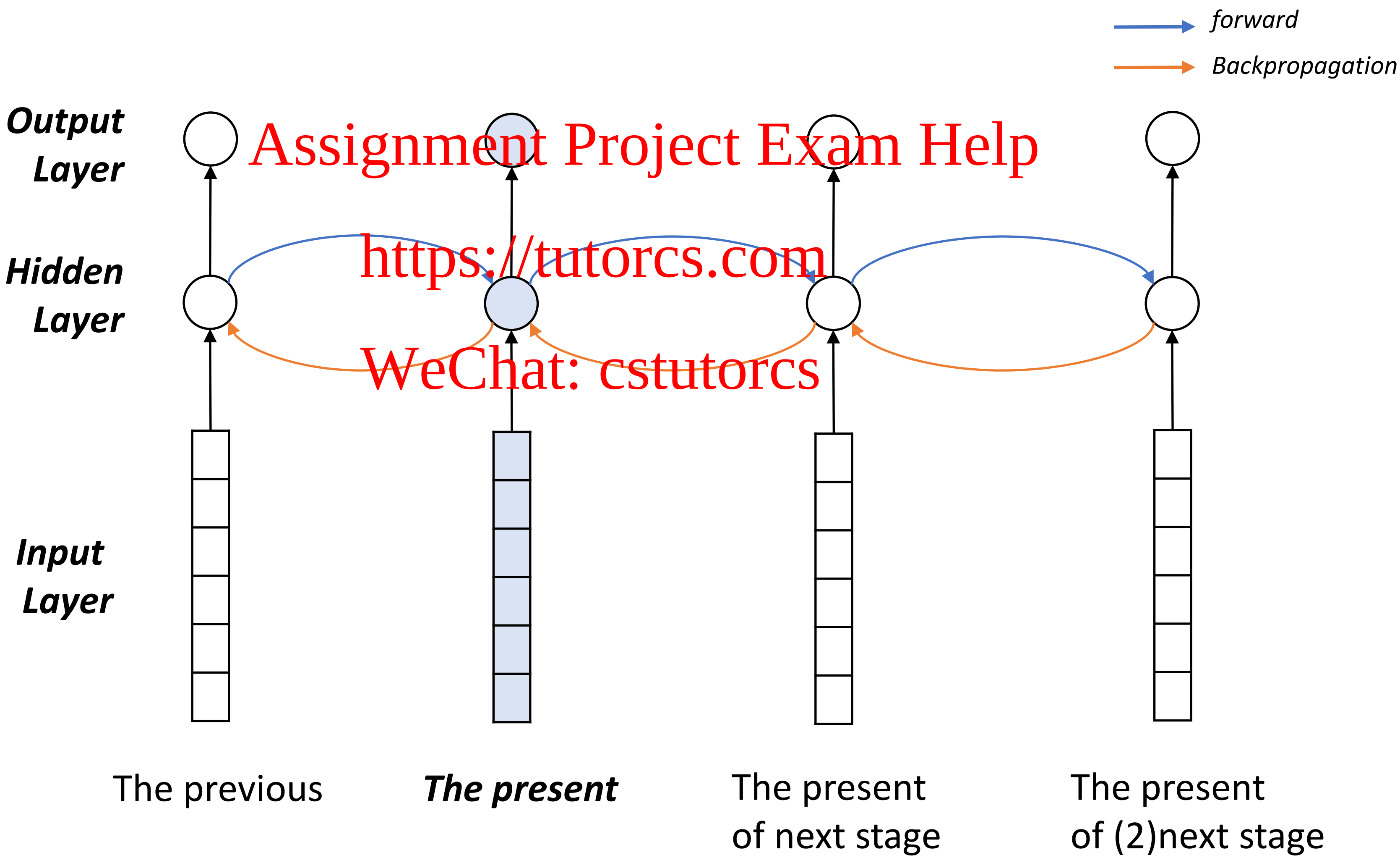
Neural Network + Memory

Memory is vital to experiences, it is the retention of information over time for the purpose of influencing future action

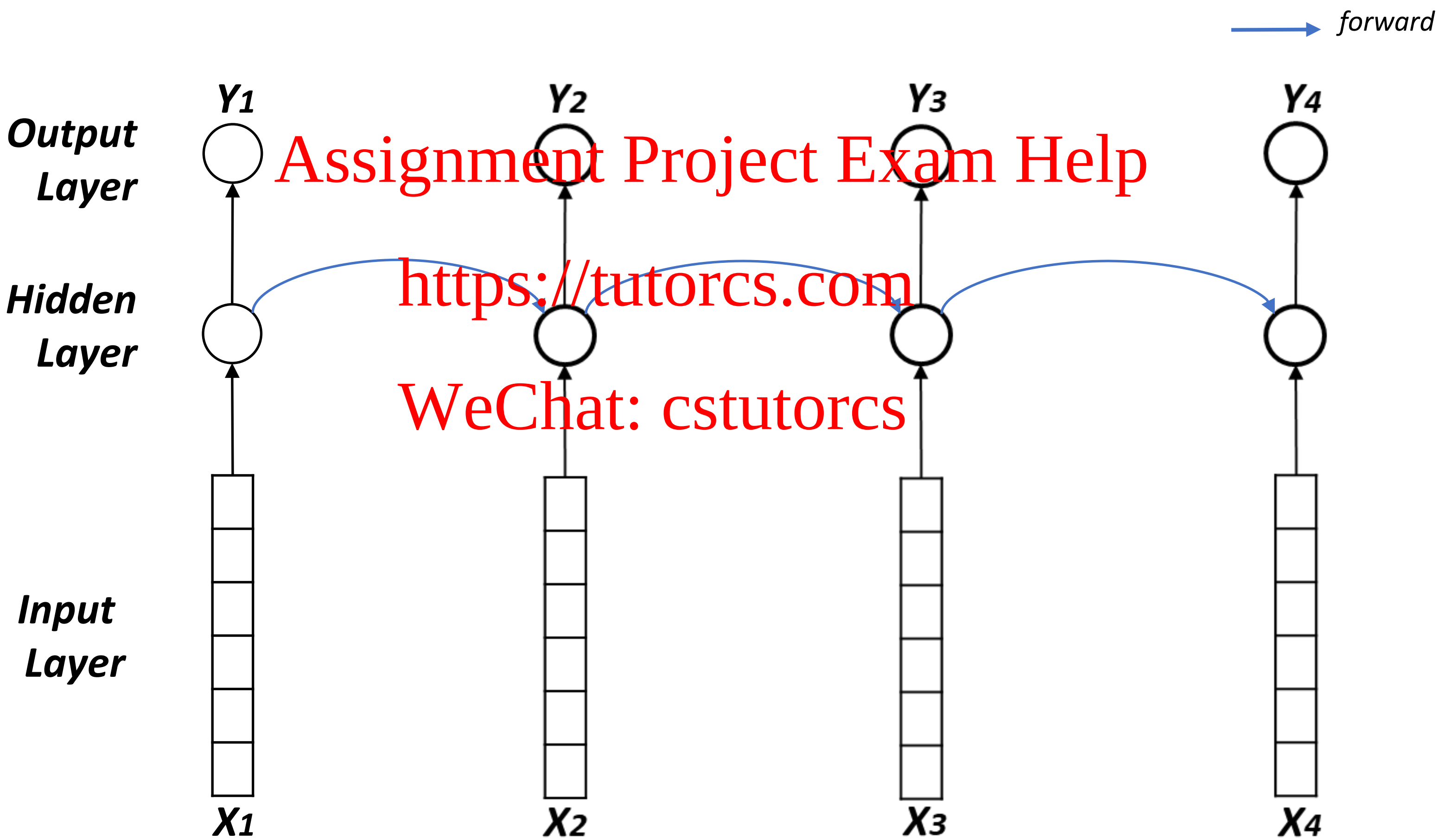


Seq2Seq with Deep Learning

Neural Network + Memory

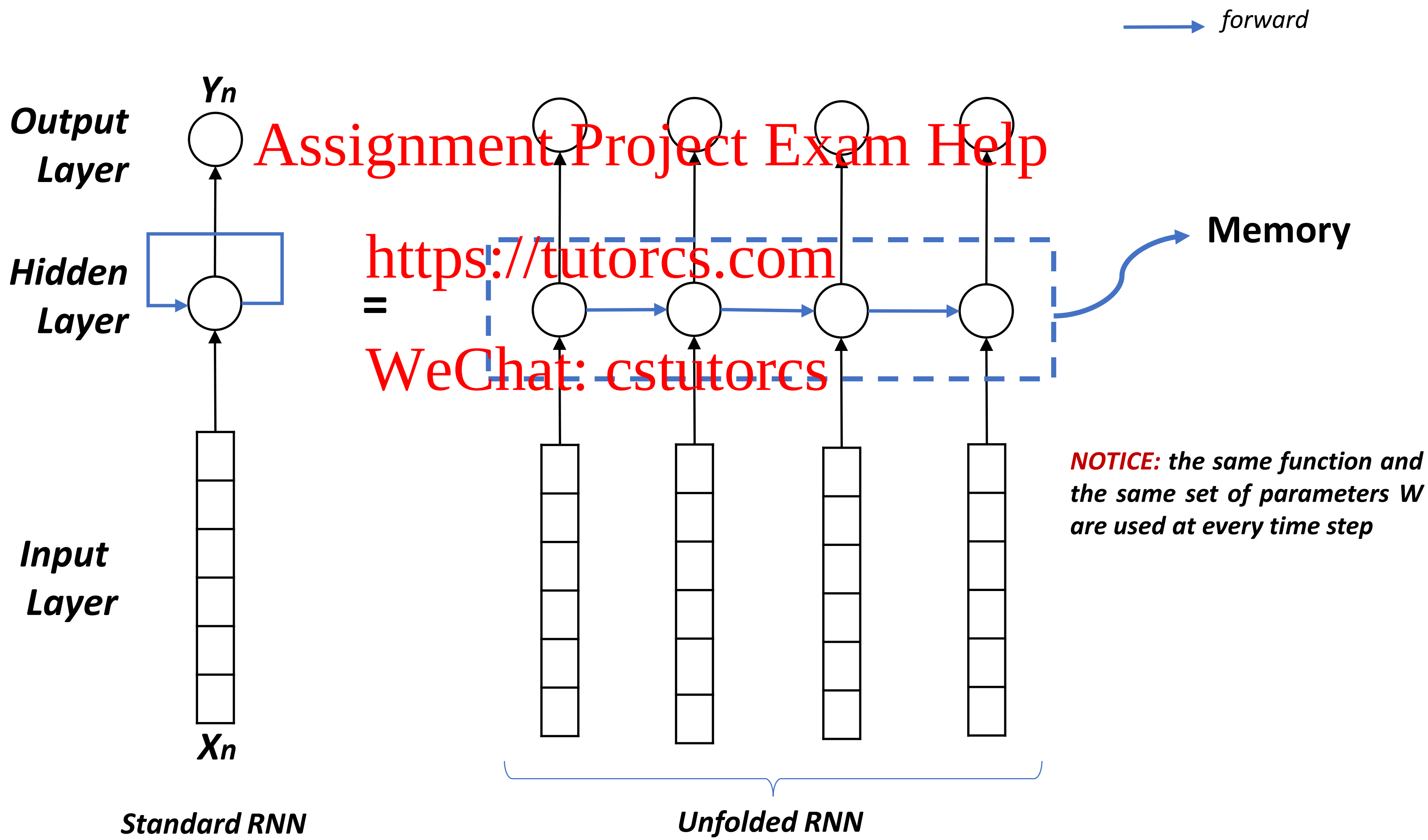


Neural Network + Memory = Recurrent Neural Network



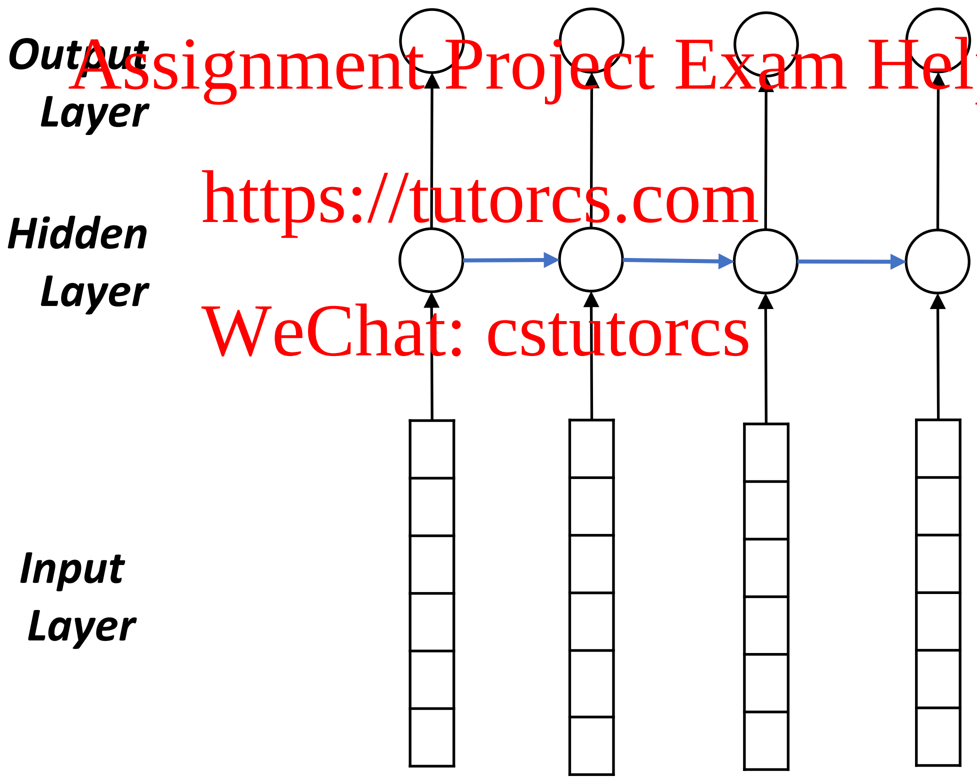
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network (RNN)



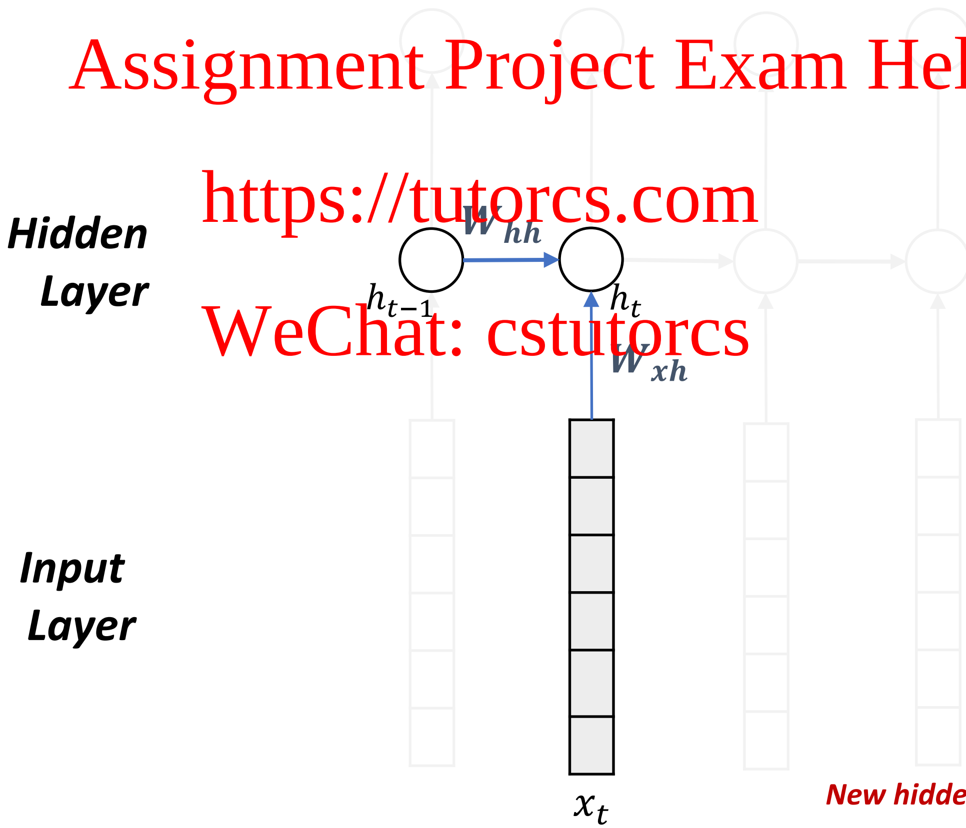
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network (RNN)



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



Previous state *input*

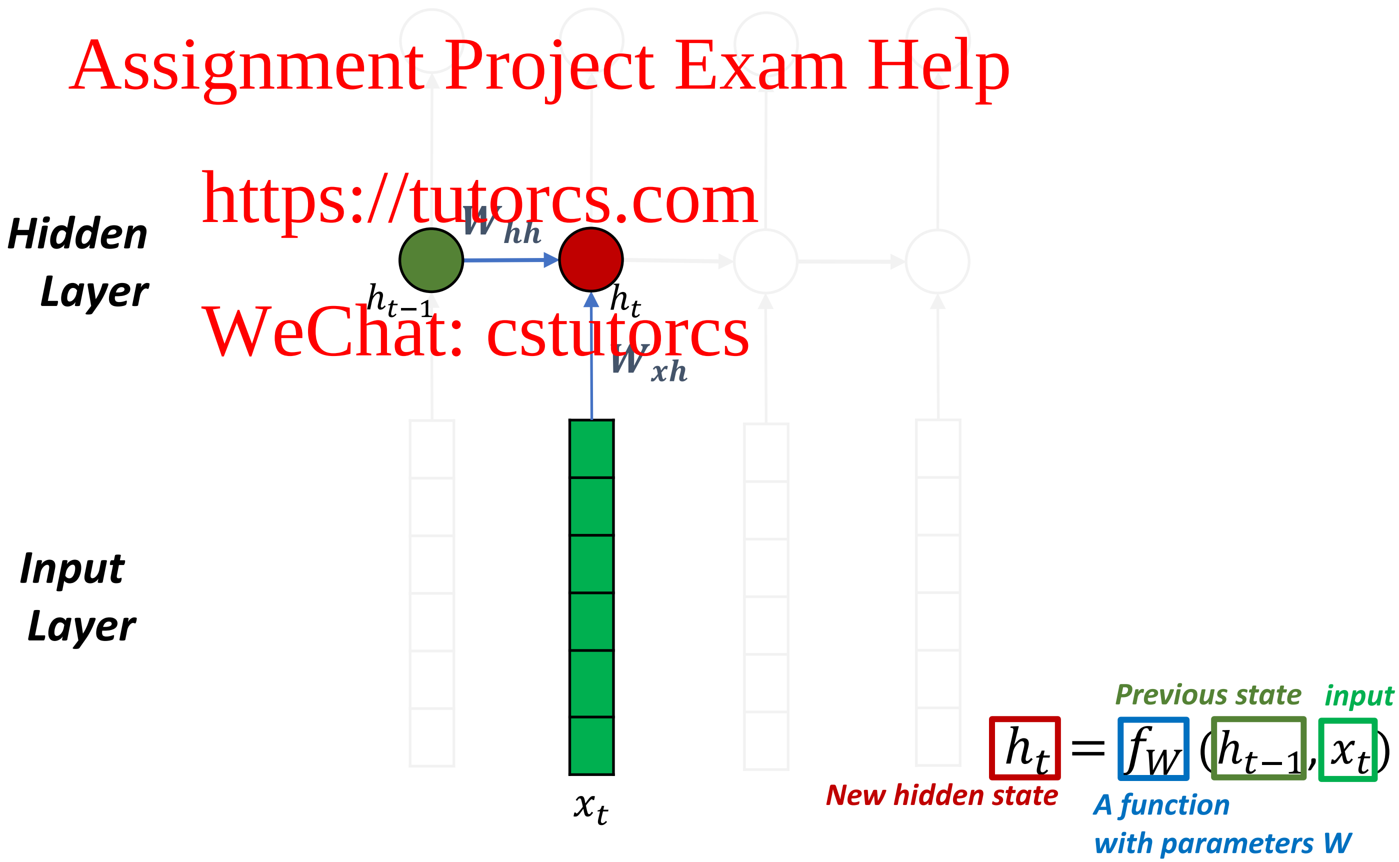
$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

New hidden state

A function with parameters W

Seq2Seq with Deep Learning

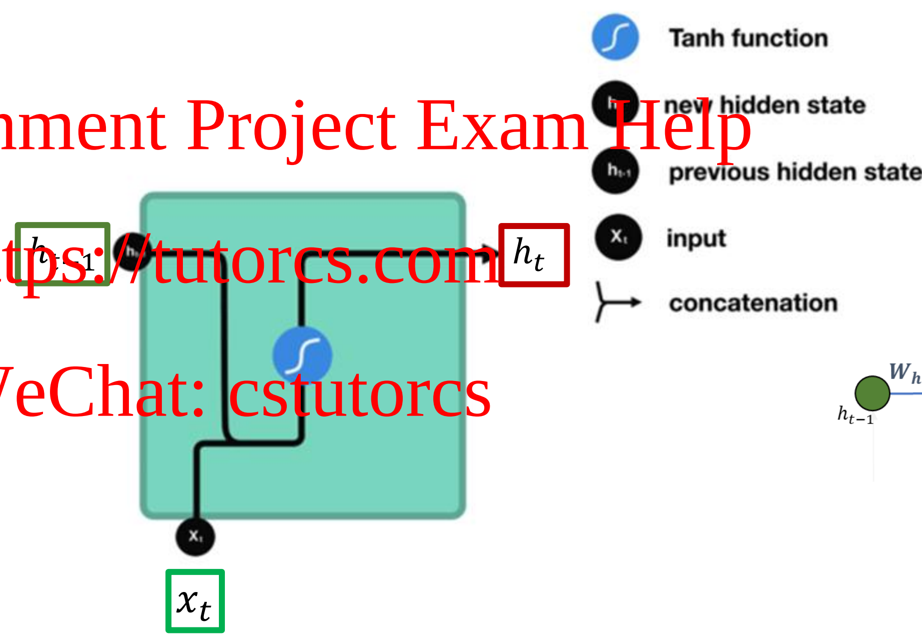
Neural Network + Memory = Recurrent Neural Network



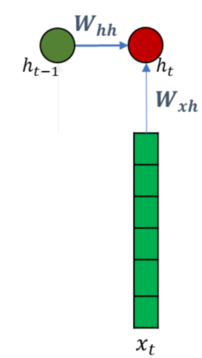
3 **Seq2Seq with Deep Learning**

Neural Network + Memory = Recurrent Neural Network

Hidden Layer



Input Layer



Previous state *input*

$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

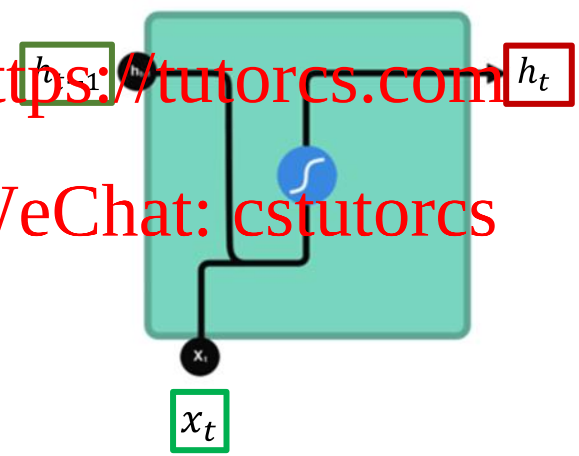
New hidden state *A function with parameters W*

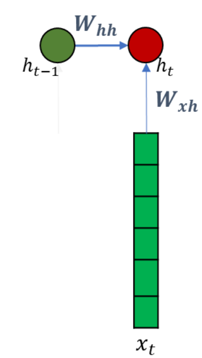
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Hidden Layer

Input Layer





$$\boxed{h_t} = \boxed{tanh}(\underline{W_{hh}}\,\underline{h_{t-1}} + \underline{W_{xh}}\,\underline{x_t} + \underline{b_h})$$

New hidden state

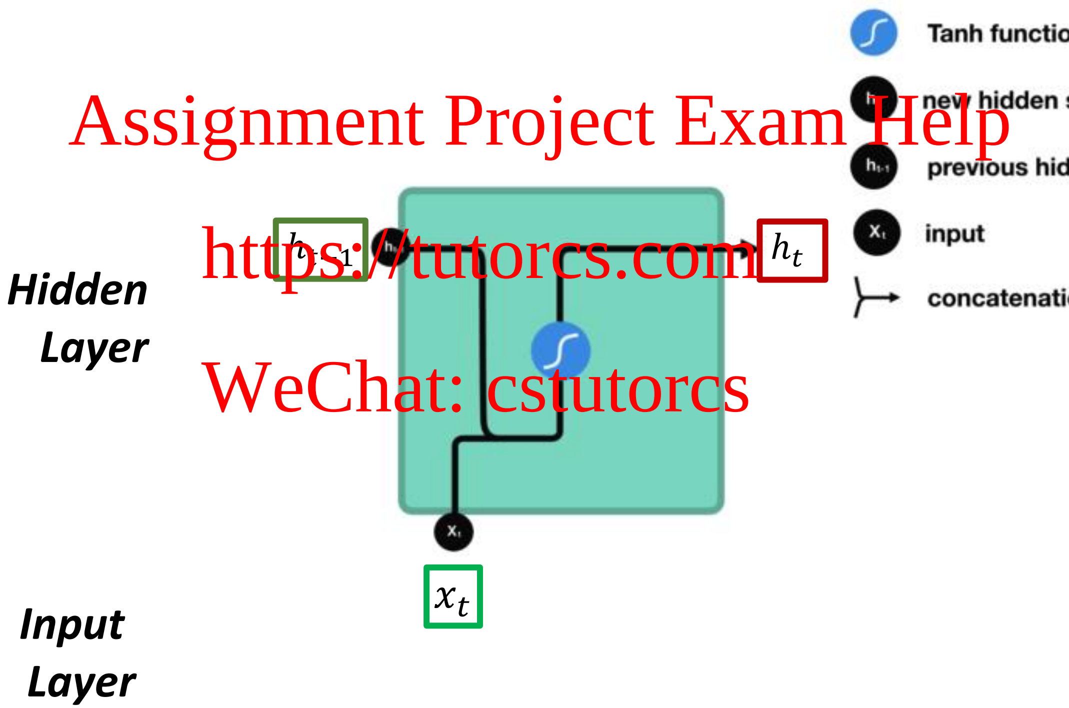
A function with parameters W

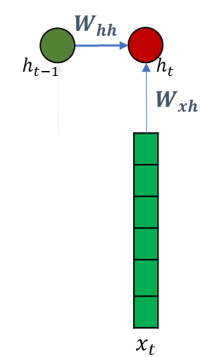
Previous state

input

3 Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



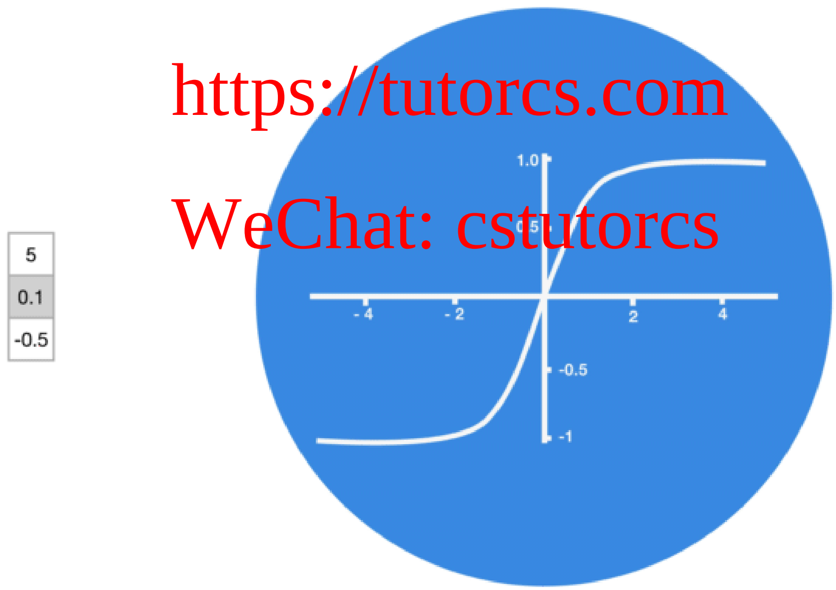


$$\underbrace{\boxed{h_t}}_{\text{New hidden state}} = \underbrace{\boxed{tanh}}_{\text{A function with parameters } W}(W_{hh}\underbrace{h_{t-1}}_{\text{Previous state}} + W_{xh}\underbrace{x_t}_{\text{input}} + b_h)$$

3 Seq2Seq with Deep Learning

Tanh activation

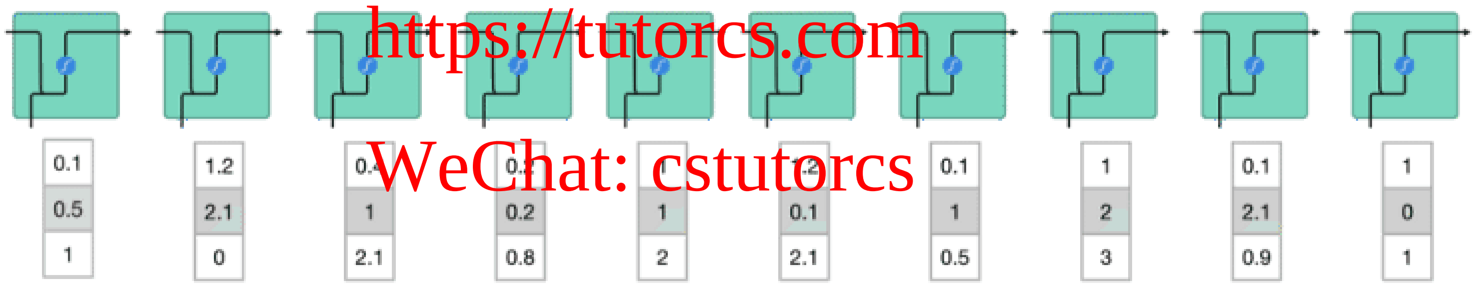
The tanh activation is used to help regulate the values flowing through the network. The tanh function squishes values to always be between -1 and 1.



3 Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

With Sequence Input

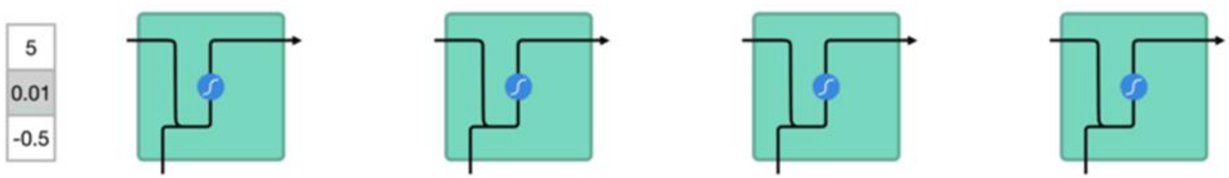


3 Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Q: Why do we need tanh function?

Vector Transformations without tanh



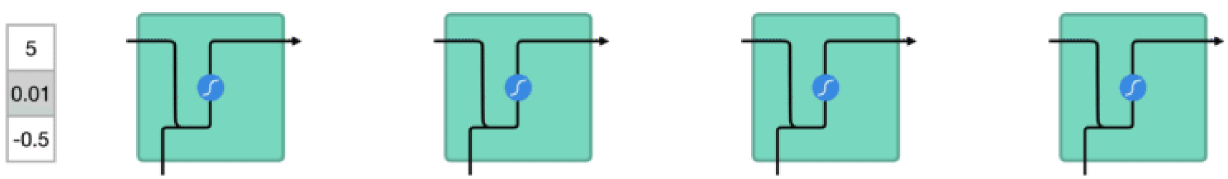
Vector Transformations with tanh

3 Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Q: Why do we need tanh function?

Vector Transformations without tanh

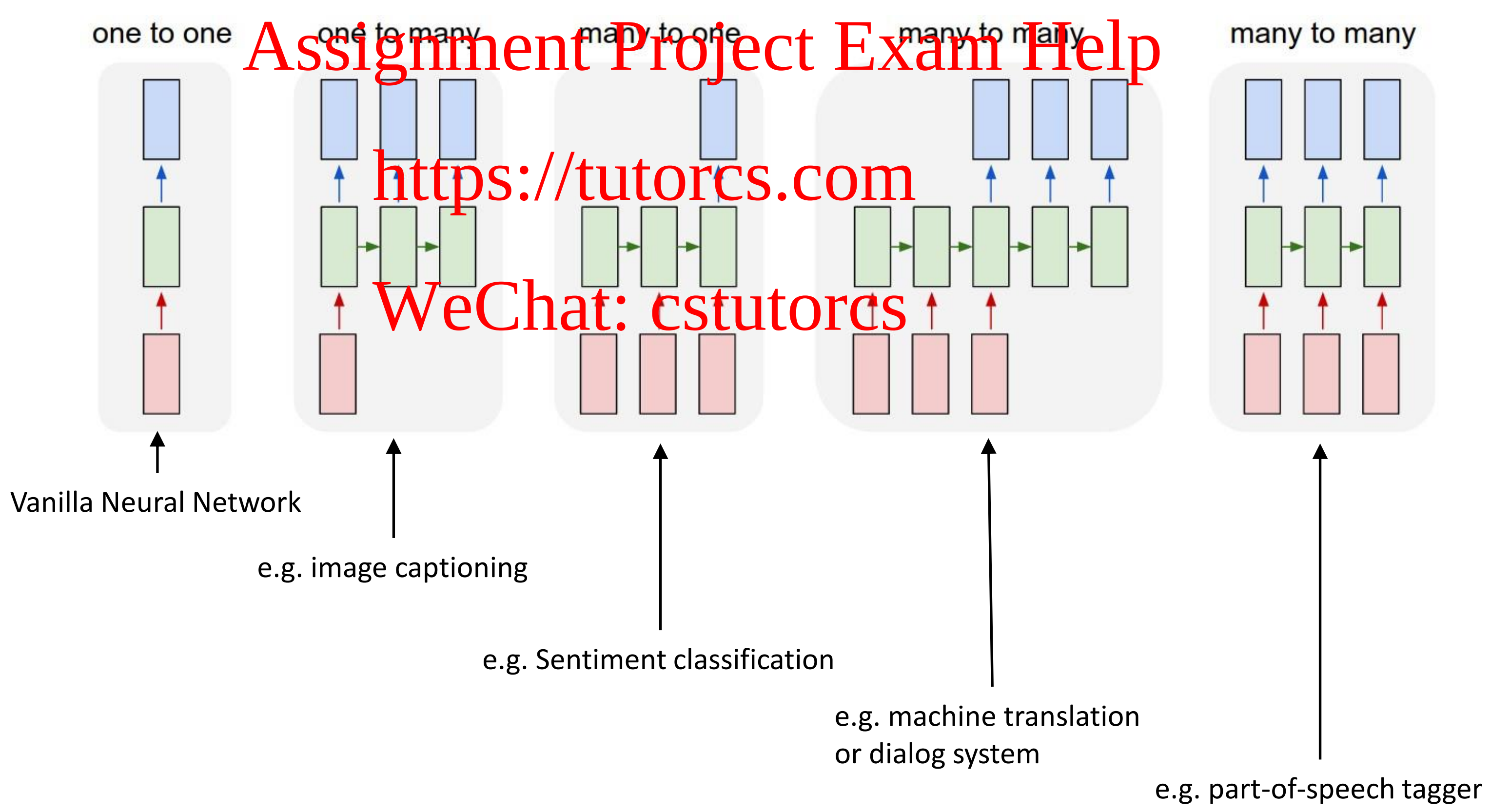


Vector Transformations with tanh

3 Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

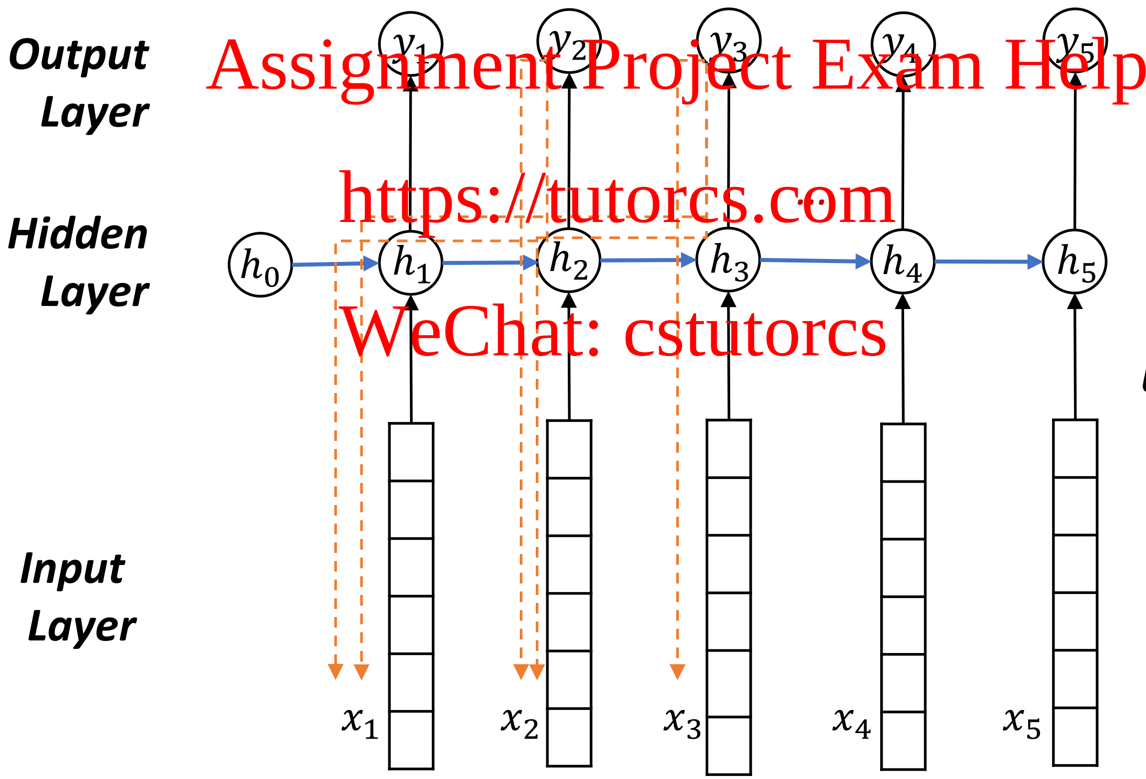
Several Variants of RNN



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Backpropagation through time



What about ***very long sequence?***

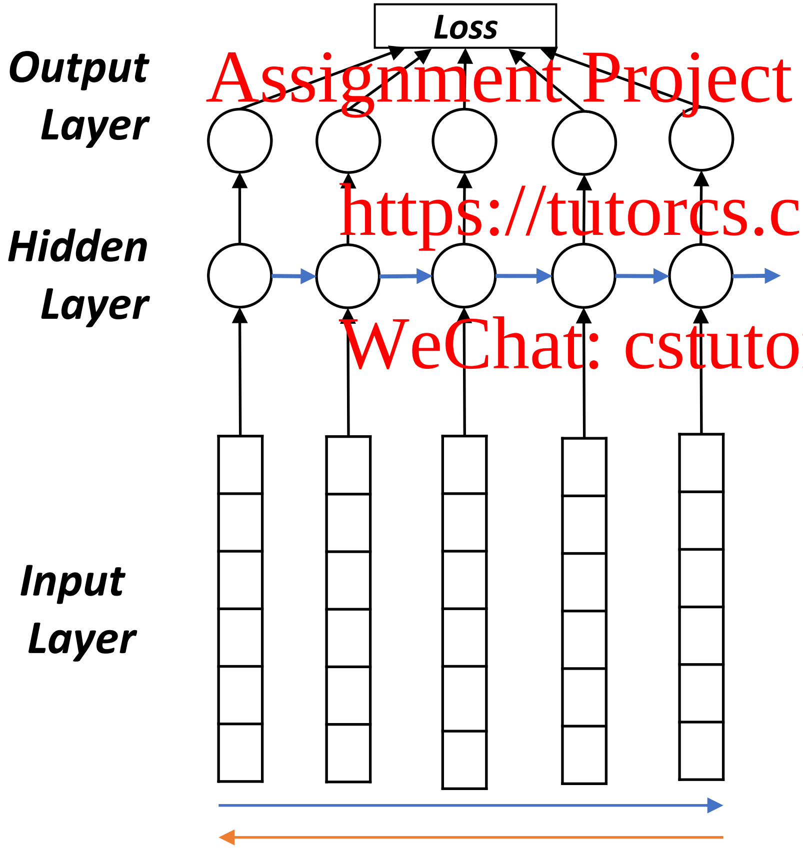
Use ***Truncate Backpropagation*** *through time*

- *Similar as* ***standard backpropagation*** *on unrolled network*
- *Similar as* ***training very deep networks*** *with tied parameters*

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Truncated Backpropagation through time

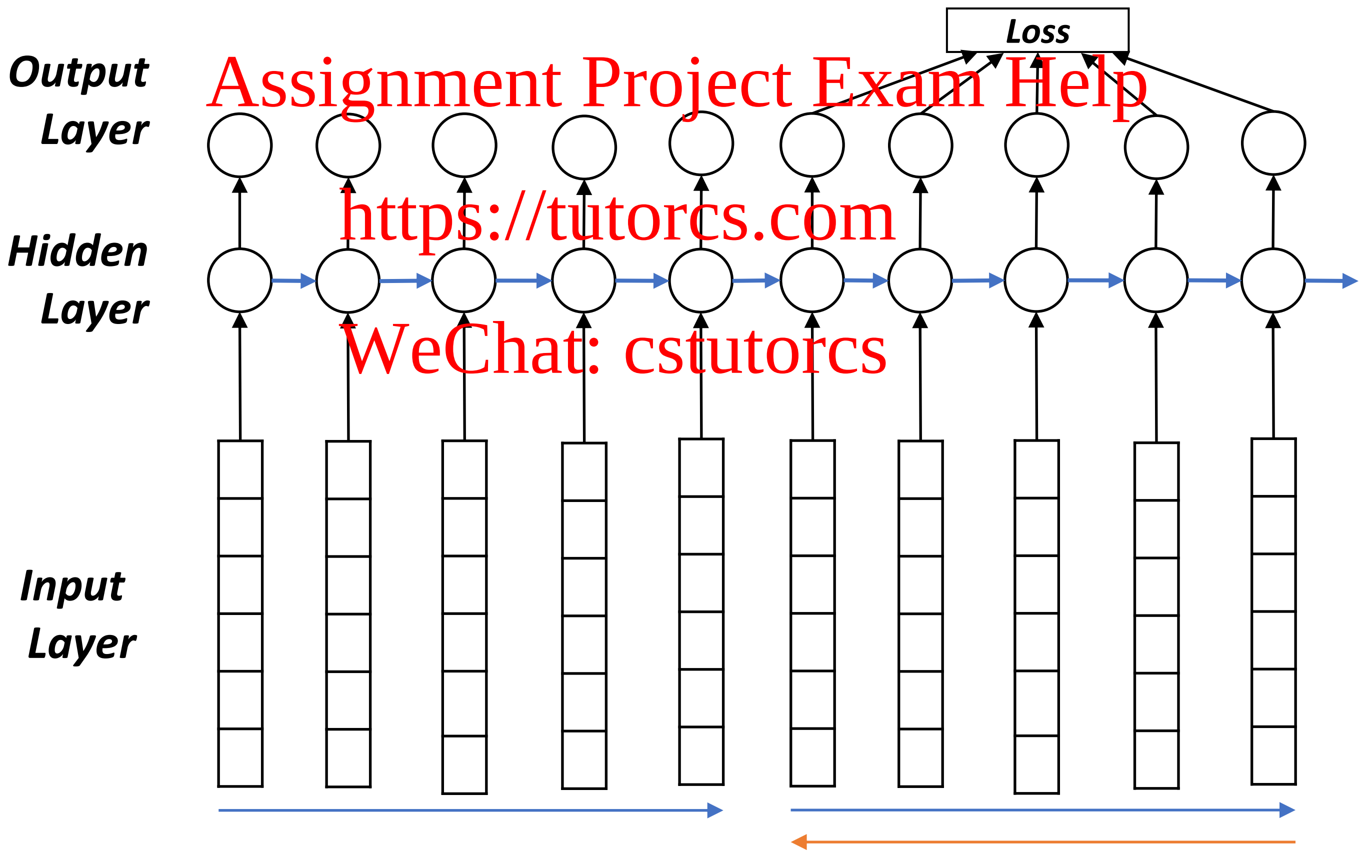


Run forward and backward through chunks of the sequence instead of whole sequence

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Truncated Backpropagation through time

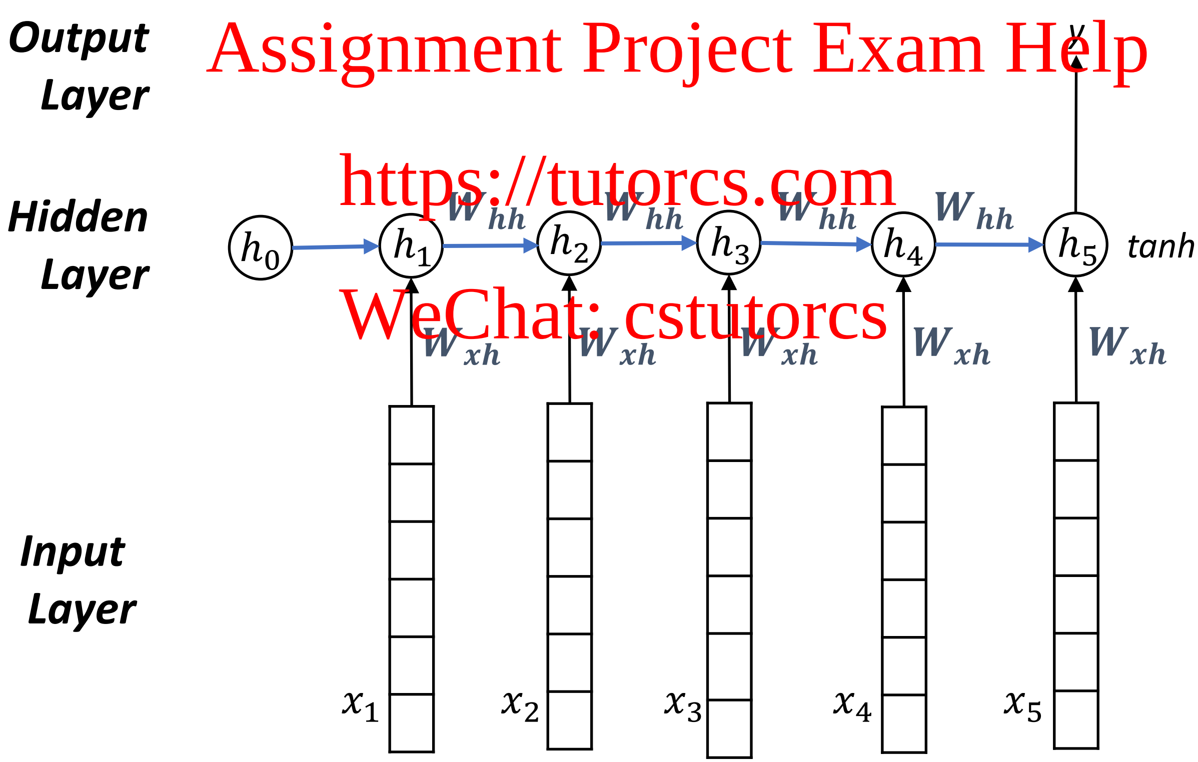


Carry hidden states forward in time forever, but only backpropagate for some smaller number of steps

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

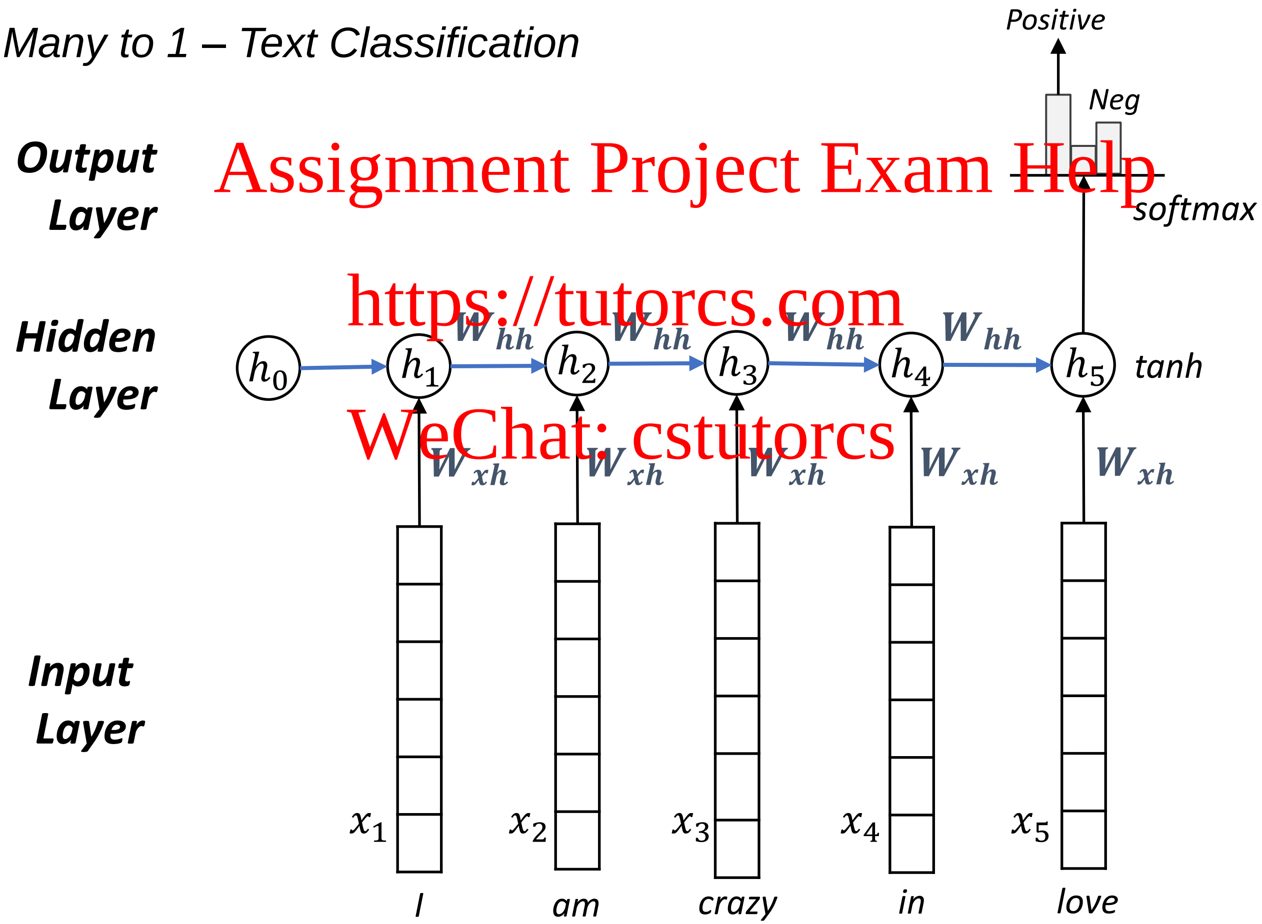
Many to 1



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

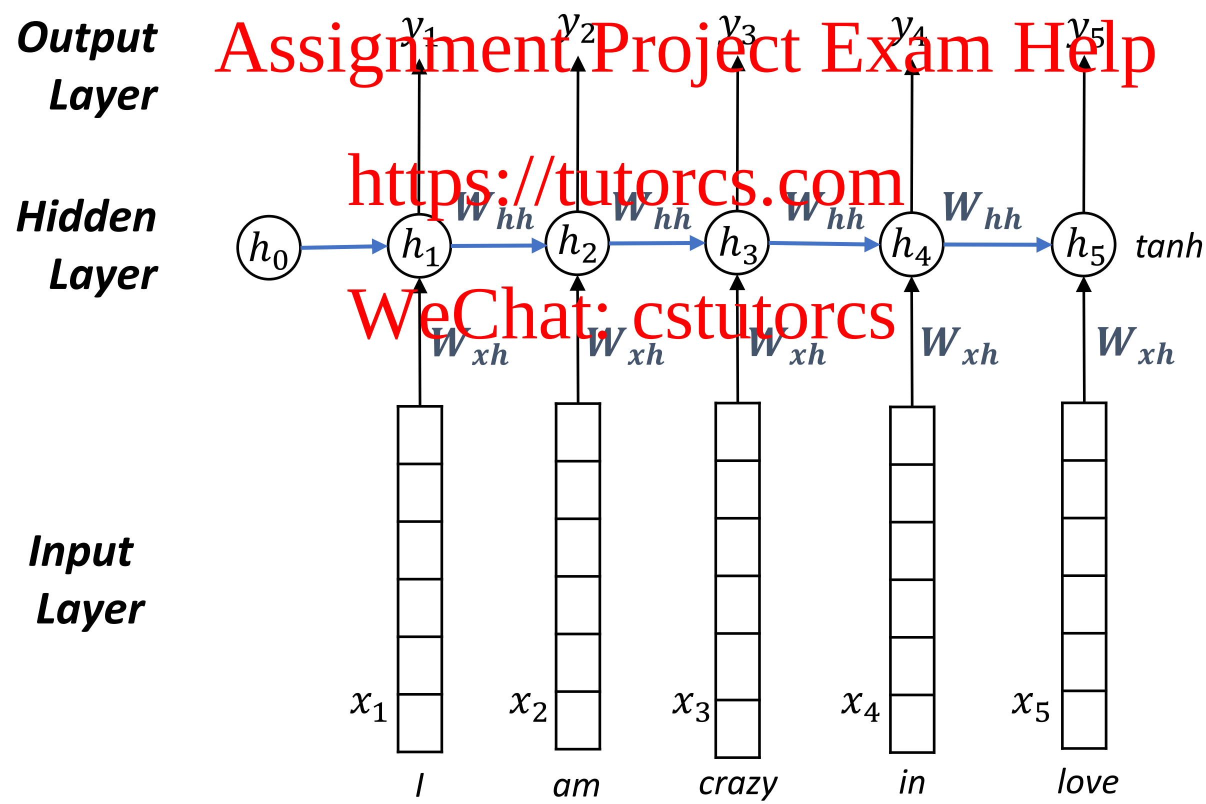
Many to 1 – Text Classification



3 **Seq2Seq with Deep Learning**

Neural Network + Memory = Recurrent Neural Network

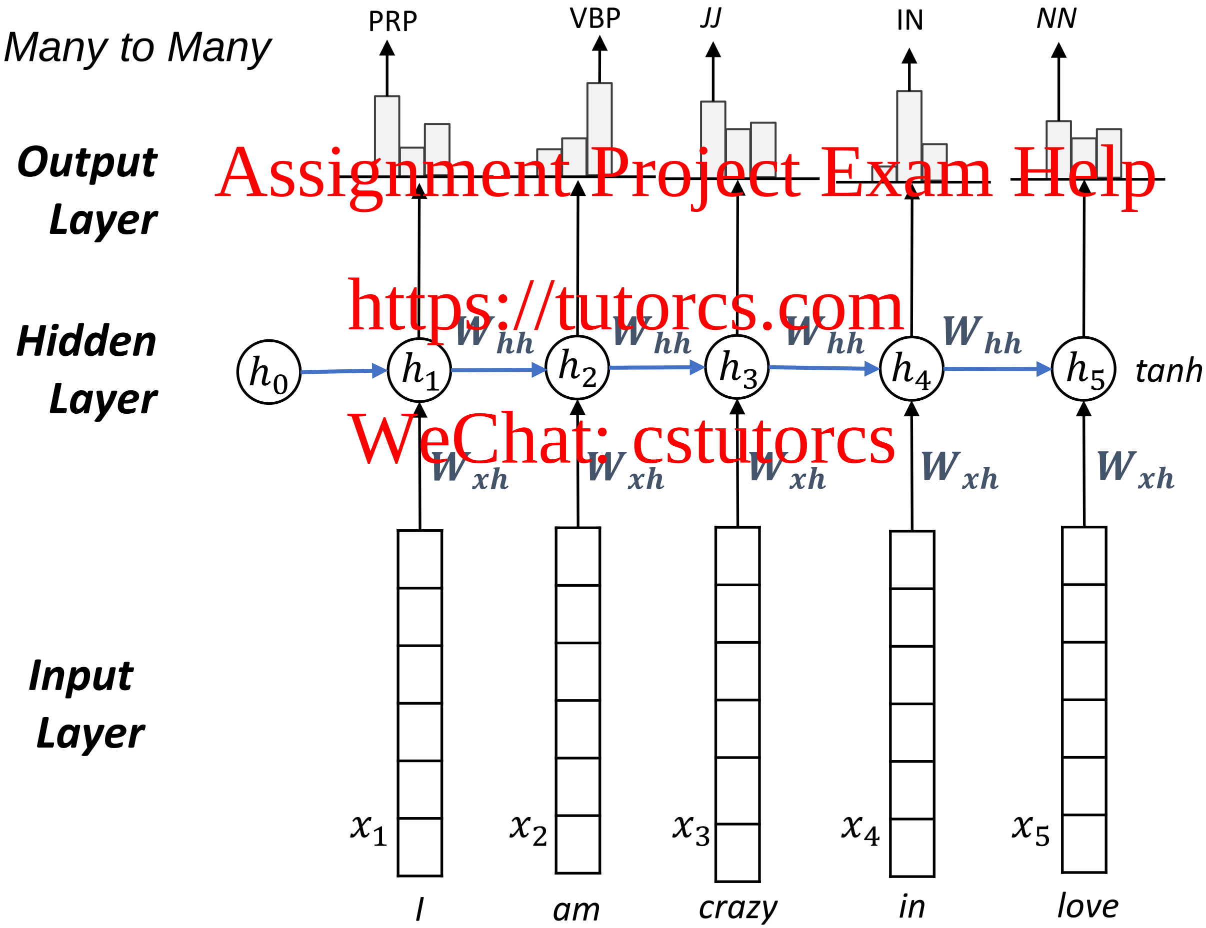
Many to Many



3 Seq2Seq with Deep Learning

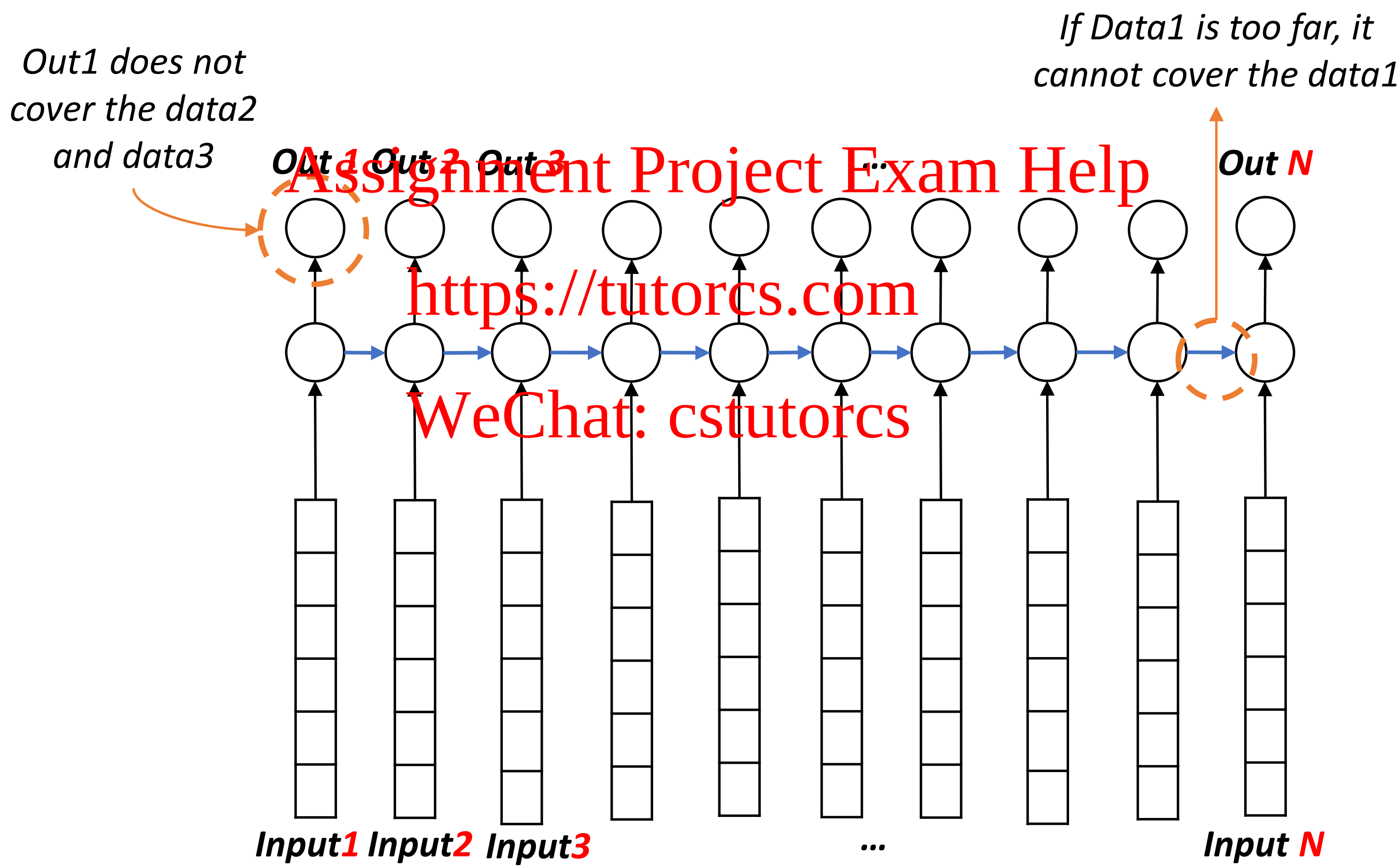
Neural Network + Memory = Recurrent Neural Network

Many to Many



THE UNIVERSITY OF SYDNEY

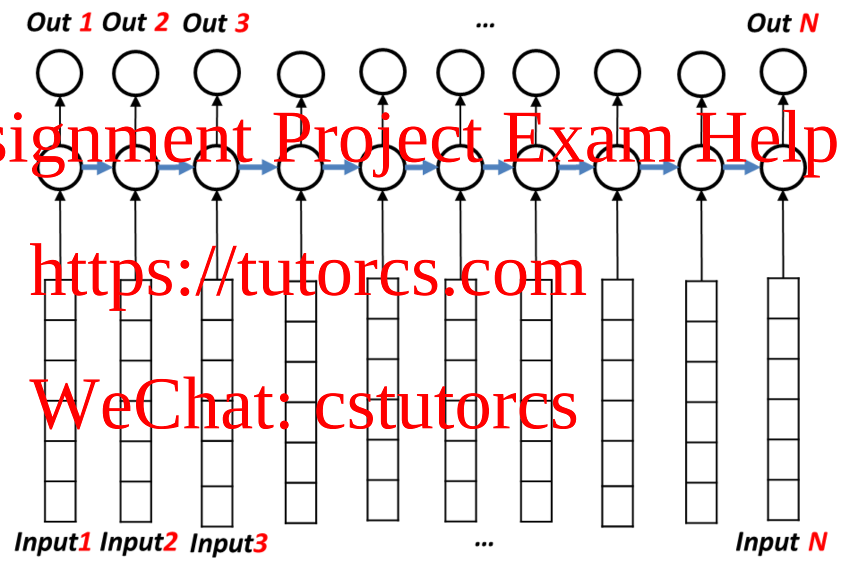
Limitation of Vanilla RNN



The Problem of Learning Long-Range Dependencies

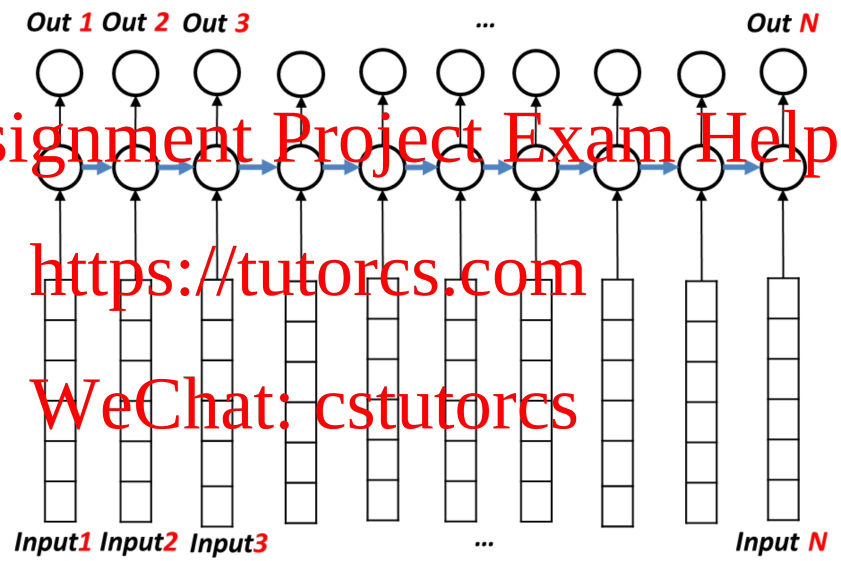
Seq2Seq with Deep Learning

Limitation of Vanilla RNN



"I grew up in Italy … (5 more sentences)… My grandma's house was very cosy and… (5 more sentences)… I speak fluent ____"

Limitation of Vanilla RNN

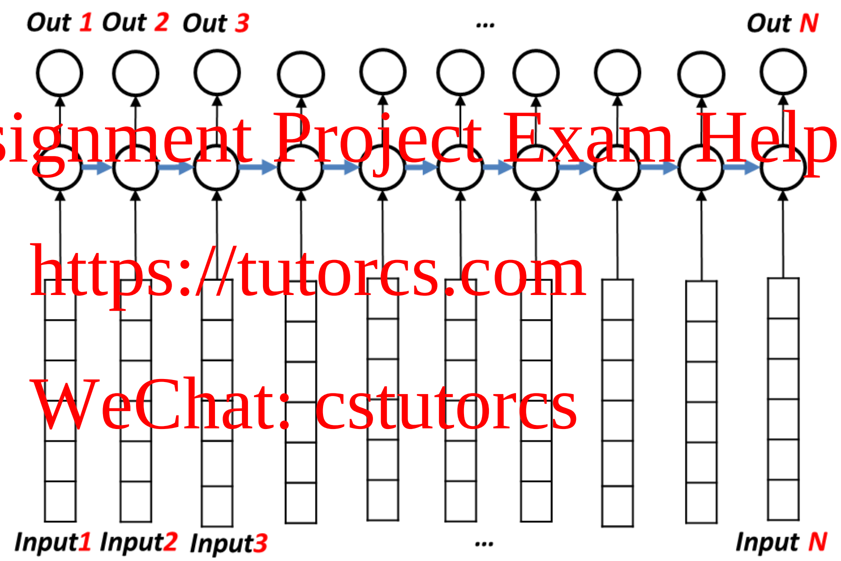


$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\|_2 < 1$$

Limitation1: Vanishing Gradient Issue

During back-propagation and calculating gradients, it tends to get smaller and smaller as we keep on moving backward in the Network. This means that the neurons in the Earlier layers learn very slowly as compared to the neurons in the later layers in the Hierarchy.

Limitation of Vanilla RNN

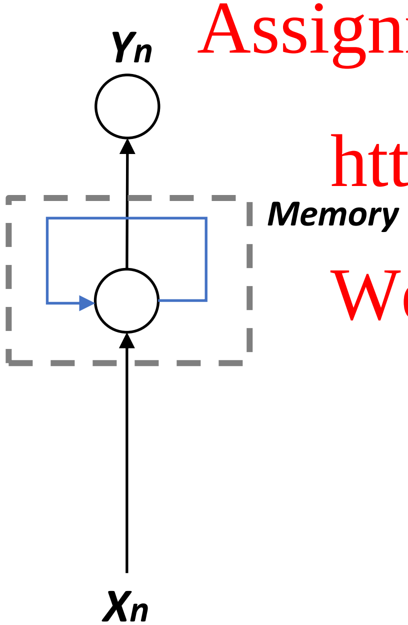


$$\left\|\frac{\partial h_i}{\partial h_{i-1}}\right\|_2 > 1$$

Limitation2: Exploding Gradient

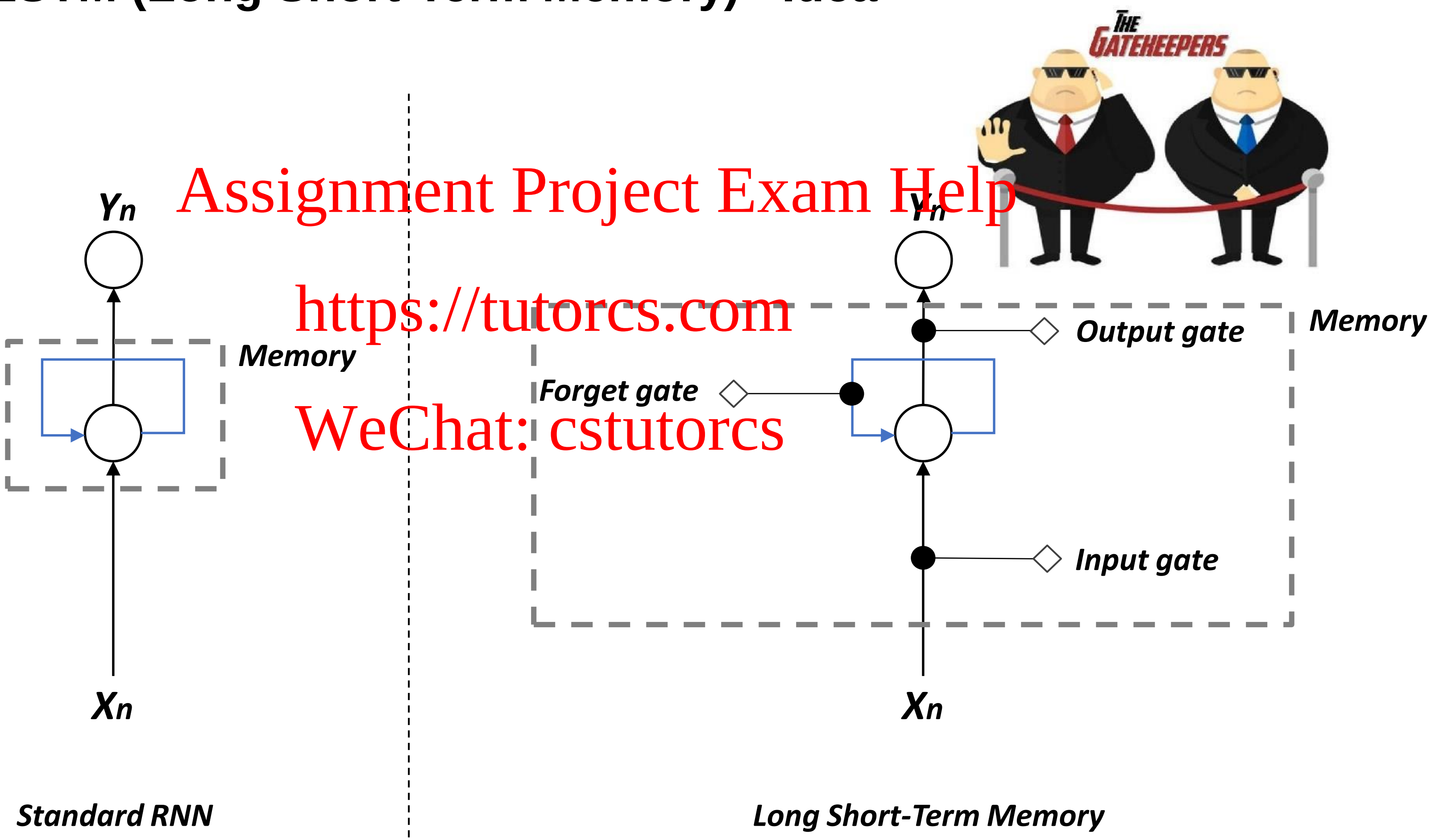
In RNN, error gradients can accumulate during an update and result in very large gradients. These in turn result in large updates to the network weights, and an unstable network. At an extreme, the values of weights can become so large as to overflow and result in NaN weight values that can no longer be updated.

LSTM (Long Short-Term Memory) - Idea



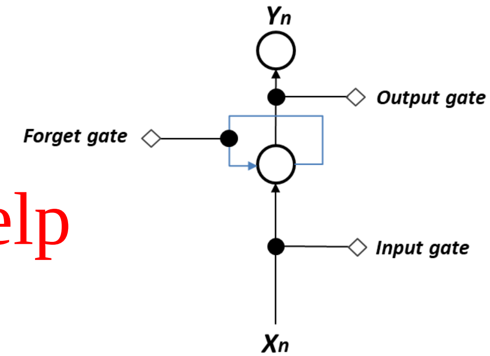
Standard RNN

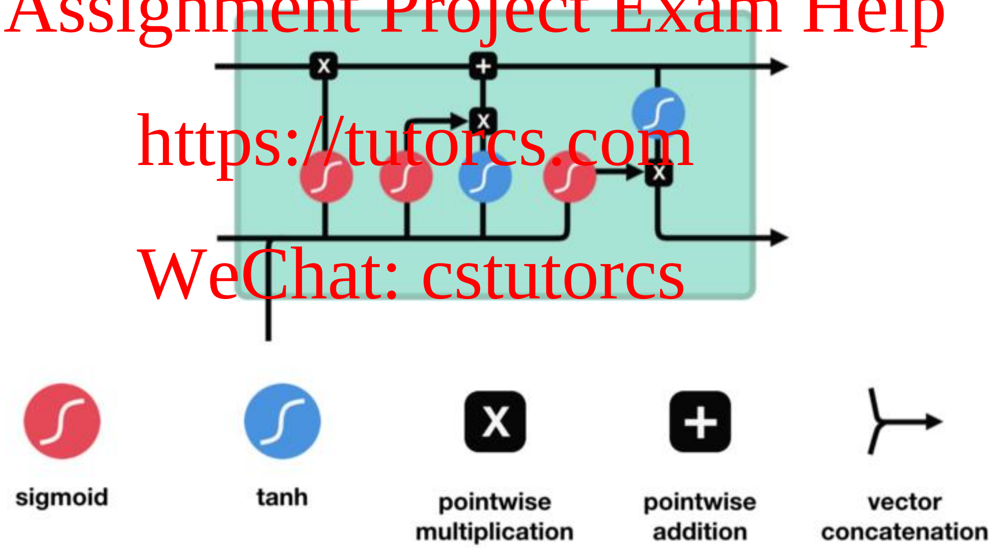
LSTM (Long Short-Term Memory) - Idea



3 Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) - Idea





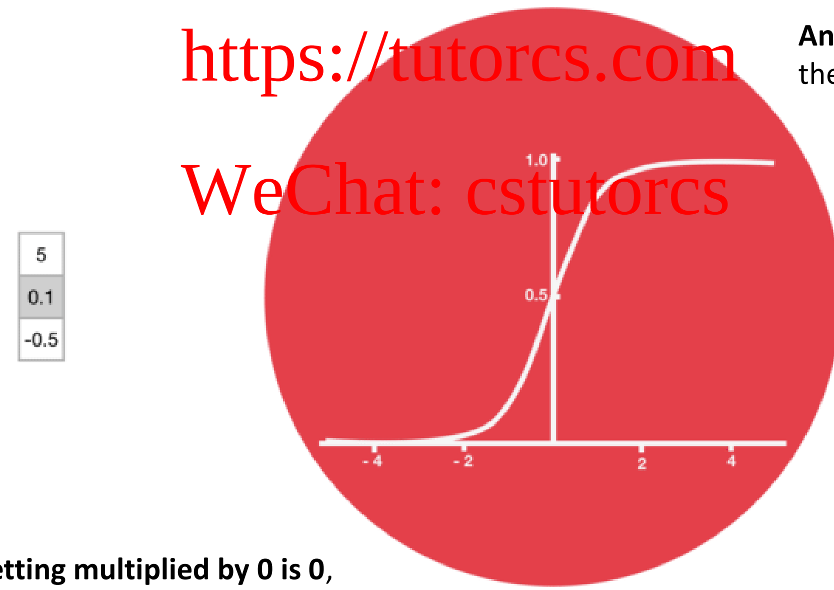
- 4 times more parameters than RNN
- Mitigates **vanishing gradient** problem through **gating**
- Widely used and was SOTA in many sequence learning problems

State-Of-The-Art

3 Seq2Seq with Deep Learning

Sigmoid activation

A sigmoid activation is similar to the tanh activation. Instead of squishing values between -1 and 1, it squishes values between 0 and 1.

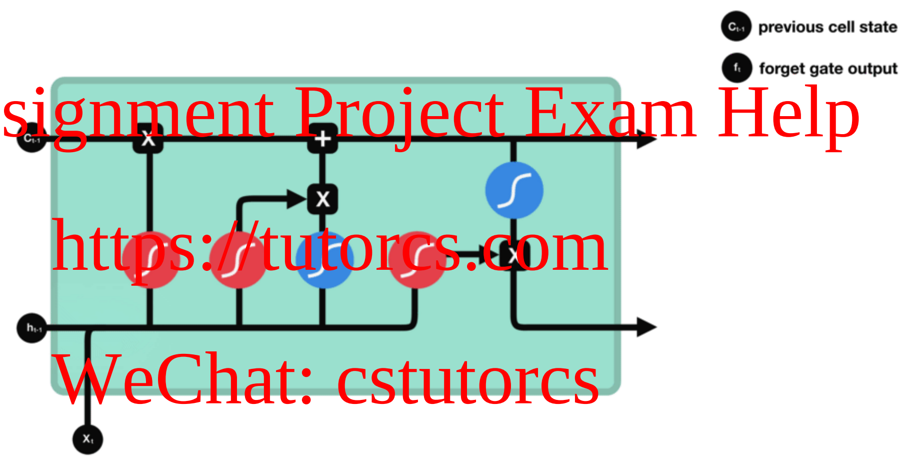


Any number multiplied by 1 is the same value therefore that value stays **the same or is "kept."**

Any number getting multiplied by 0 is 0, causing **values to disappears or be "forgotten."**

3 Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Forget Gate

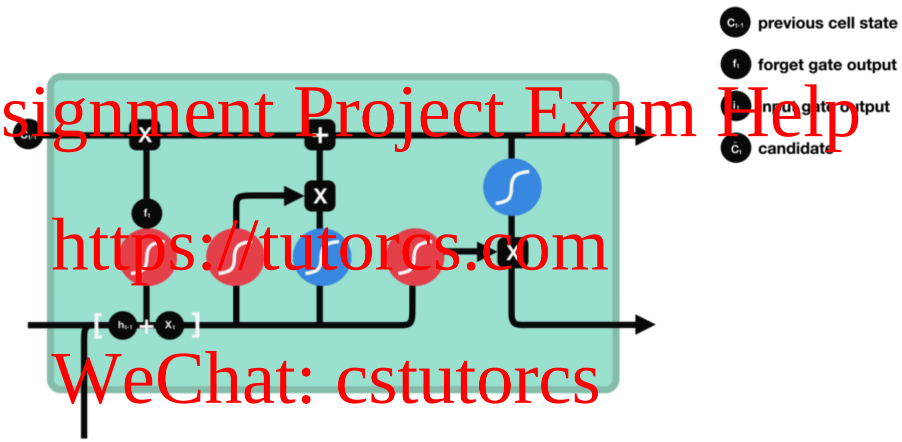


$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Decides what information should be thrown away or kept

Information from the **previous hidden state** and information from the **current input** is passed through the sigmoid function. Values come out between 0 and 1. The closer to 0 means to forget, and the closer to 1 means to keep.

LSTM (Long Short-Term Memory) – Input Gate



$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

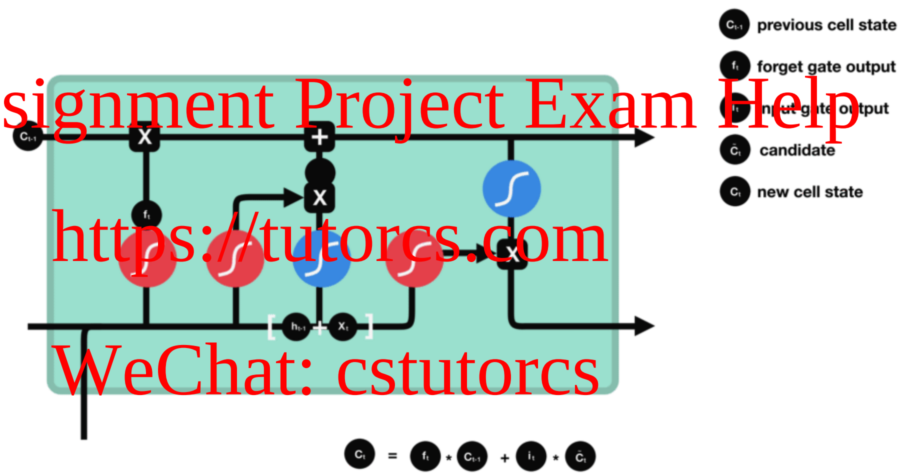
$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$$

1. *Pass the previous hidden state and current input into a sigmoid function*
2. *Pass the hidden state and current input into the tanh function to squish values between -1 and 1 to help regulate the network*
3. *Multiply the tanh output with the sigmoid output*

**sigmoid output will decide which information is important to keep from the tanh output*

Seq2Seq with Deep Learning

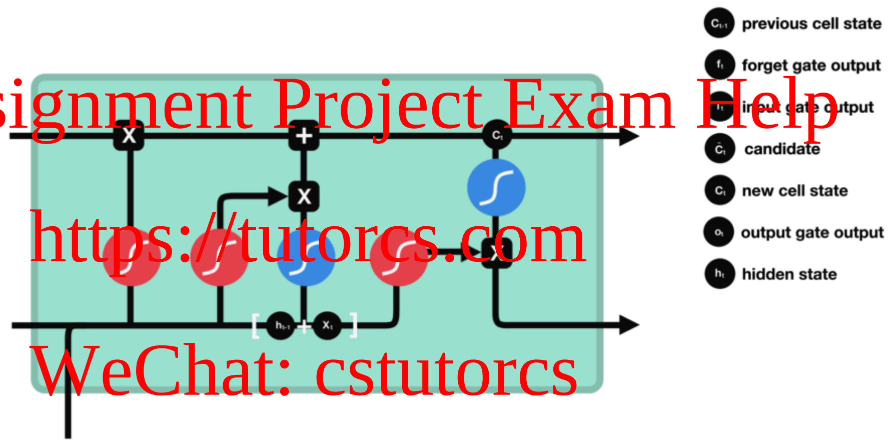
LSTM (Long Short-Term Memory) – Cell States



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- *the cell state gets pointwise multiplied by the forget vector*
- *take the output from the input gate and do a pointwise addition which updates the cell state to new values that the neural network finds relevant*
- *That gives us our new cell state*

LSTM (Long Short-Term Memory) – Output Gate



$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

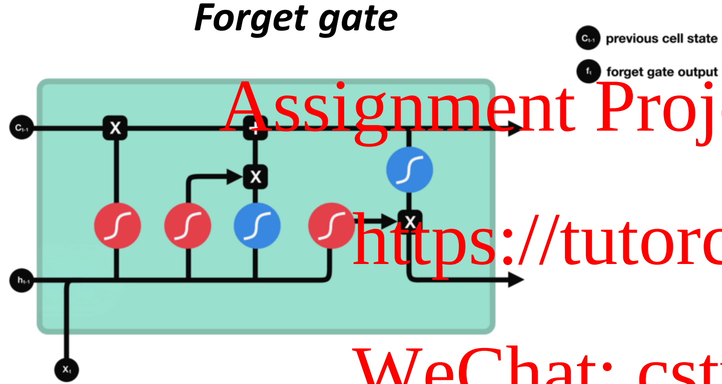
$$h_t = o_t * \tanh(C_t)$$

decides what the next hidden state should be.

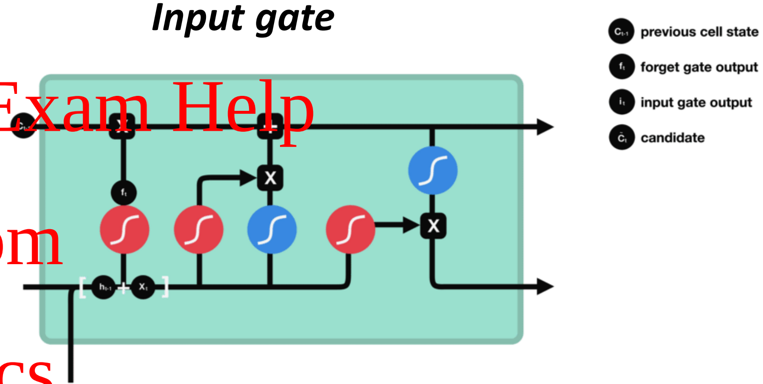
- pass the previous hidden state and the current input into a sigmoid function
- pass the newly modified cell state to the tanh function
- multiply the tanh output with the sigmoid output to decide what information the hidden state should carry

LSTM (Long Short-Term Memory) - Overall

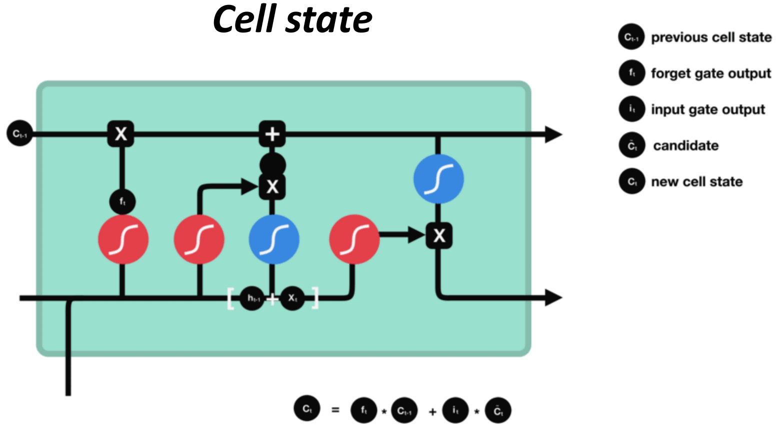
Forget gate



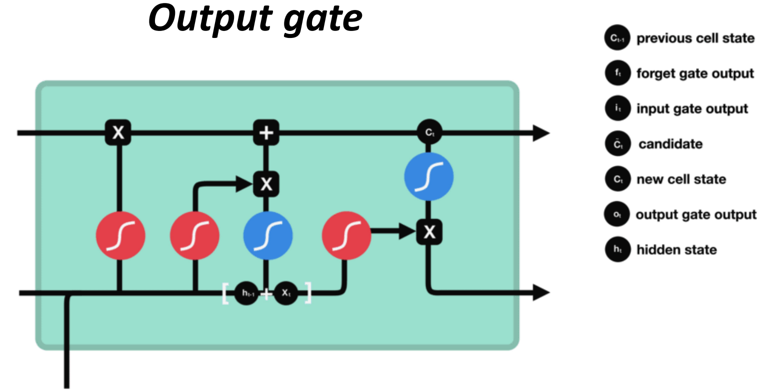
Input gate



Cell state

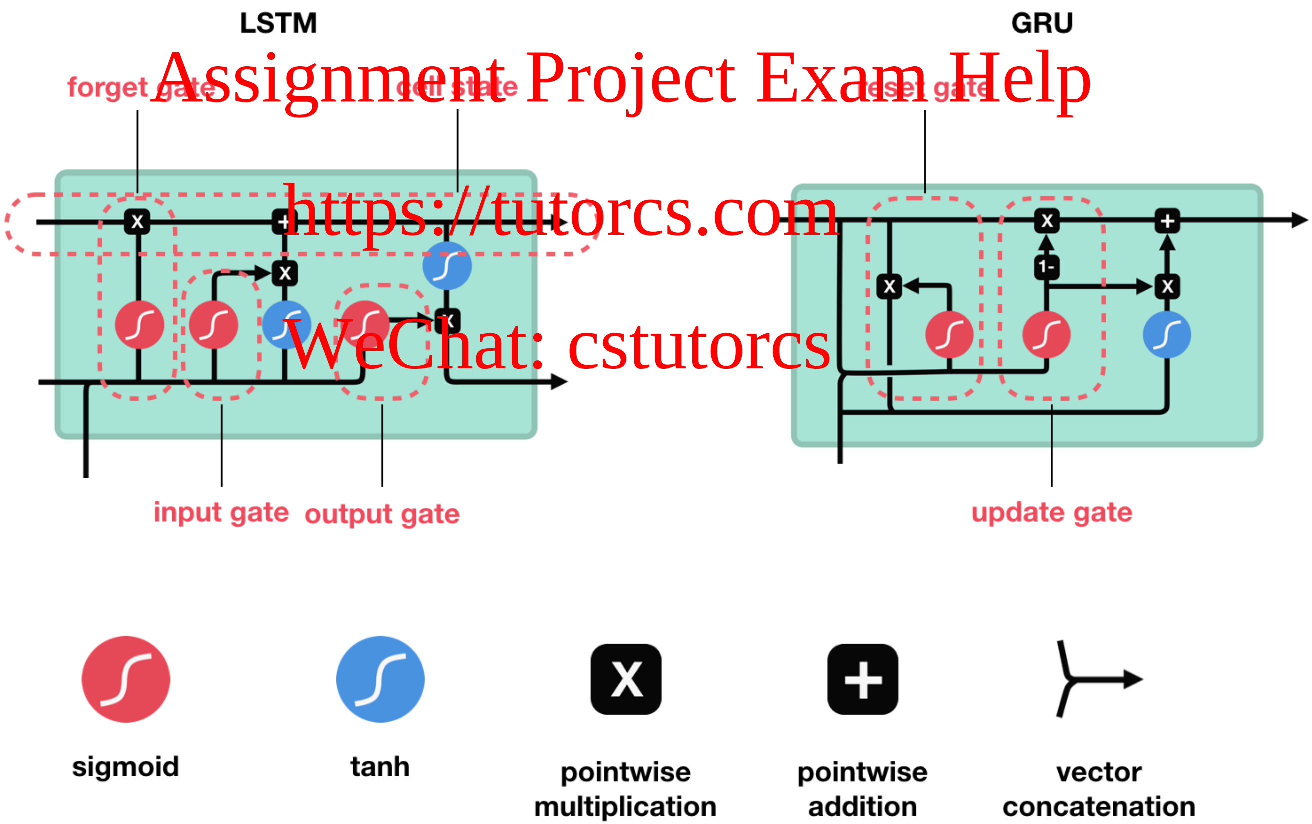


Output gate



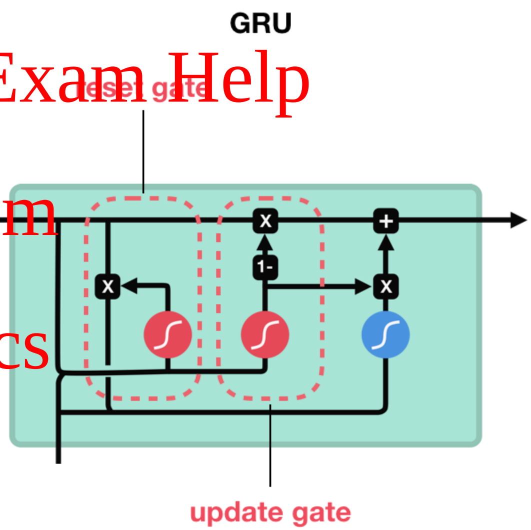
3 Seq2Seq with Deep Learning

Gated Recurrent Unit



Gated Recurrent Unit

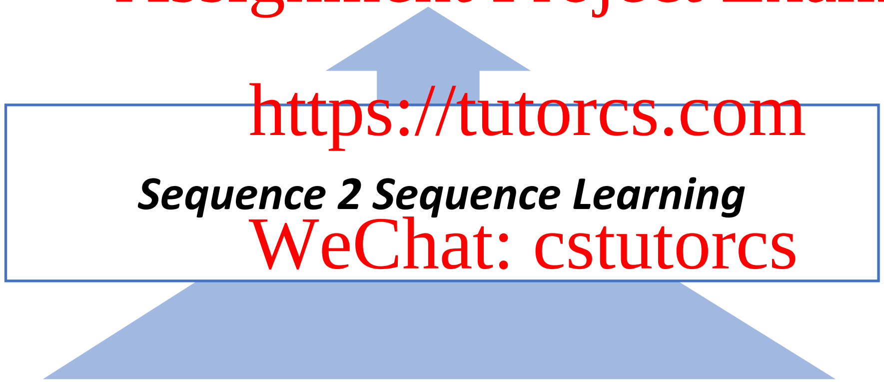
- GRU first computes an **update gate** based on **current input word vector** and **hidden state**
- Compute reset gate similarly but with different weights
  - If reset gate unit is ~0, then this ignores previous memory and only stores the new word information
- Final memory at time step combines current and previous time steps



Seq2Seq – PoS tagger

ADV VERB DET NOUN NOUN

Output: Part of Speech

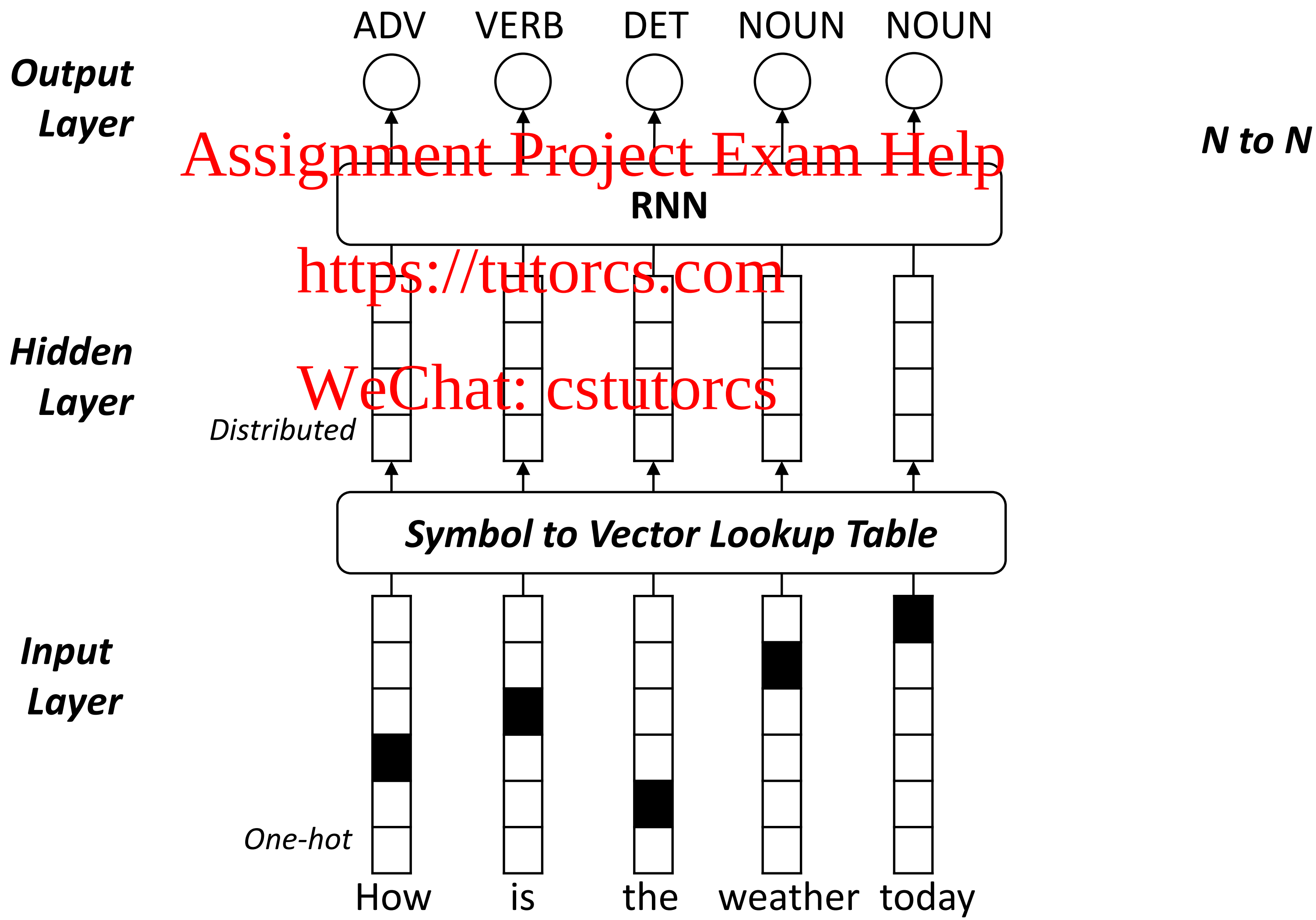


How is the weather today

Input: Text

3 Seq2Seq Modelling

Sequence Modelling for POS Tagging



0 LECTURE PLAN

Lecture 4: Word Classification and Machine Learning 2

1. Machine Learning and NLP: Finish
2. Seq2Seq Learning
3. Seq2Seq Deep Learning
    1. RNN (Recurrent Neural Network)
    2. LSTM (Long Short-Term Memory)
    3. GRU (Gated Recurrent Unit)
4. **Data Transformation for Deep Learning NLP**
5. Next Week Preview
    - Natural Language Processing Stack

4 Data Transformation for Deep Learning NLP

ImageNet: Image Classification

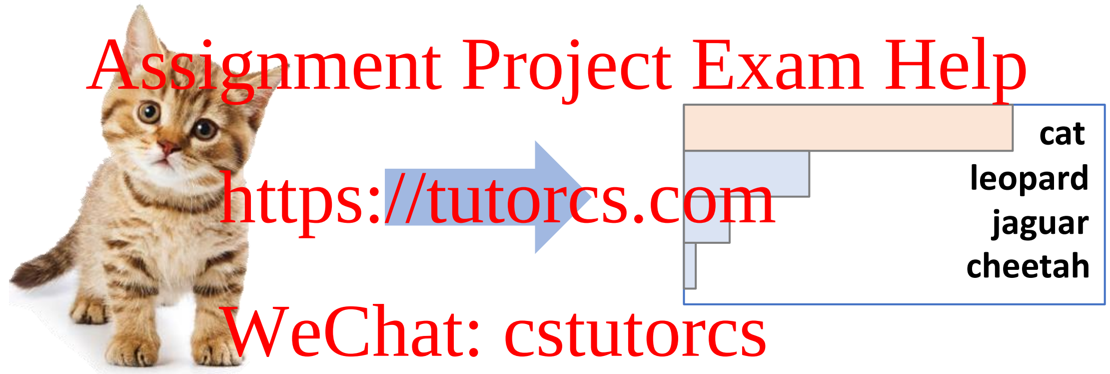
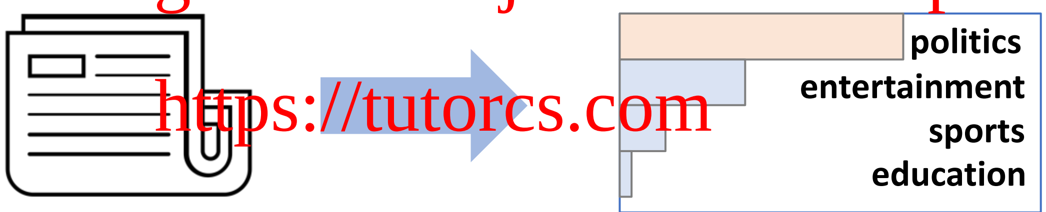


Image Pixel

4 Data Transformation for Deep Learning NLP

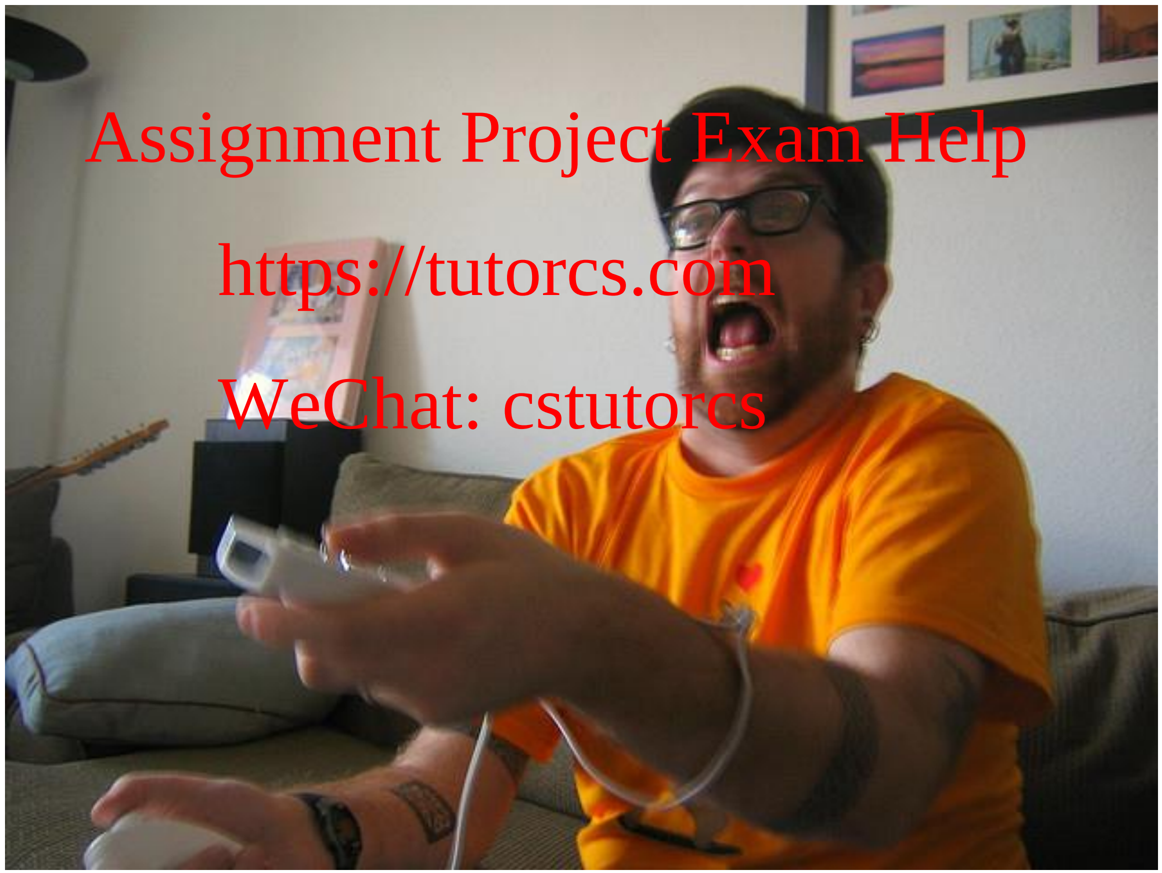
Topic Classification



News Articles

4 Data Transformation for Deep Learning NLP

Visual Question Answering



Visual Question Answering

What color of the shirt does he wear

Submit

Predicted top-5 answers with confidence:

orange	99.999%
yellow	0.001%
orange and white	0.000%
yellow and orange	0.000%
orange and black	0.000%

Data Transformation for Deep Learning NLP

Visual Question Answering

Visual Question Answering

Why is he surprised

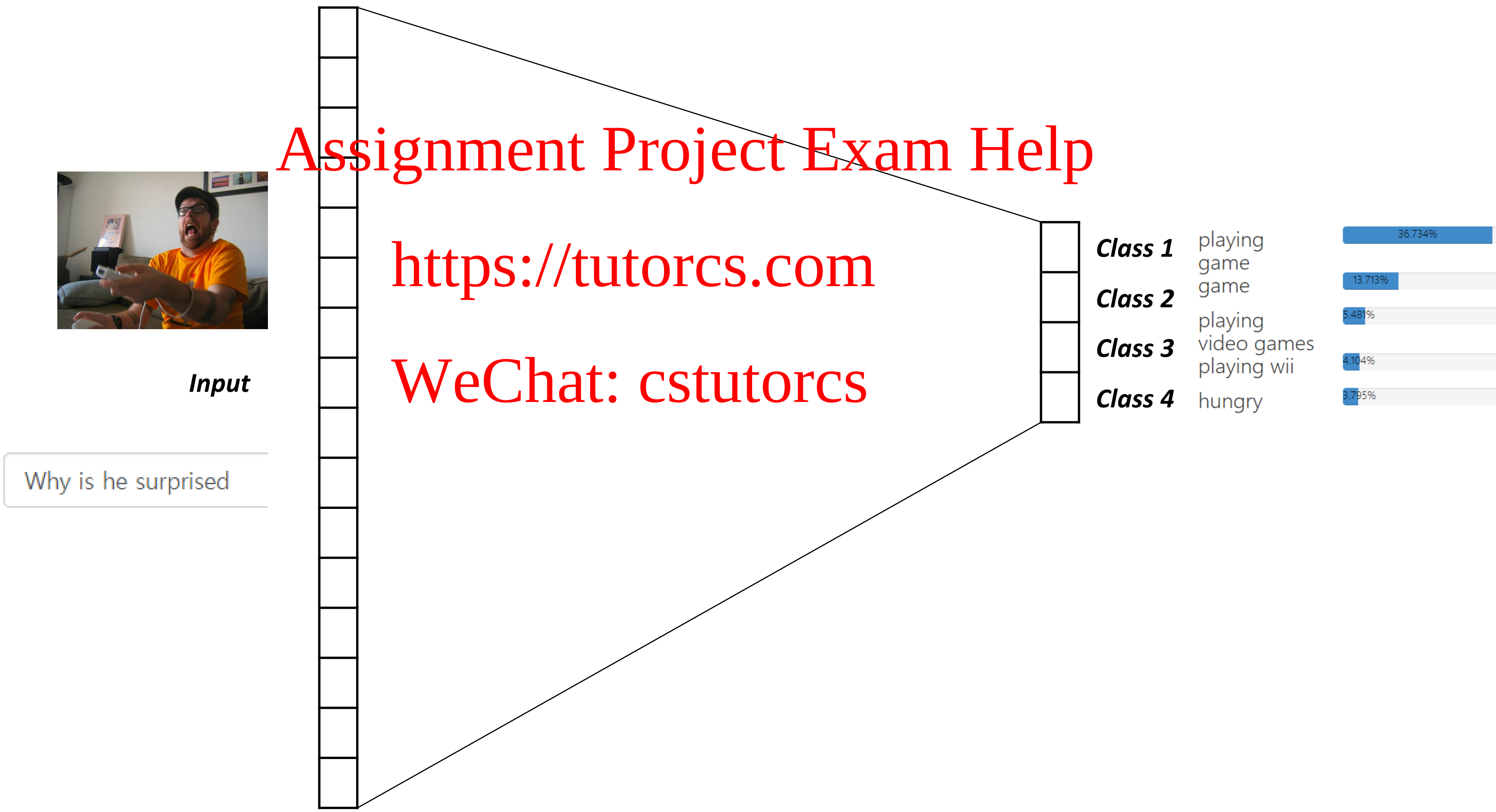
Submit

Predicted top-5 answers with confidence:

playing game	36.734%
game	13.713%
playing video games	5.481%
playing wii	4.104%
hungry	3.795%

4 Data Transformation for Deep Learning NLP

Classification Formulation

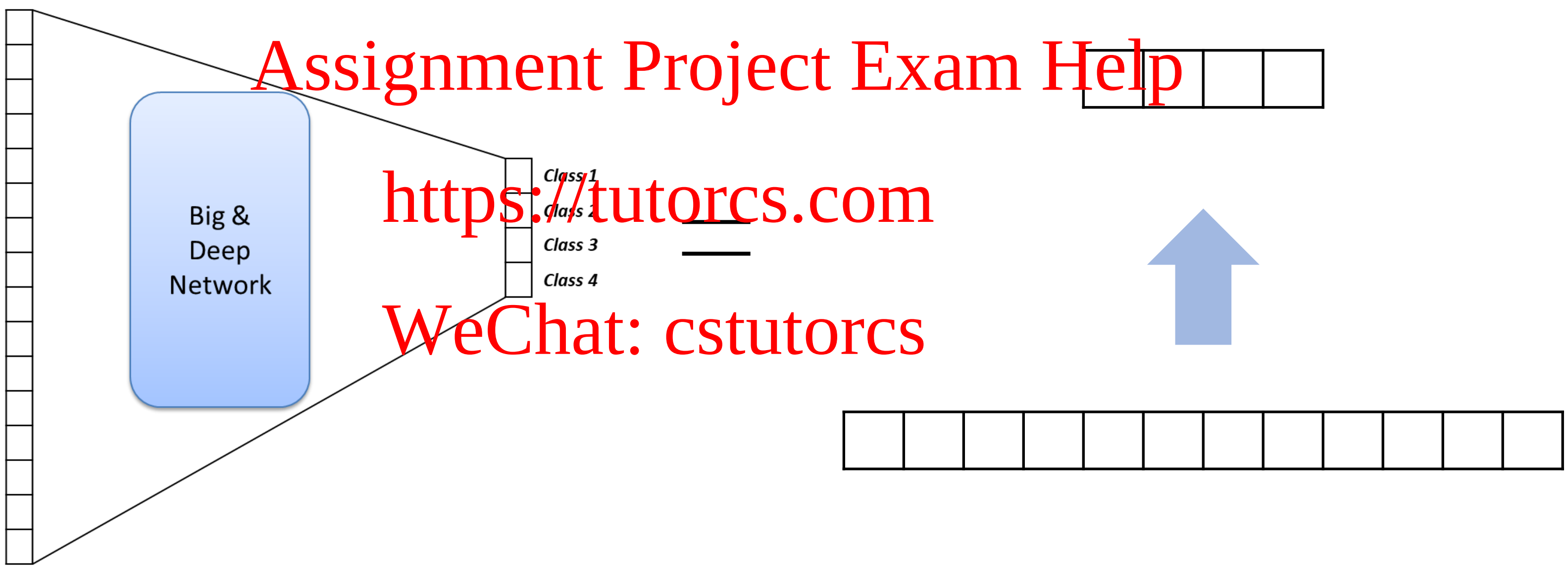


4 Data Transformation for Deep Learning NLP

Classification Formulation

4 Data Transformation for Deep Learning NLP

Classification



Graphical Notation for Data

Data X

10
2
8
2
15
3
5
1
5

=

Data X

10	2	8
2	15	3
5	1	5

4 Data Transformation for Deep Learning NLP

V to 1

10
2
8
2
15
3
5
1
5

=

10	2	8
2	15	3
5	1	5

?

Data Transformation for Deep Learning NLP

V to 1 – Simple Method

center one

10	2	8
2	15	3
5	1	5

→ 15

average

10	2	8
2	15	3
5	1	5

→ 5.6

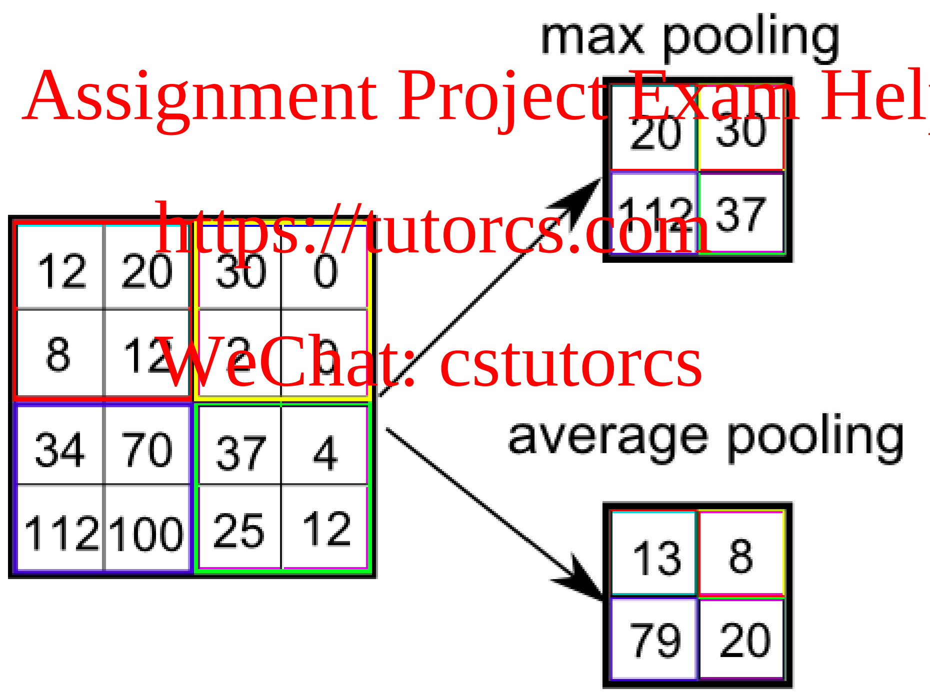
median

10	2	8
2	15	3
5	1	5

→ 5

4 Data Transformation for Deep Learning NLP

V to 1 – Simple Method



Data Transformation for Deep Learning NLP

V to 1 – Weighted Method

Weighted Sum

10	2	8
2	15	3
5	1	5

Value

3	1	5
2	6	3
9	3	6

Weight

→ 253

Weighted Average

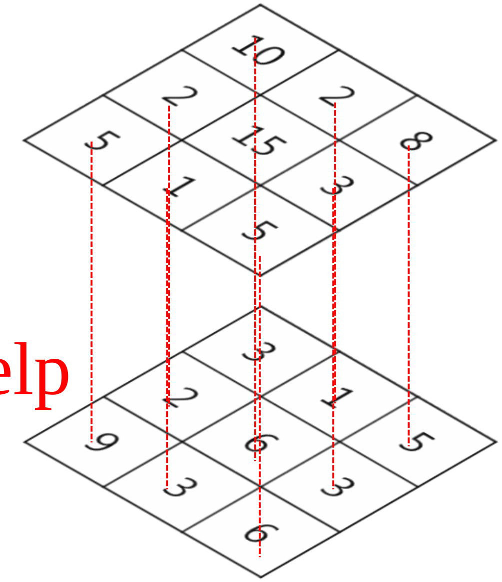
10	2	8
2	15	3
5	1	5

Value

3/9	1/9	5/9
2/9	6/9	3/9
9/9	3/9	6/9

Weight

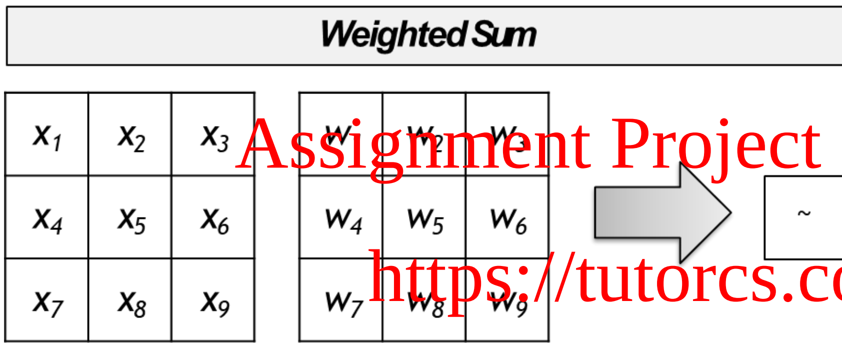
→ 6.65



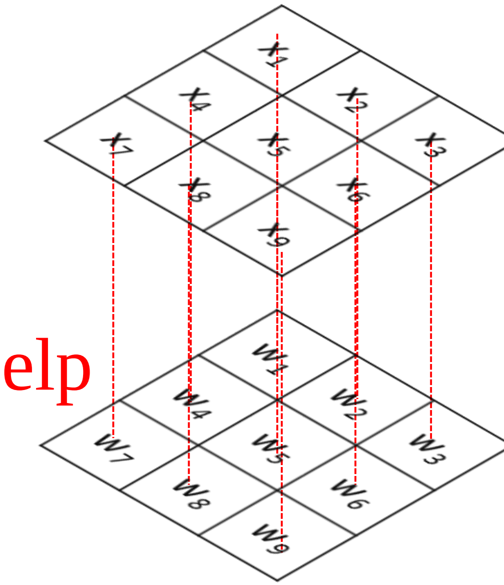
Element-wise multiplication

4 Data Transformation for Deep Learning NLP

V to 1 – General Form



$$v = x_1 * w_1 + x_2 * w_2 + \cdots + x_9 * w_9$$



Element-wise multiplication

Data Transformation for Deep Learning NLP

V to 1 – Linear Algebra

[1 x 9] matrix

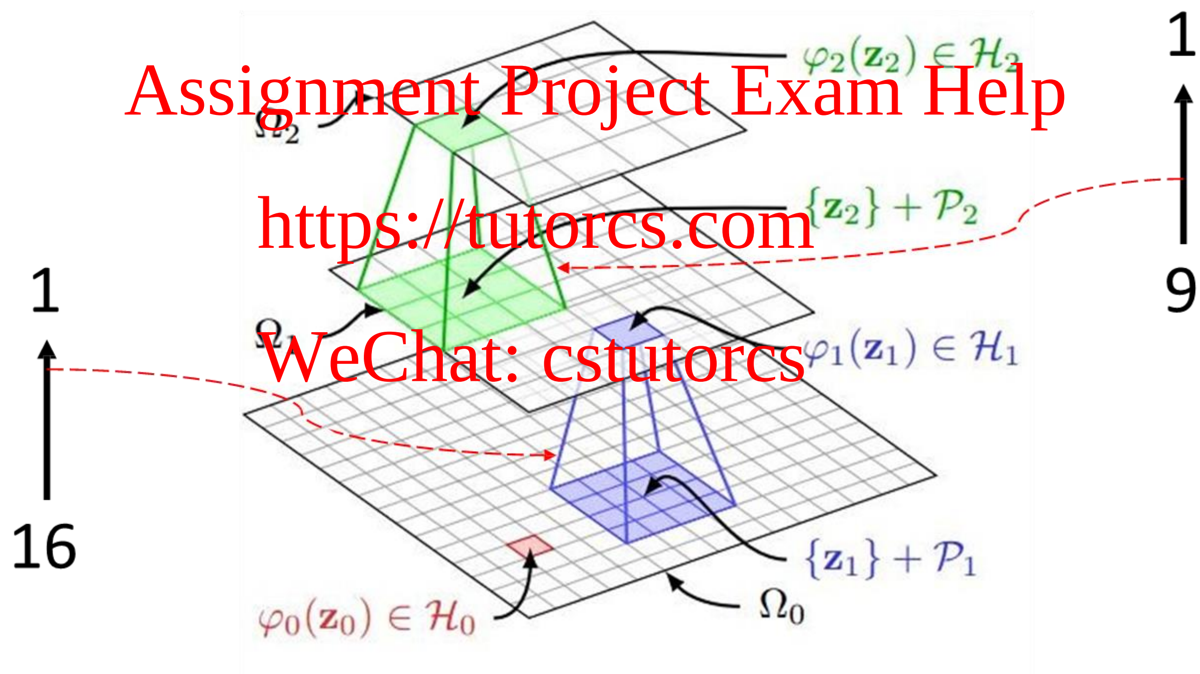
[9x1] matrix

[1x1] matrix

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \end{bmatrix} \text{ x } \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ w_9 \end{bmatrix} = \sum_{i}^{9} x_i * w_i$$

4 Data Transformation for Deep Learning NLP

Convolution Neural Network (1)

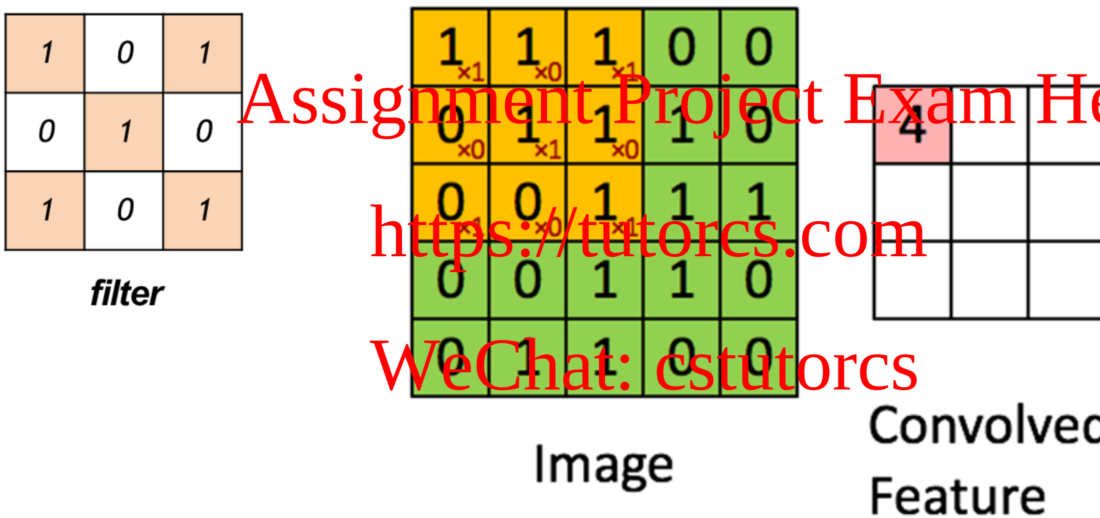


(a) Hierarchy of Image Feature Maps

Data Abstraction

Mairal et al., Convolutional Kernel Networks, 2014

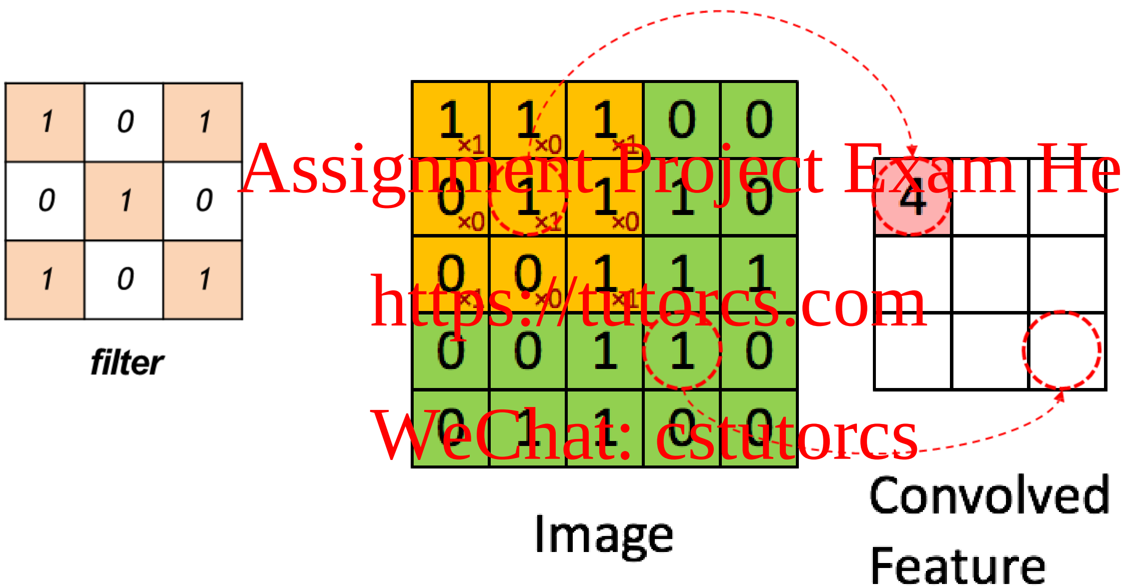
Convolution Neural Network (2)



Mairal et al., Convolutional Kernel Networks, 2014

THE UNIVERSITY OF SYDNEY

Convolution Neural Network (2)



Mairal et al., Convolutional Kernel Networks, 2014

4 Data Transformation for Deep Learning NLP

V to V'

V= 9

10	2	8
2	15	3
5	1	5

V'= 2

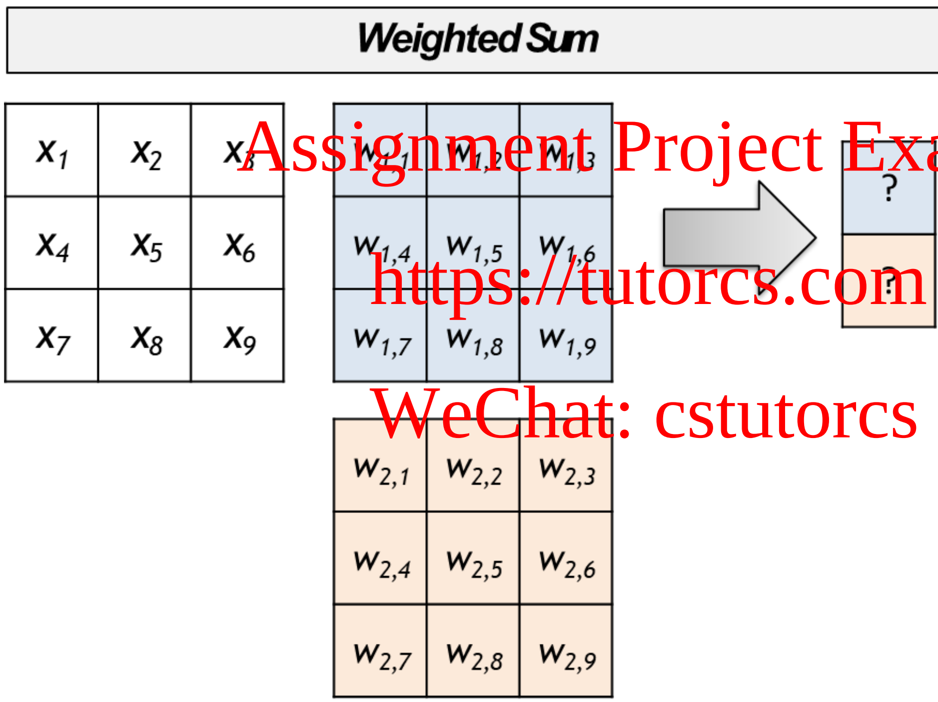
?
?

10	*2*	*8*	*2*	*15*	*3*	*5*	*1*	*5*

?	*?*

4 Data Transformation for Deep Learning NLP

V to V' – generalized method



$$\begin{bmatrix} v_1 = x_1 * w_{1,1} + x_2 * w_{1,1} + \cdots + x_9 * w_{1,9} \\ v_2 = x_1 * w_{2,1} + x_2 * w_{2,1} + \cdots + x_9 * w_{2,9} \end{bmatrix}$$

Data Transformation for Deep Learning NLP

V to V' – generalized method

Weighted Sum

[1 x 9] matrix

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

X

[9x2] matrix

$w_{1,1}$	$w_{2,1}$
$w_{1,2}$	$w_{2,2}$
$w_{1,3}$	$w_{2,3}$
$w_{1,4}$	$w_{2,4}$
$w_{1,5}$	$w_{2,5}$
$w_{1,6}$	$w_{2,6}$
$w_{1,7}$	$w_{2,7}$
$w_{1,8}$	$w_{2,8}$
$w_{1,9}$	$w_{2,9}$

=

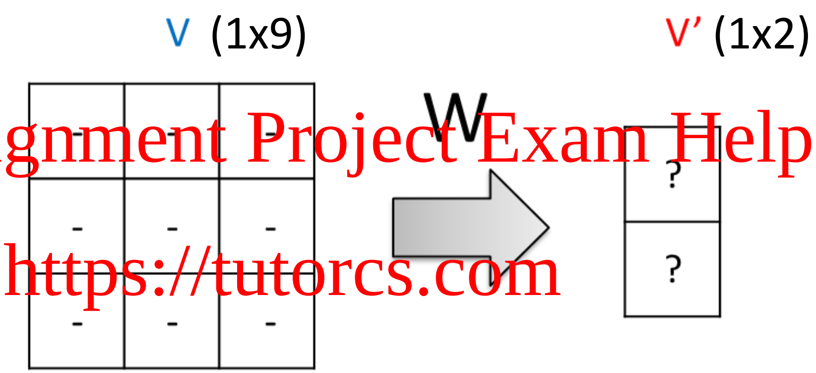
[1x2] matrix

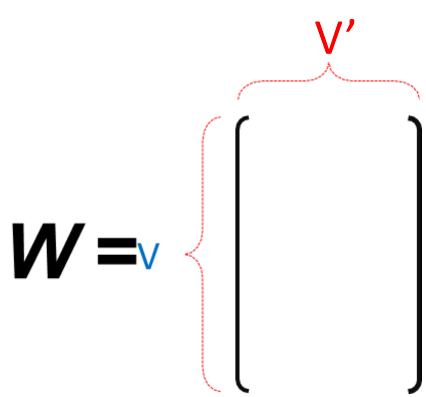
$$\left[\sum_i^9 x_i * w_{1,i} \; , \; \sum_i^9 x_i * w_{2,i} \right]$$

Fully Connected Network

4 Data Transformation for Deep Learning NLP

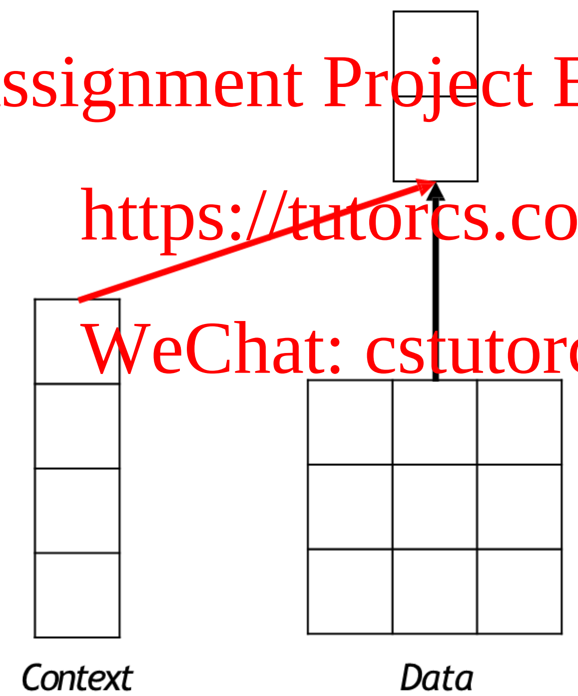
V to V’ – Projection Notation



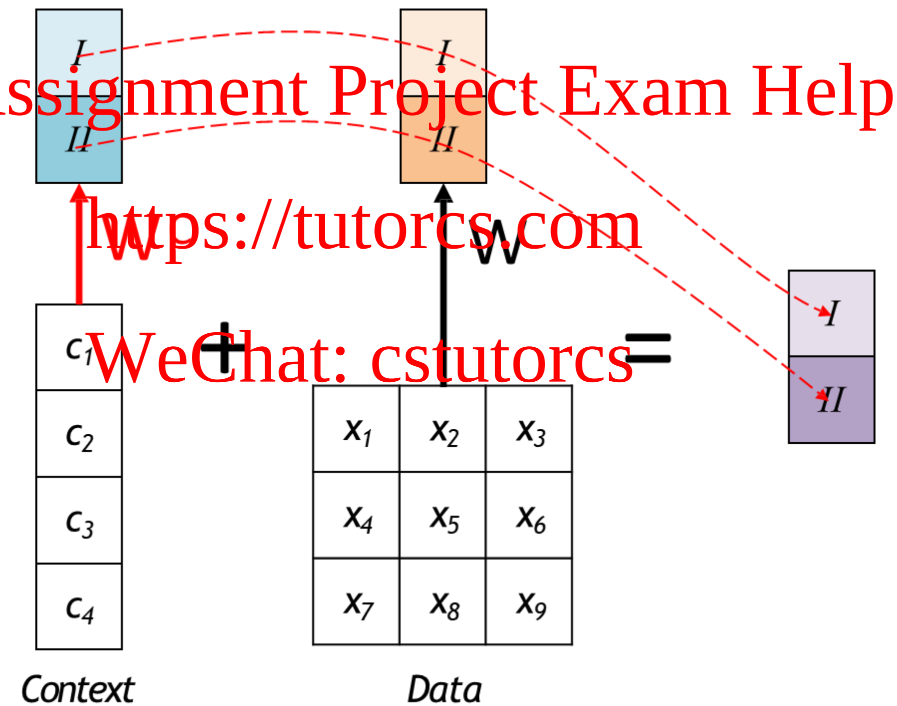


Data Transformation for Deep Learning NLP

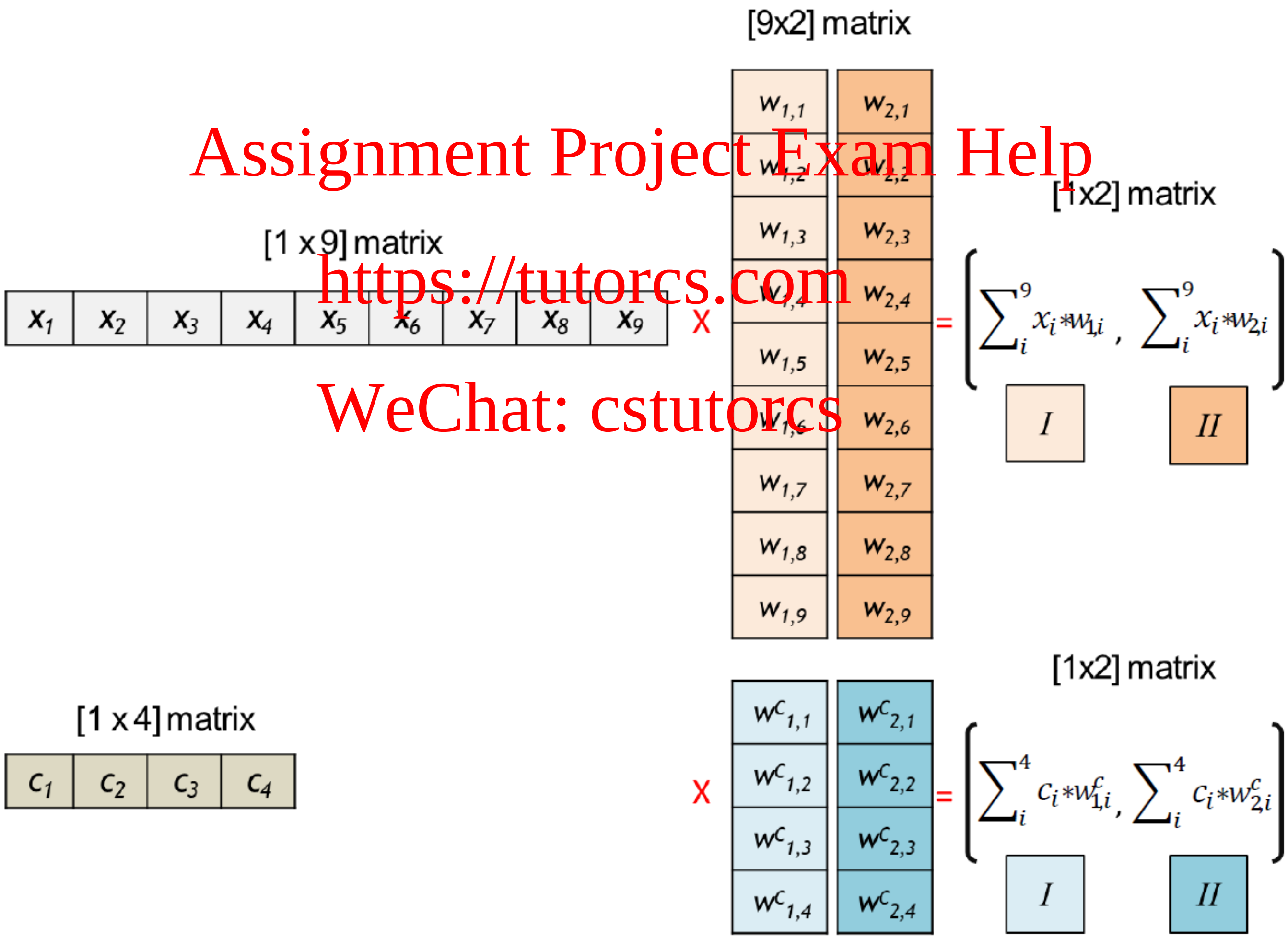
V to V' – Projection with Context (1)



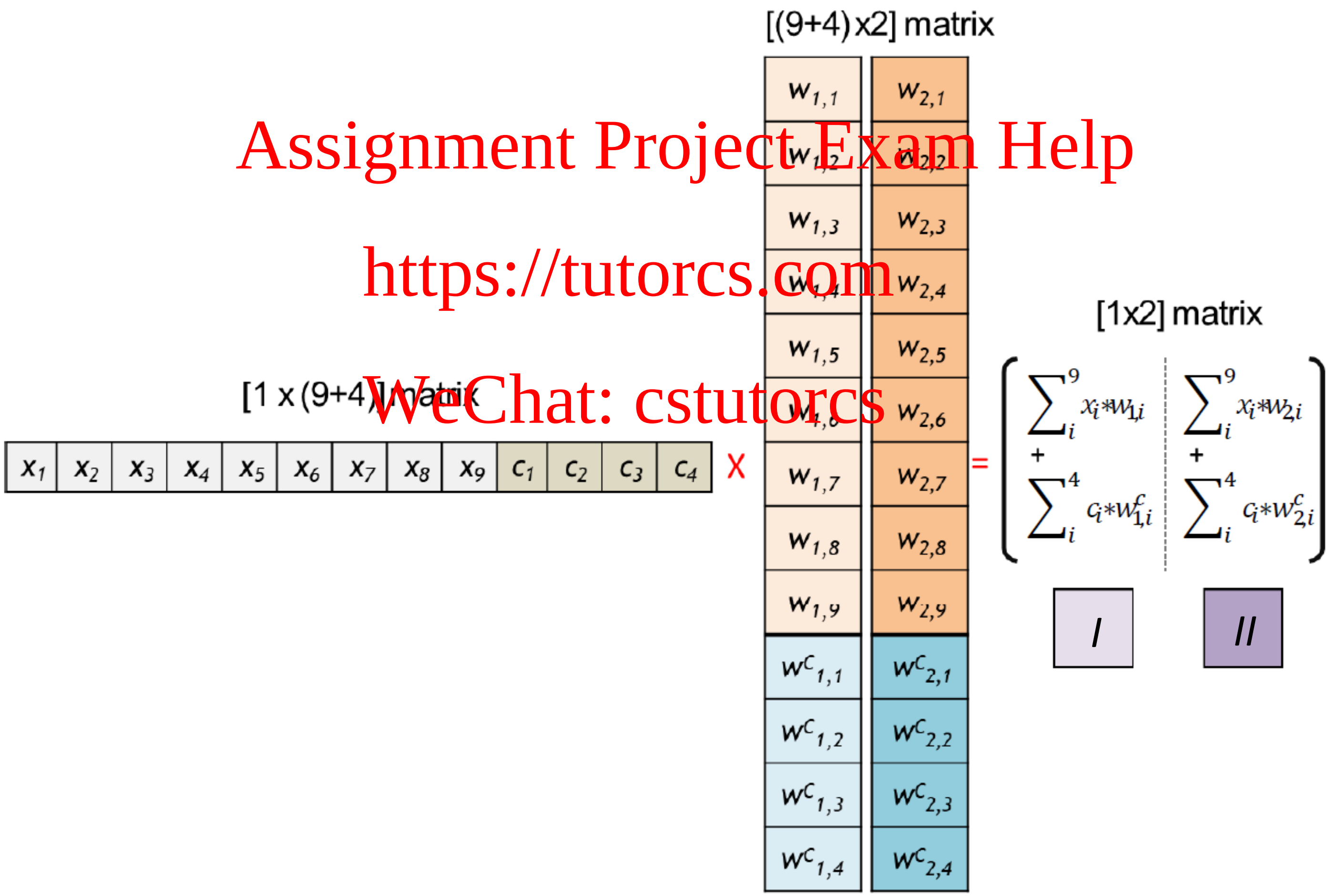
V to V' – Projection with Context (2)



V to V' with Context - Linear Algebra

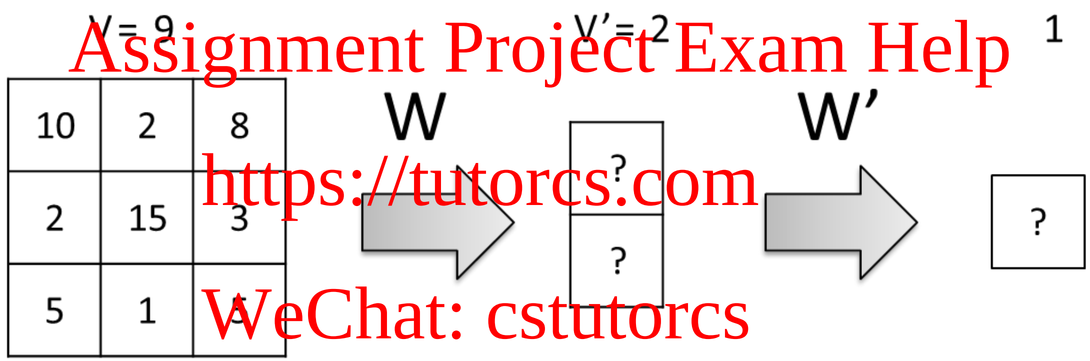


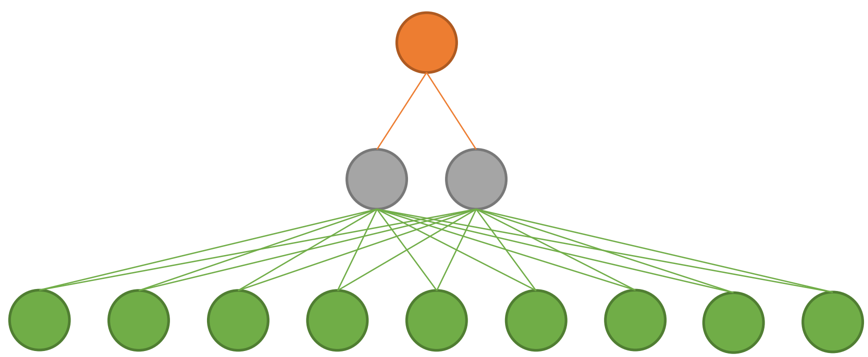
V to V’ with Context - Linear Algebra (Simplified)



4 Data Transformation for Deep Learning NLP

V → V' → 1





4 Data Transformation for Deep Learning NLP

V → V' → 1

Single Layer

V →V' →1

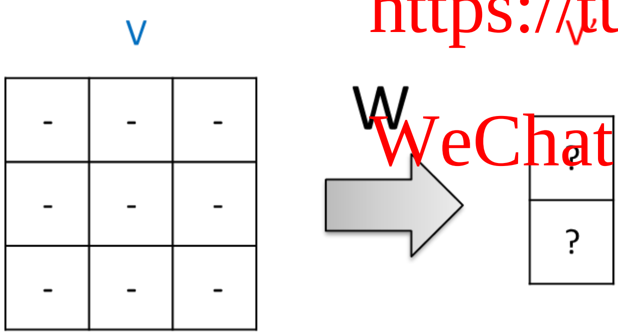
Multilayer

V →V' →V'' →1

Seq2Seq Encoding

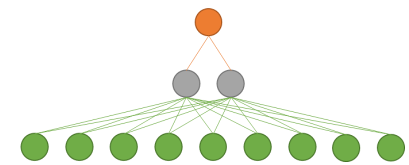
Single Item Summarisation

Multiple Item Summarisation



?

4 Data Transformation for Deep Learning NLP



Multiple Item Summarisation

10	2	8
2	15	3
5	1	5

Data 1

13	4	8
4	5	2
1	45	31

Data 2

6	3	4
1	7	1
3	4	0

Data 3

→

?	?	?
?	?	?
?	?	?

V

→

?
?

V'

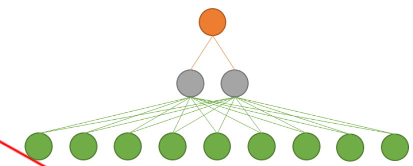
→

?

1

4 Data Transformation for Deep Learning NLP

(10+ 13+ 6)/3 (8+8+4)/3



Vs to V'

10	2	8
2	15	3
5	1	5

13	4	8
4	5	2
1	45	31

6	3	4
1	7	1
3	4	0

9.6	?	6.6
?	?	?
?	?	?

Element-wise Average

Data Transformation for Deep Learning NLP

Vs to V'

10	2	8
2	15	3
5	1	5

13	4	8
4	5	2
1	45	31

6	3	4
1	7	1
3	4	0

x $w^1 = 0.2$ x $w^2 = 0.4$ x $w^3 = 0.4$

Element-wise multiplication

2	0.4	1.6
0.4	3	0.6
1	0.2	1.0

+

5.2	1.6	3.2
1.6	2	0.8
0.4	18	12.4

+

2.4	1.2	1.6
0.4	2.8	0.4
1.2	1.6	0

9.6	3.2	6.4
2.4	7.8	1.8
2.6	19.8	13.4

Element-wise summation

Data Transformation for Deep Learning NLP

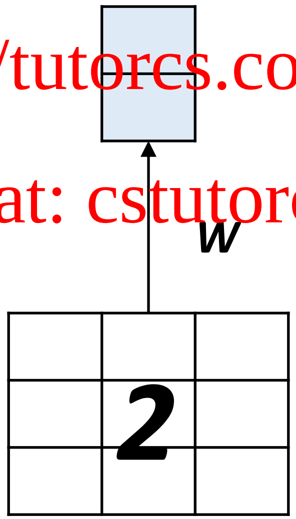
Temporal Summarisation

Context

How to include Temporal information?

4 Data Transformation for Deep Learning NLP

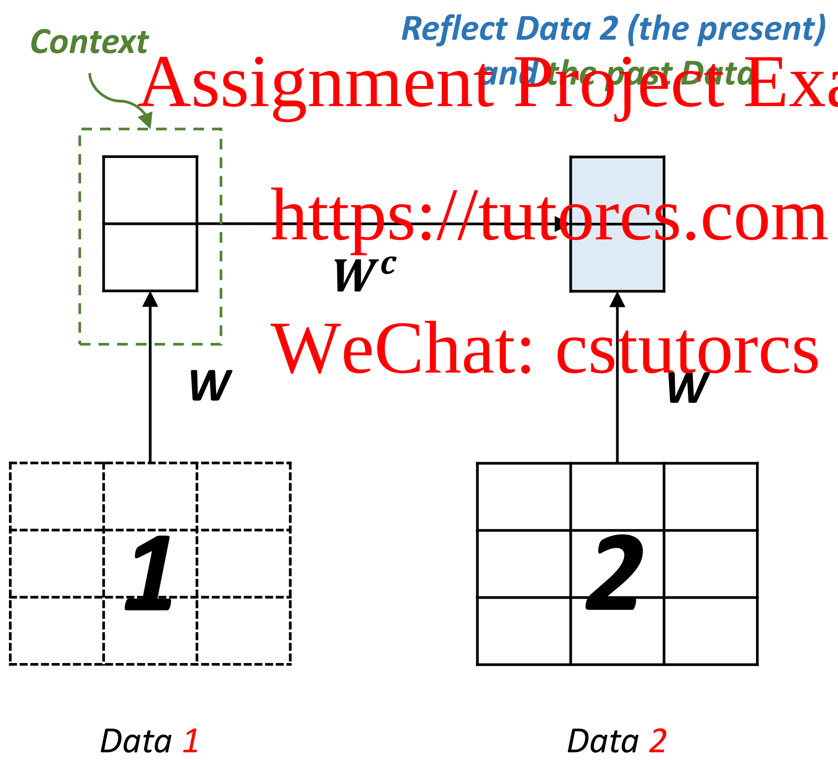
Vs → V's → V'



Data 2

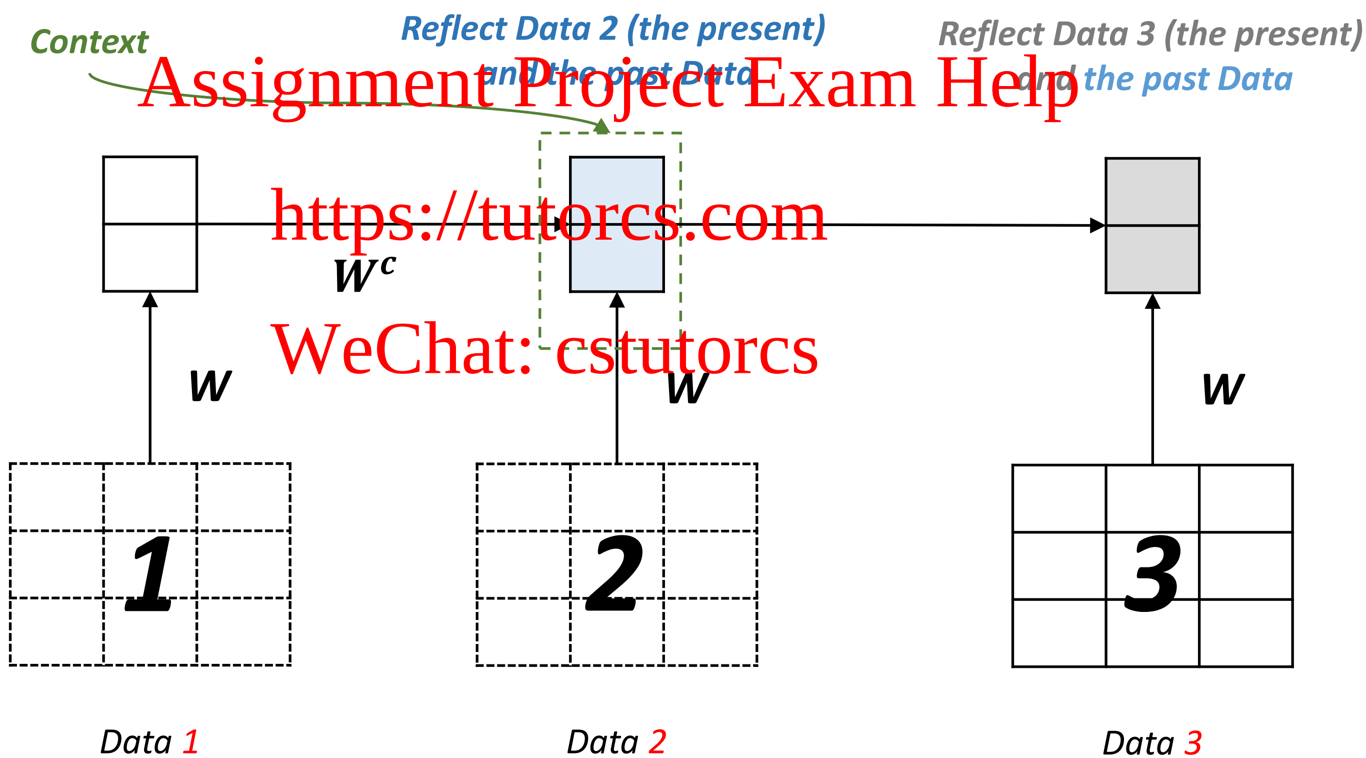
4 Data Transformation for Deep Learning NLP

Vs → V's → V'



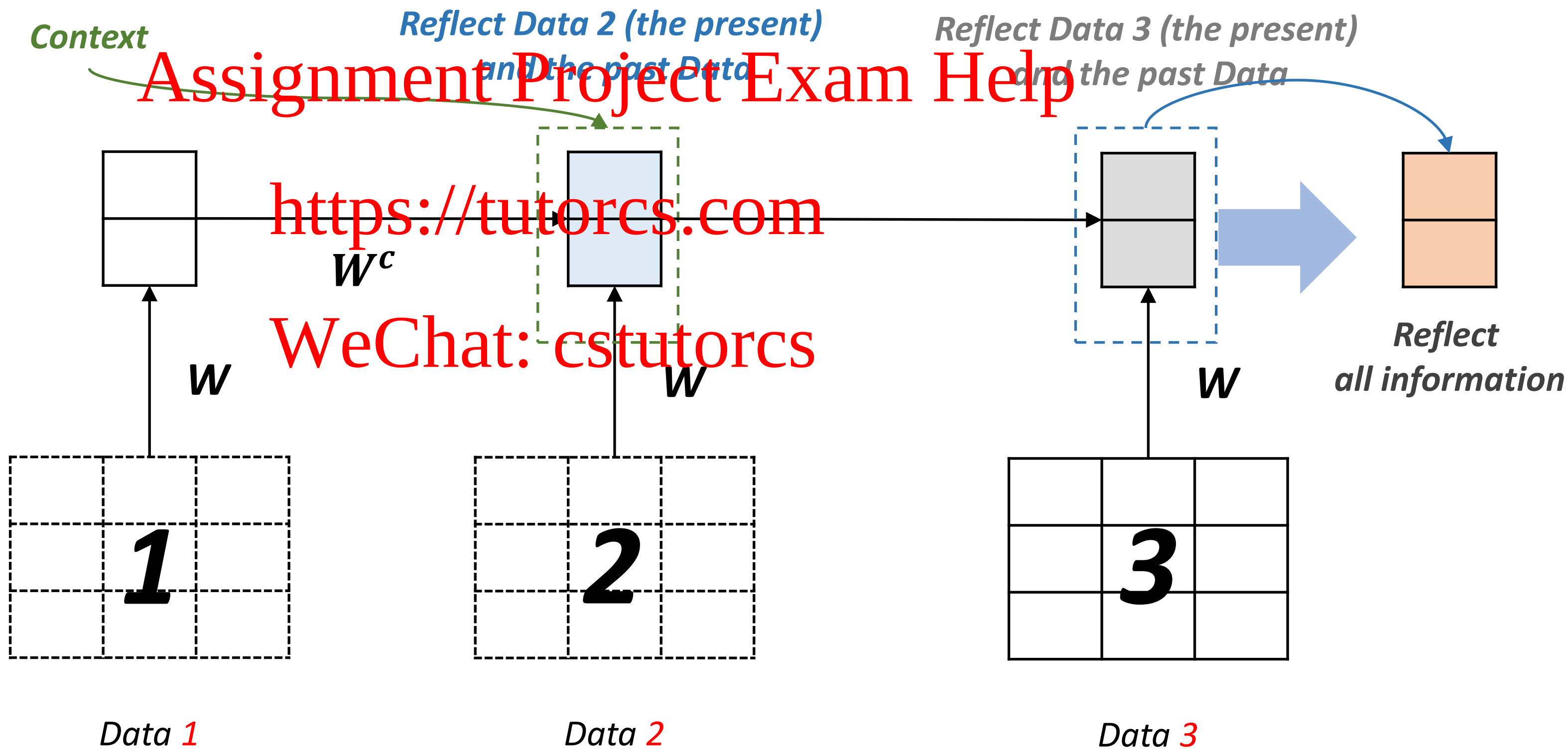
4 Data Transformation for Deep Learning NLP

Vs → V's → V'



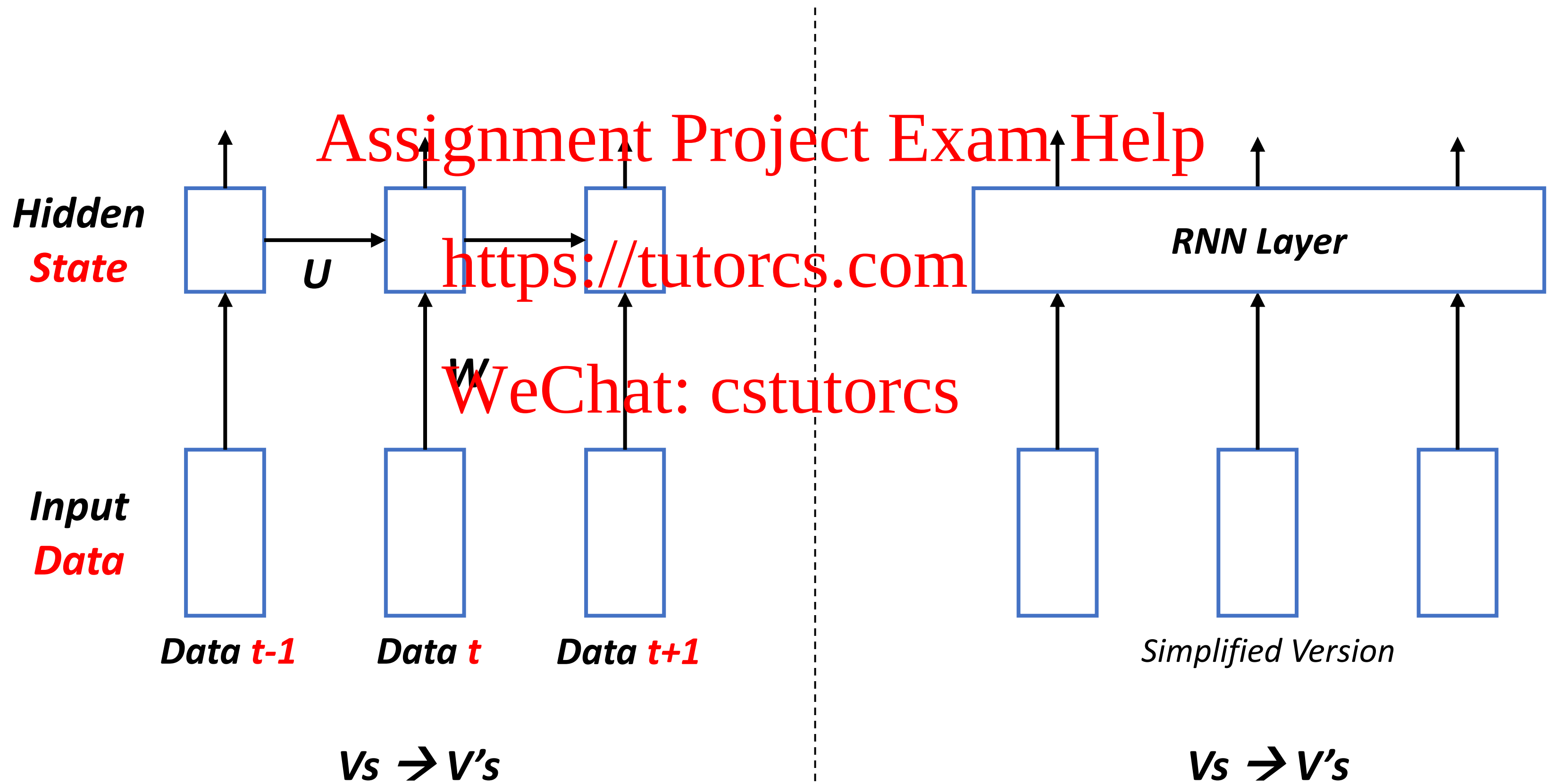
4 Data Transformation for Deep Learning NLP

Vs → V's → V'



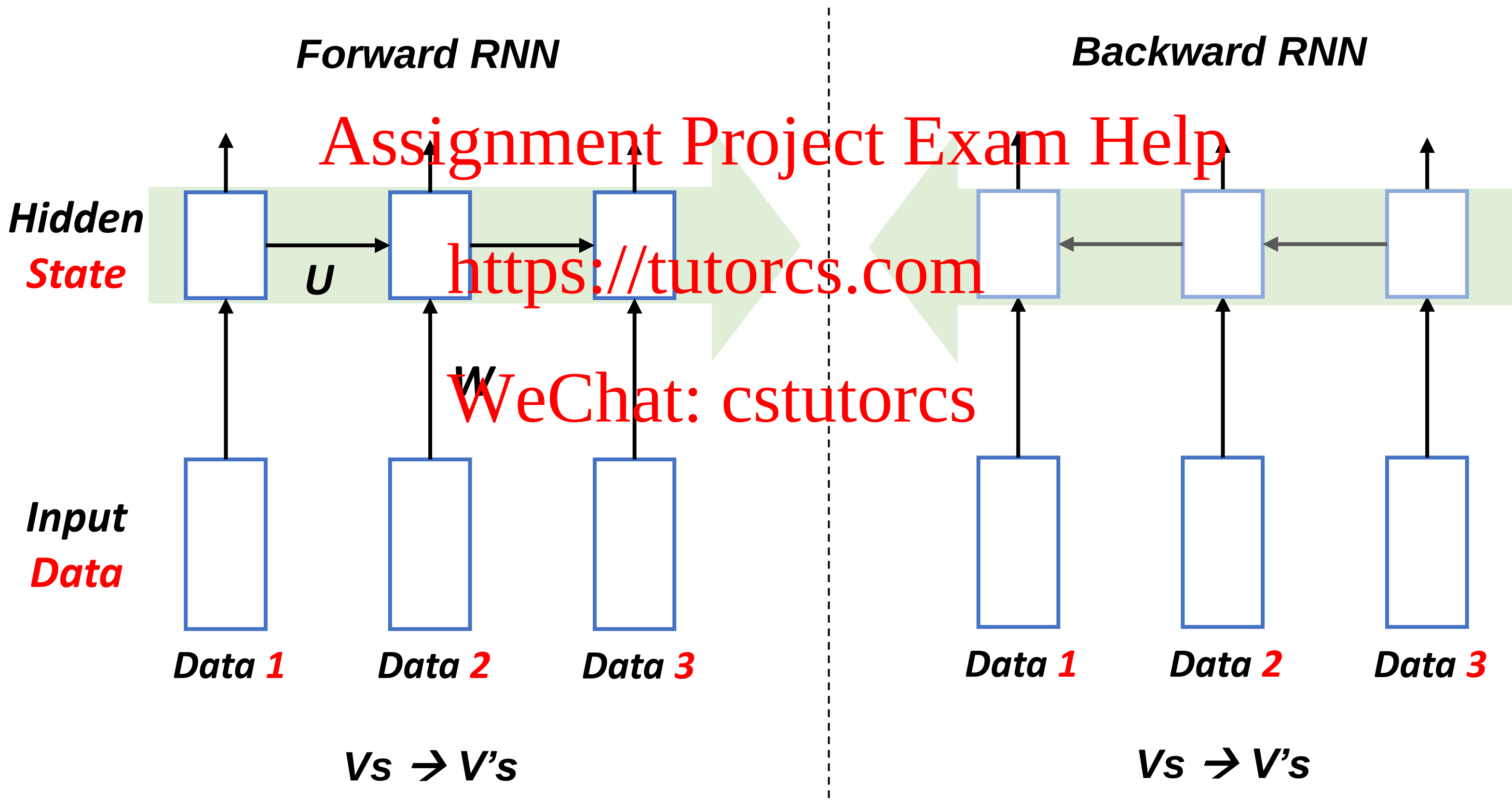
4 Data Transformation for Deep Learning NLP

Graphical Notation

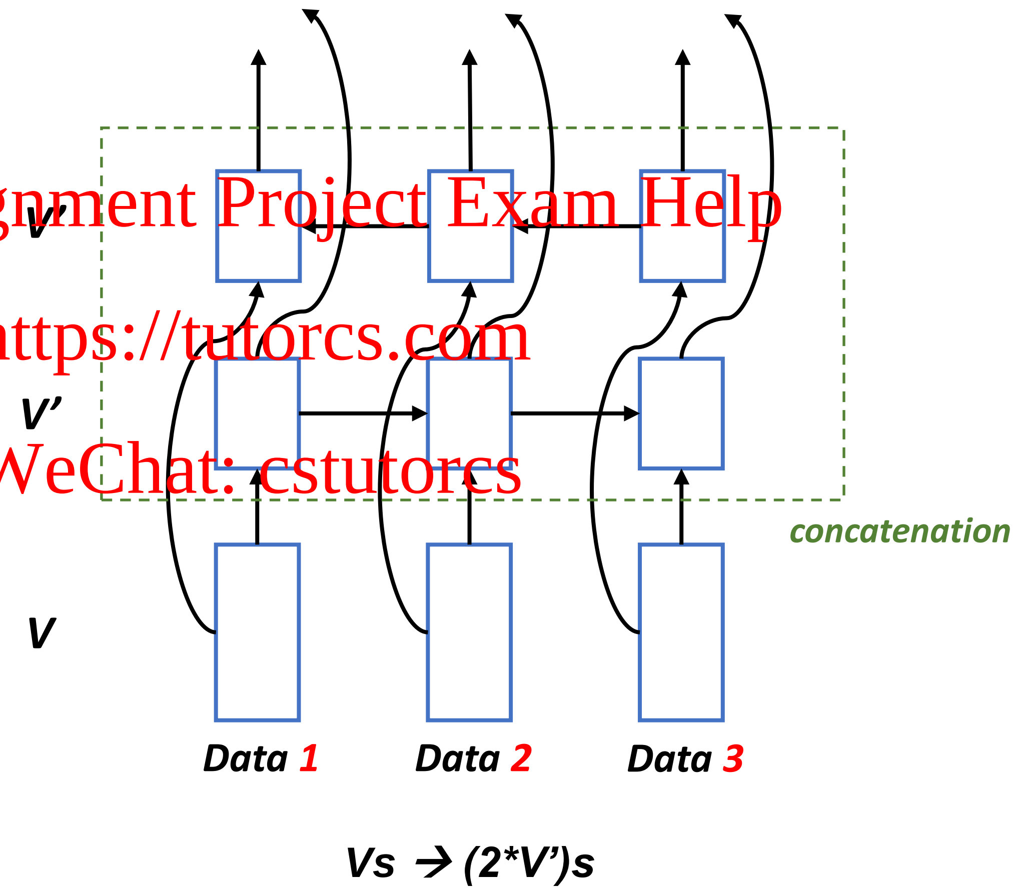


4 Data Transformation for Deep Learning NLP

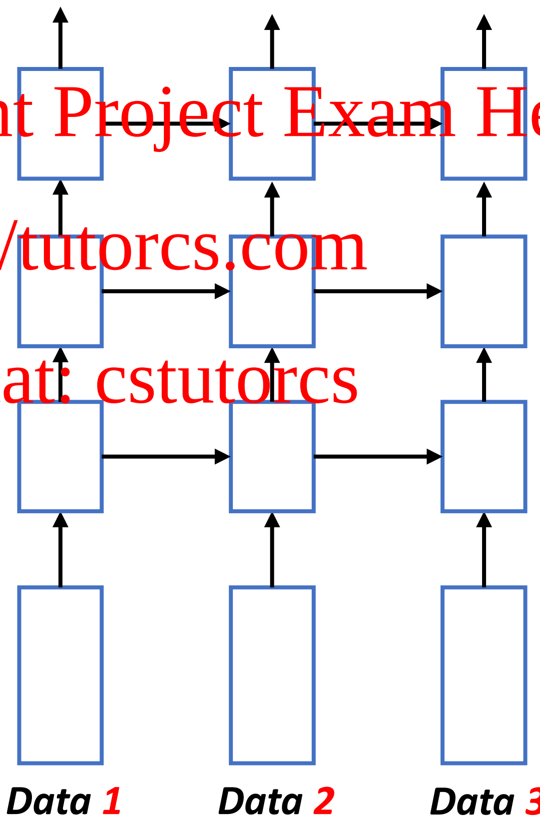
Forward/Backward RNN



Bidirectional RNN



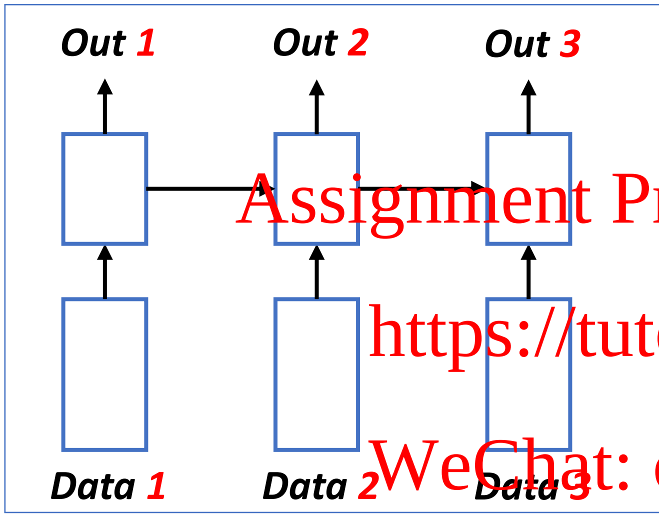
Stacking RNN



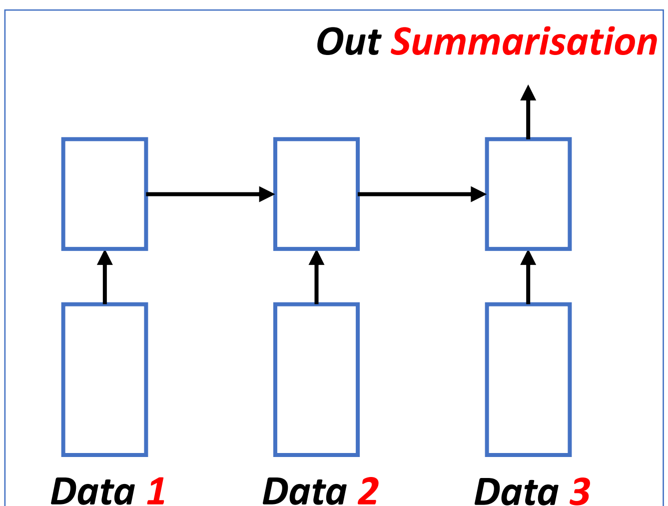
Vs → V's

4 Data Transformation for Deep Learning NLP

RNN: Input and Output



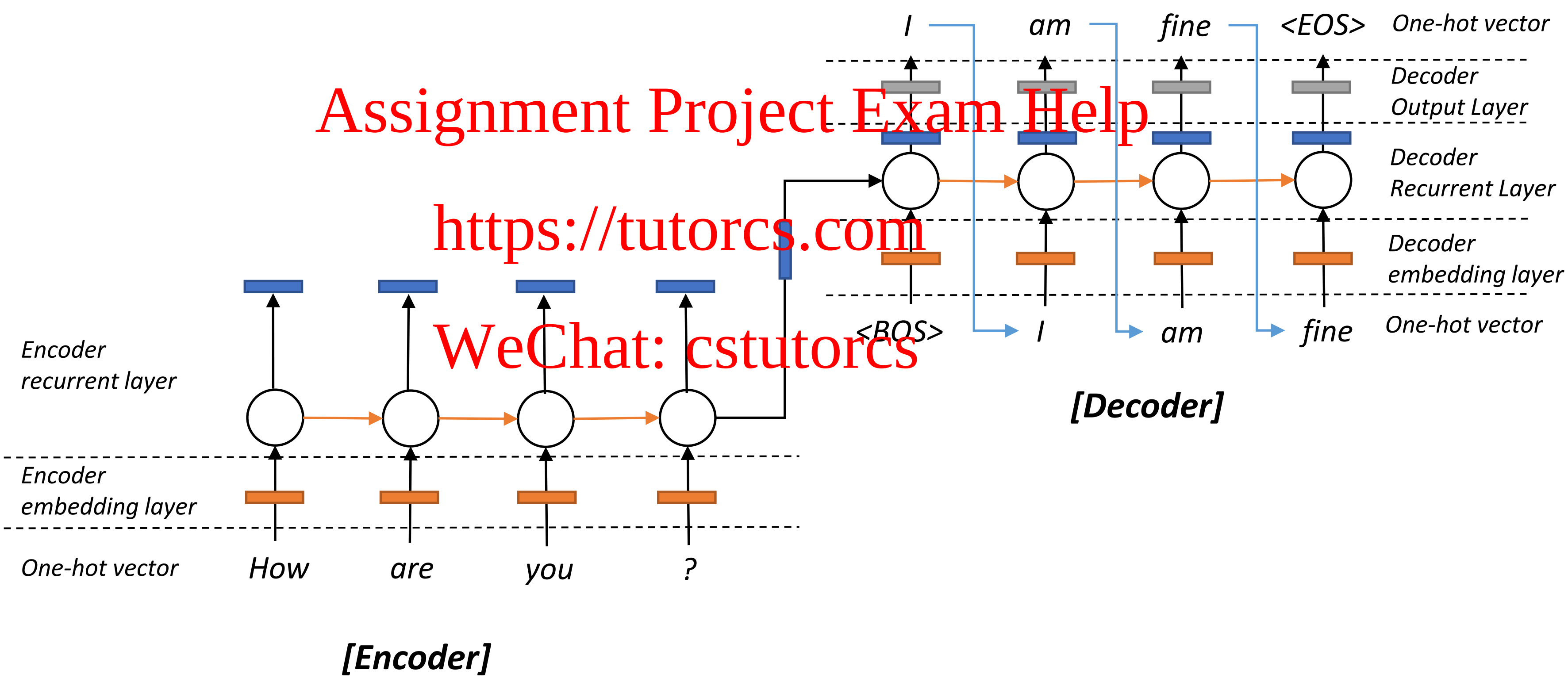
- ✓ *Vs → V's*
- ✓ *Len(Vs) → Len(V's)*



- ✓ *Vs → 1*

Data Transformation for Deep Learning NLP

Seq2Seq Encoding and Decoding- Dialog System



0 LECTURE PLAN

Lecture 4: Word Classification and Machine Learning 2

1. Machine Learning and NLP: Finish
2. Seq2Seq Learning
3. Seq2Seq Deep Learning
    1. RNN (Recurrent Neural Network)
    2. LSTM (Long Short-Term Memory)
    3. GRU (Gated Recurrent Unit)
4. Data Transformation for Deep Learning NLP
5. **Next Week Preview**
    - Natural Language Processing Stack

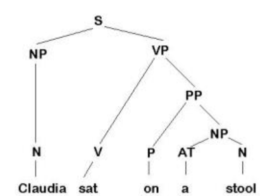
Next Week Preview

The purpose of Natural Language Processing: Overview

Application

Understanding			***Searching***
		Dialog	
Sentiment Analysis	***Topic Classification***	***Translation***	***Search***
Topic Modelling	***....***		***....***

NLP Stack

Entity Extraction	When Sebastian Thrun …	When Sebastian Thrun PERSON started at Google ORG in 2007 DATE
Parsing	Claudia sat on a stool	
PoS Tagging	She sells seashells	**[she/PRP] [sells/VBZ] [seashells/NNS]**
Stemming	Drinking, Drank, Drunk	**Drink**
Tokenisation	How is the weather today	**[How] [is] [the] [weather] [today]**

Reference

Reference for this lecture

- Deng, L., & Liu, Y. (Eds.). (2018). Deep Learning in Natural Language Processing. Springer.
- Rao, D., & McMahan, B. (2019). Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning. " O'Reilly Media, Inc.".
- Manning, C. D., Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press.
- Blunsom, P 2017, Deep Natural Language Processing, lecture notes, Oxford University
- Manning, C 2017, Natural Language Processing with Deep Learning, lecture notes, Stanford University
- Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Grue Simonsen, J., & Nie, J. Y. (2015, October). A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 553-562). ACM.

Figure Reference

- https://towardsdatascience.com/types-of-optimization-algorithms-used-in-neural-networks-and-ways-to-optimize-gradient-95ae5d39529f
- https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21