COMP5046
Natural Language Processing

Lecture 4: Word Classification and Machine Learning 2

Semester 1, 2020

School of Computer Science

The University of Sydney, Australia

THE UNIVERSITY OF SYDNEY

Dr Caren Han

Caren.Han@sydney.edu.au

0 LECTURE PLAN

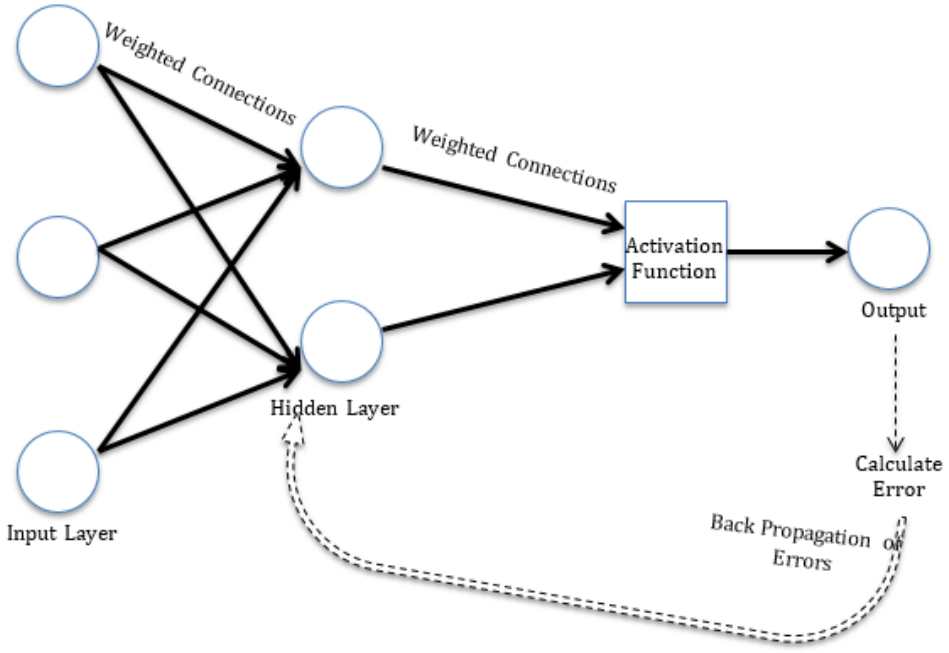
THE UNIVERSITY OF SYDNEY

Lecture 4: Word Classification and Machine Learning 2

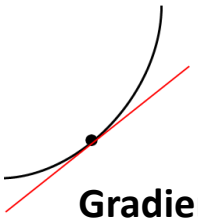
1. **Machine Learning and NLP: Finish**
2. Seq2Seq Learning
3. Seq2Seq Deep Learning
    1. RNN (Recurrent Neural Network)
    2. LSTM (Long Short-Term Memory)
    3. GRU (Gated Recurrent Unit)
4. Data Transformation for Deep Learning NLP
5. Next Week Preview
    - Natural Language Processing Stack

Machine Learning and NLP

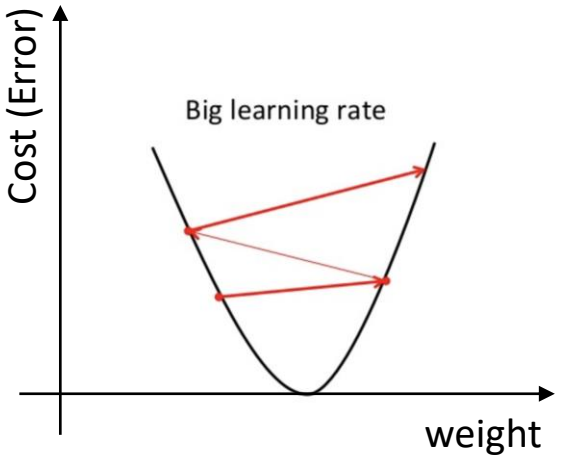
Prediction – Review (Back propagation)

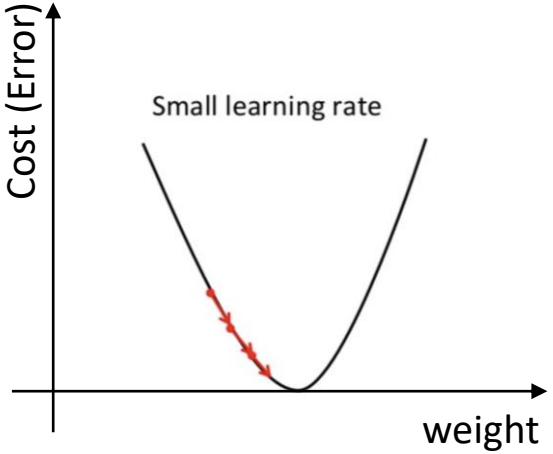


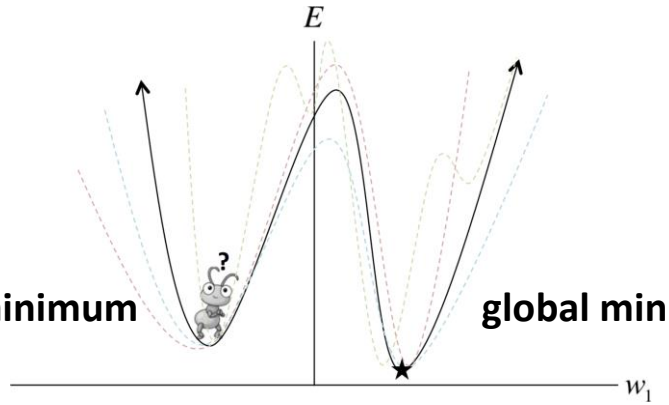
Prediction – Review (Gradient)



Gradient
= slope

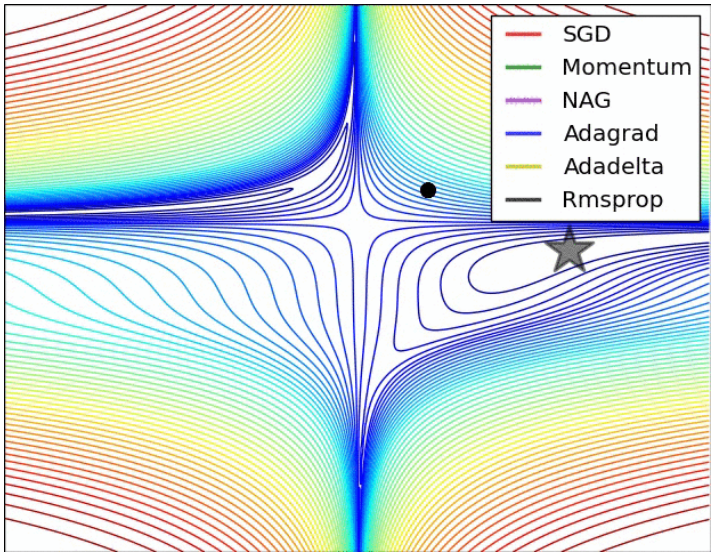




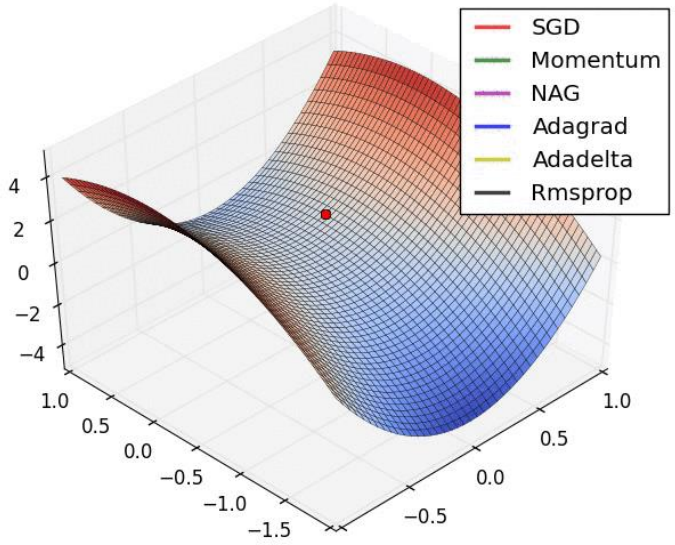


Local minimum **global minimum**

Prediction – Review (Optimizer)



Optimisation on loss surface contours

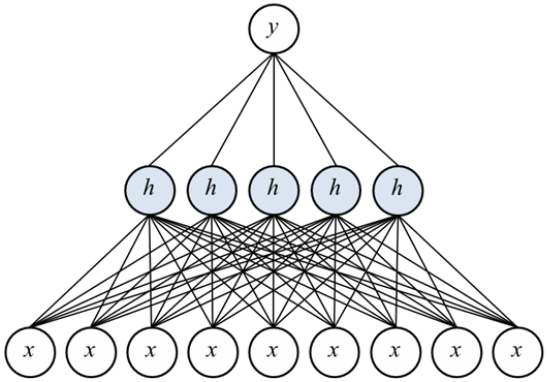


Optimisation on saddle points

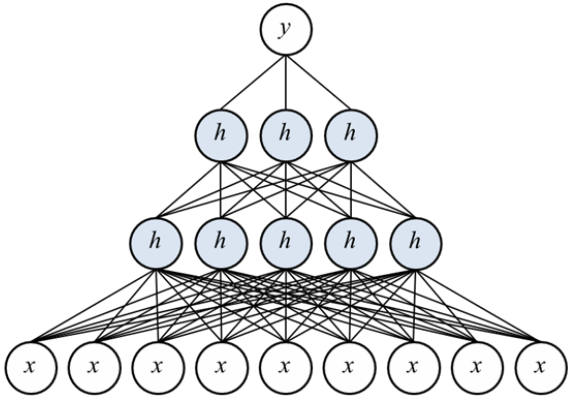
There are different types of Gradient descent Optimization algorithm

SGD (Stochastic gradient descent), Momentum, NAG(Nesterov accelerated gradient), Adagrad, Adadelta, Adam,..

Prediction – Review (neural networks)



Single Layer



Multilayer

Machine Learning and NLP

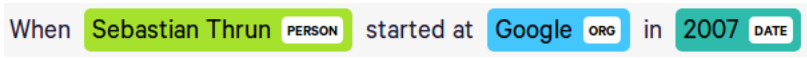
The purpose of Natural Language Processing: Overview

Application

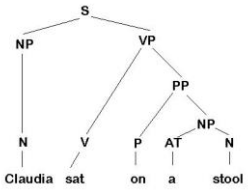
- ***Understanding***
  - ***Sentiment Analysis***
  - ***Topic Classification***
  - ***Topic Modelling***
  - ***....***
- ***Searching***
  - ***Search***
  - ***....***
- ***Dialog***
- ***Translation***

NLP Stack

- ***Entity Extraction*** — When Sebastian Thrun …



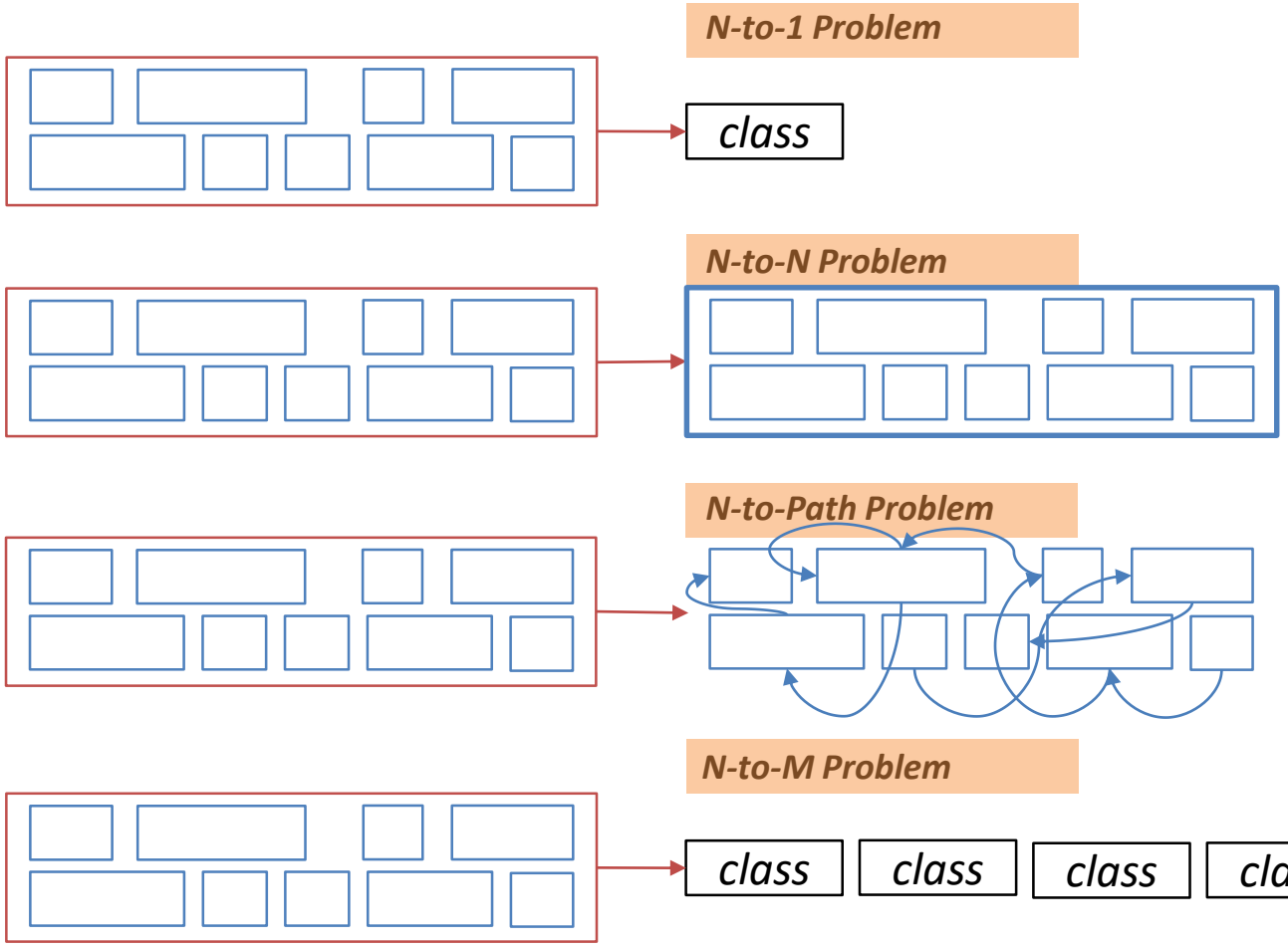
- ***Parsing*** — Claudia sat on a stool



- ***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]**

Machine Learning and NLP

Problem Abstraction



Sentiment Analysis

Topic Classification

PoS Tagging

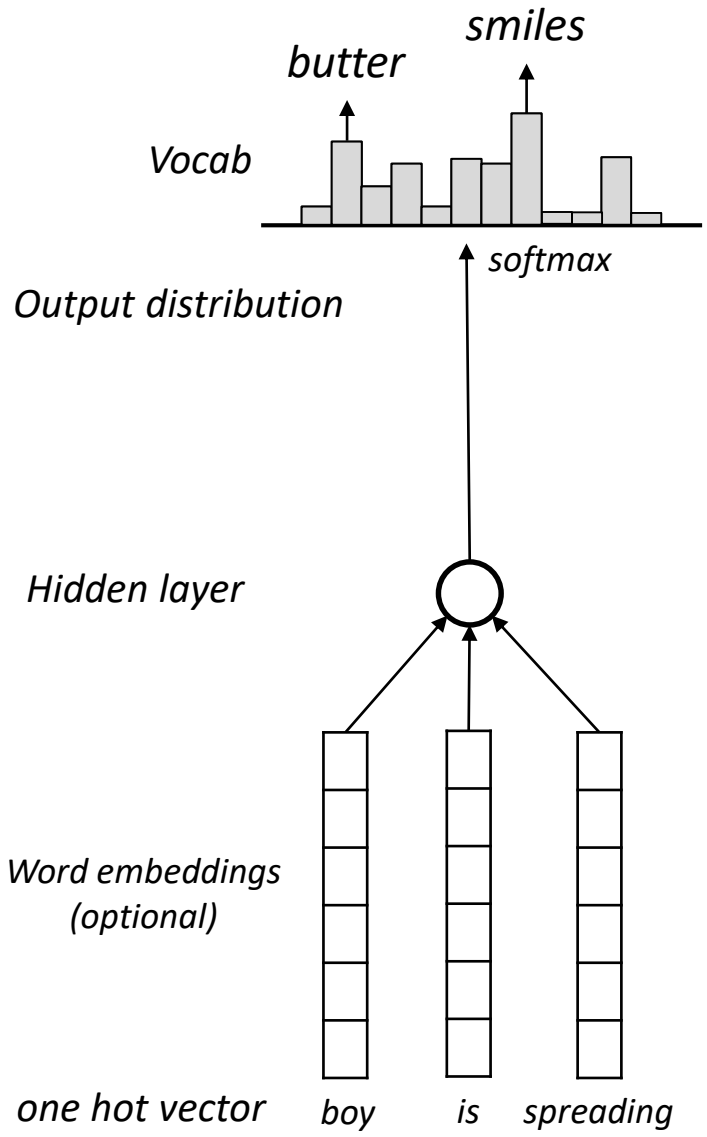
Parsing

Dialog

Translation

Machine Learning and NLP

Prediction



0 LECTURE PLAN

THE UNIVERSITY OF SYDNEY

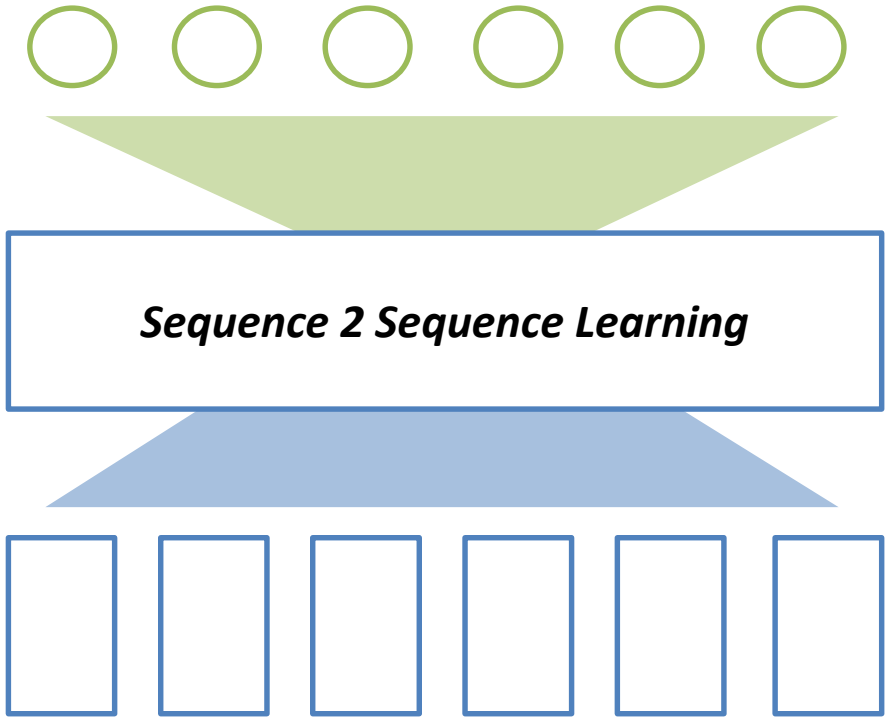
Lecture 4: Word Classification and Machine Learning 2

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

2 Sequence 2 Sequence Learning

THE UNIVERSITY OF SYDNEY

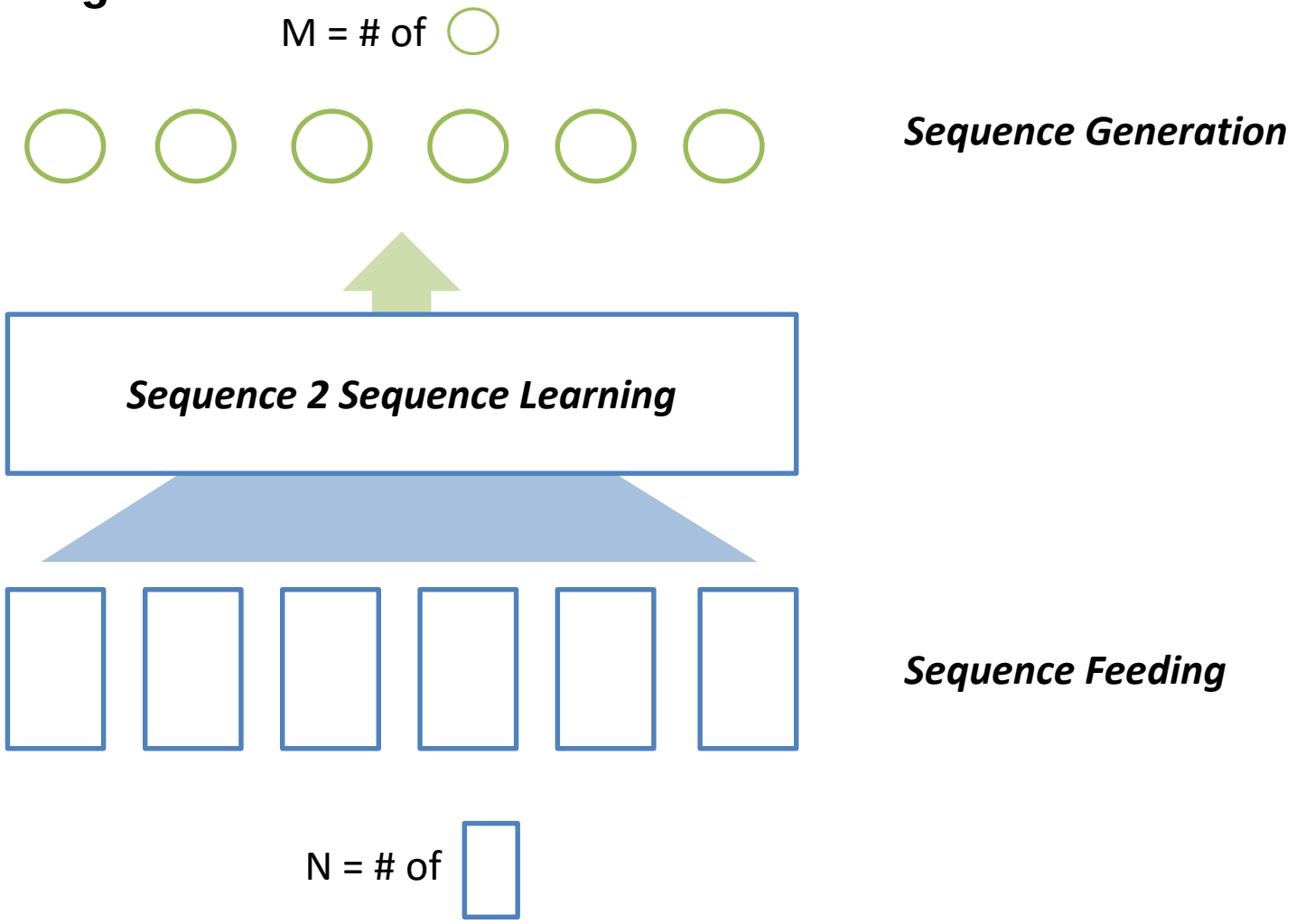
Illustration



2 Sequence 2 Sequence Learning

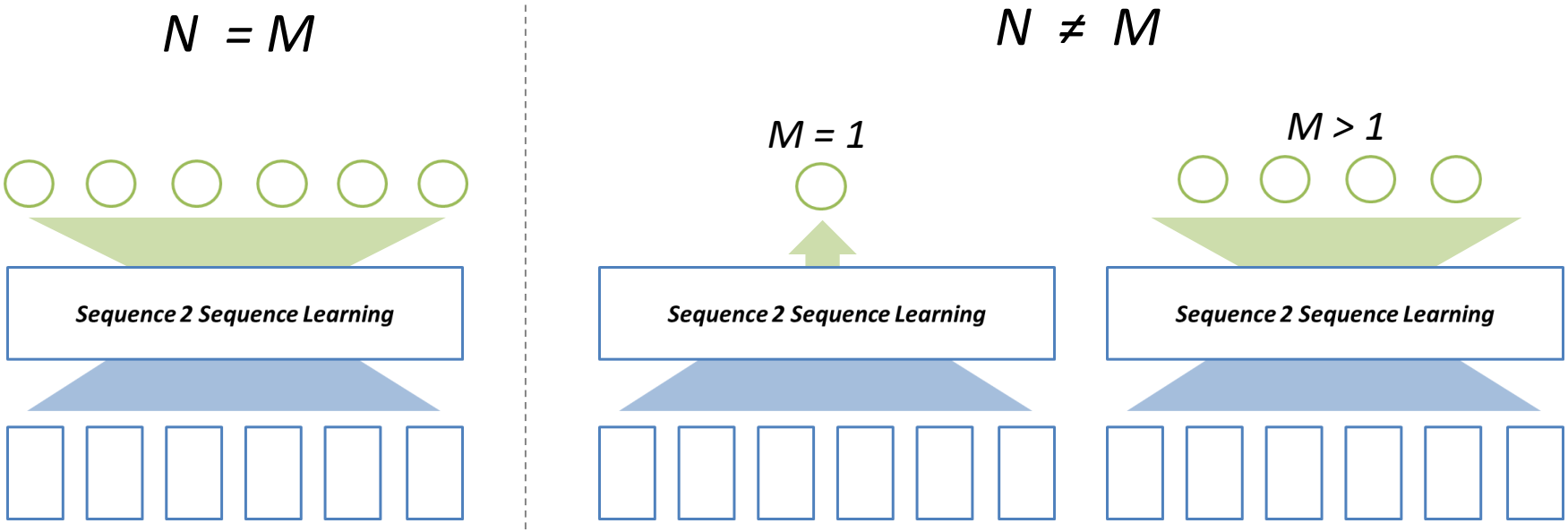
THE UNIVERSITY OF SYDNEY

Running time



Sequence 2 Sequence Learning

Sequence 2 Sequence Learning

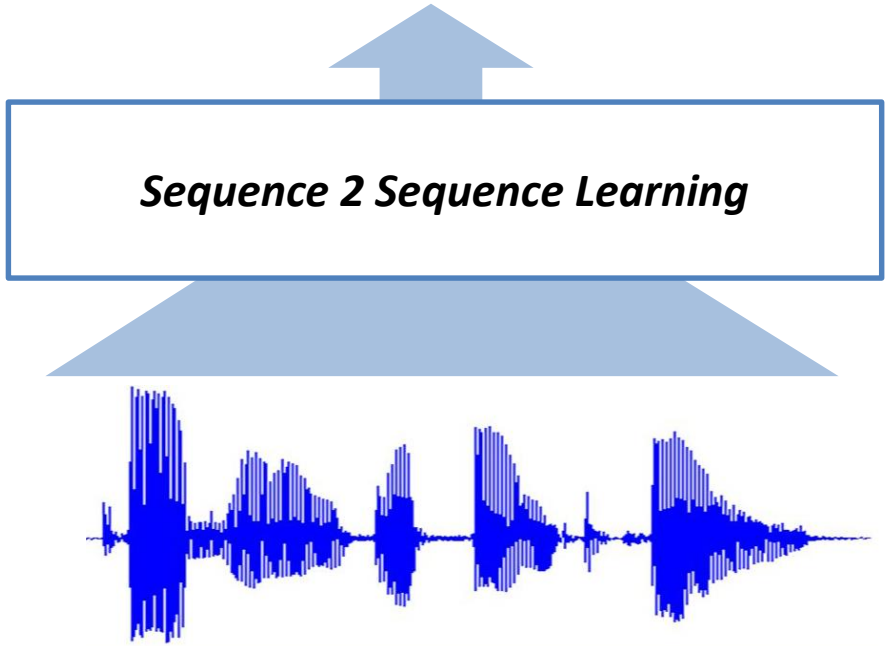


Sequence 2 Sequence Learning

Seq2Seq – Speech Recognition

How is the weather today

Output: Text



Input: Speech Signal

Sequence 2 Sequence Learning

Seq2Seq – Movie Frame Labelling

Swing Swing Hit Bat_Broken

Output: Scene Labels

Sequence 2 Sequence Learning

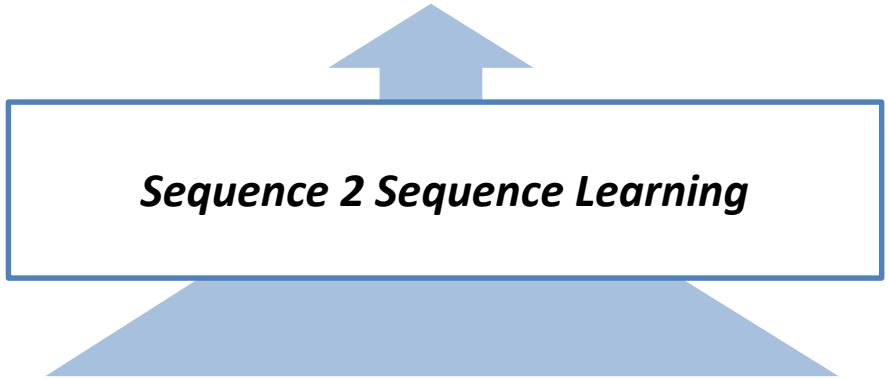
Input: Video Frame

Sequence 2 Sequence Learning

Seq2Seq – PoS Tagging

ADV VERB DET NOUN NOUN

Output: Part of Speech

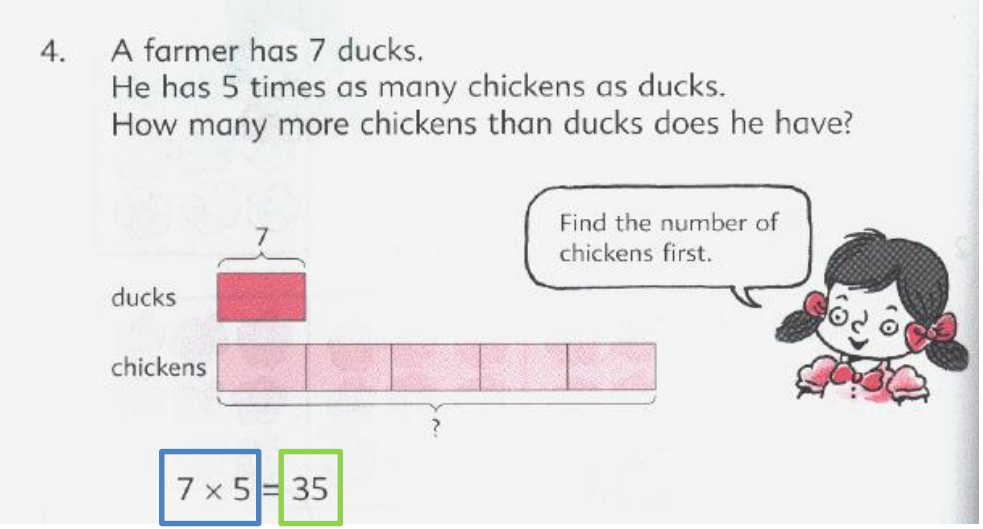


How is the weather today

Input: Text

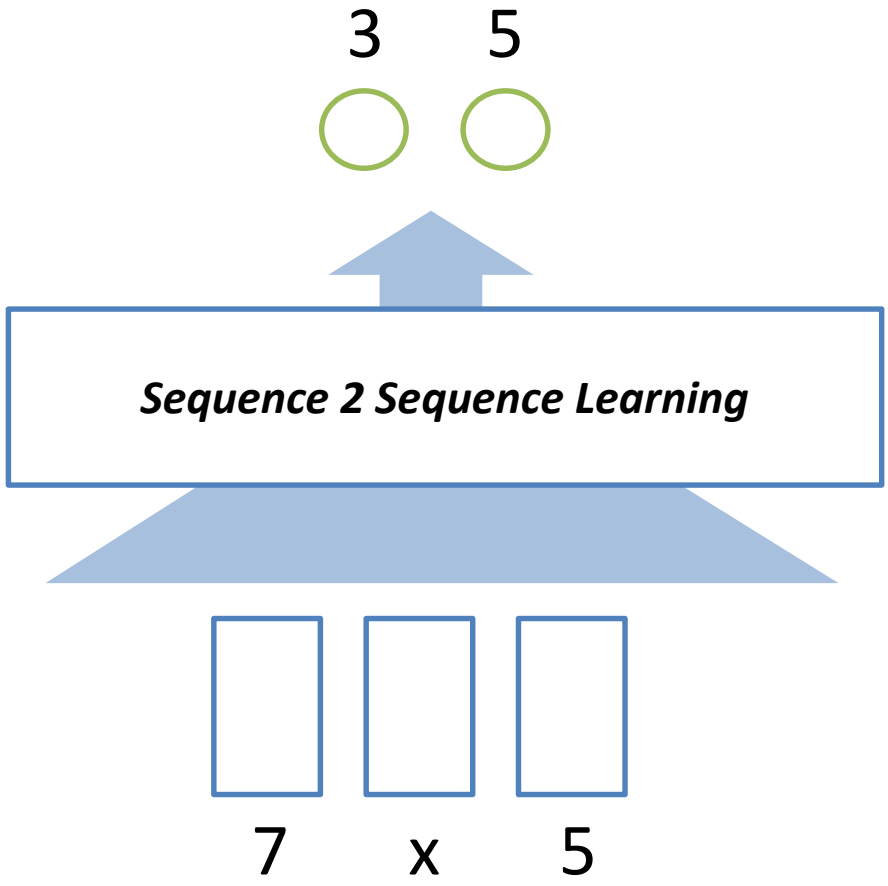
Sequence 2 Sequence Learning

Seq2Seq – Arithmetic Calculation



Sequence 2 Sequence Learning

Seq2Seq – Arithmetic Calculation



Output: Numbers

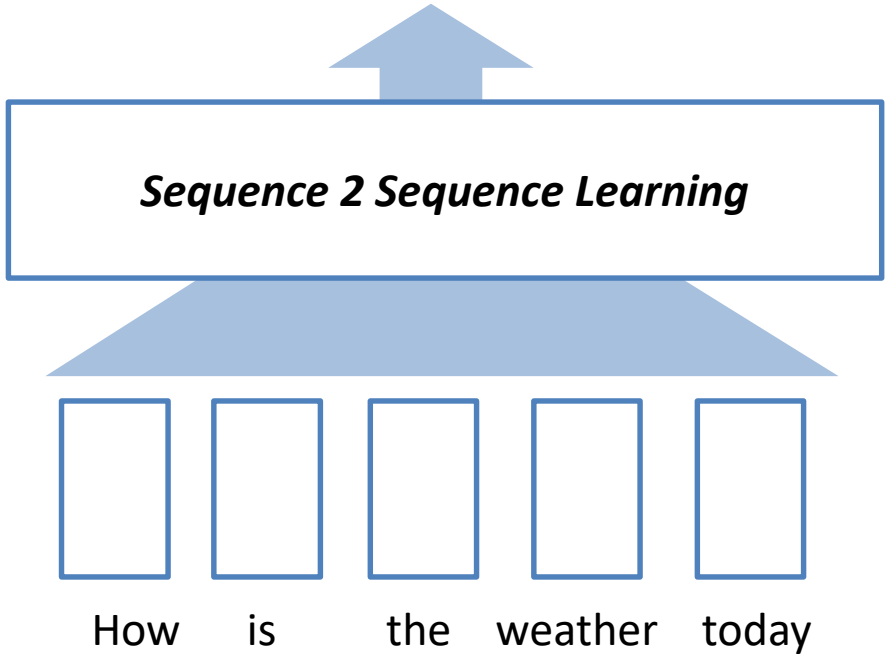
Input: Math Expression

Sequence 2 Sequence Learning

Seq2Seq – Machine Translation

今天　天气　怎么 样？

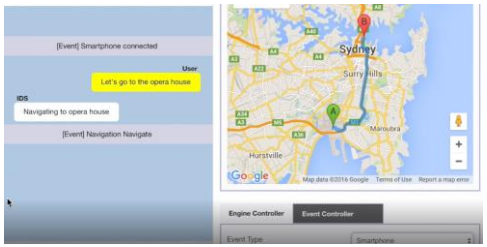
Output: Chinese Text



Input: English Text

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

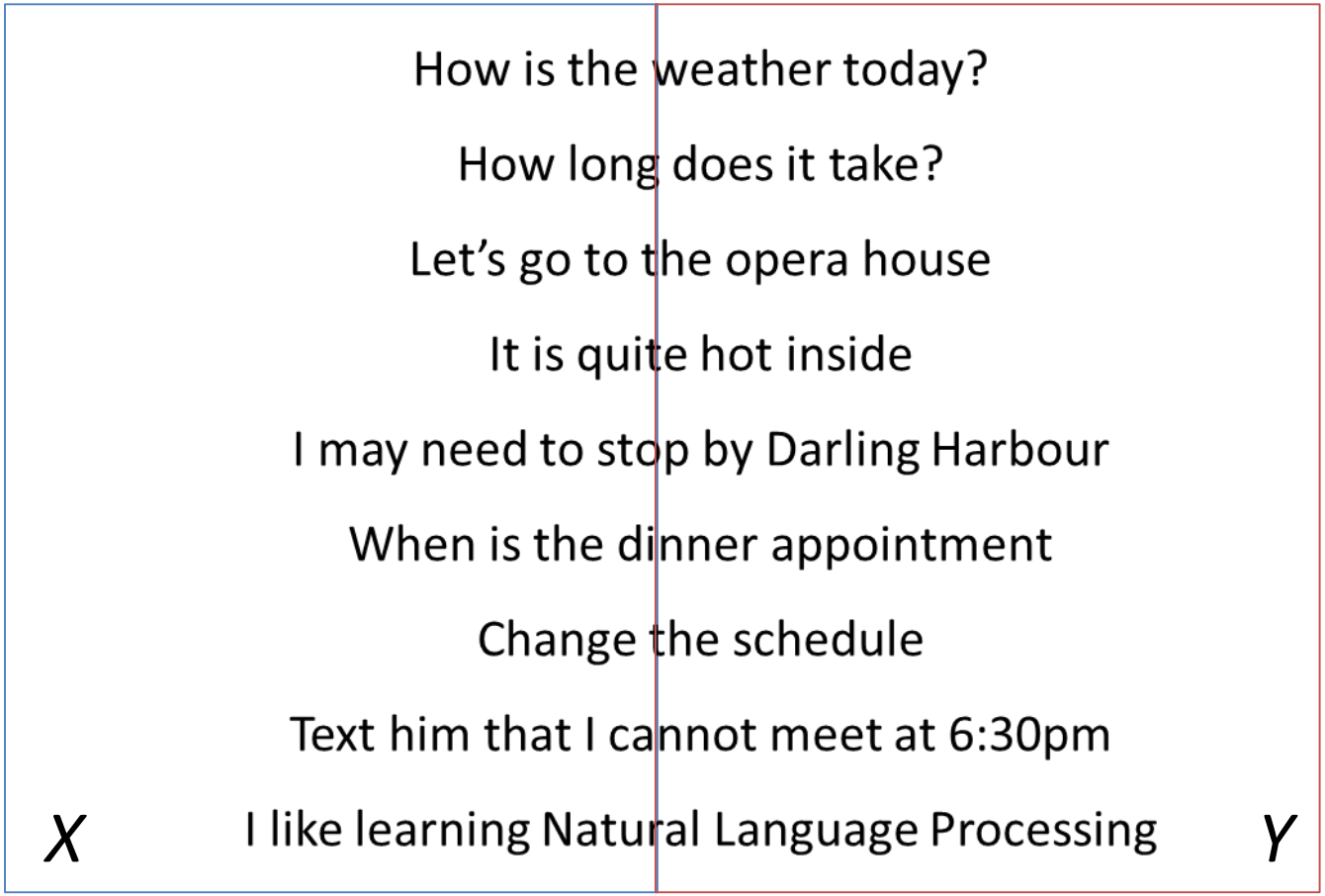
Change the schedule

Text him that I cannot meet at 6:30pm

I like learning Natural Language Processing

Sequence 2 Sequence Learning

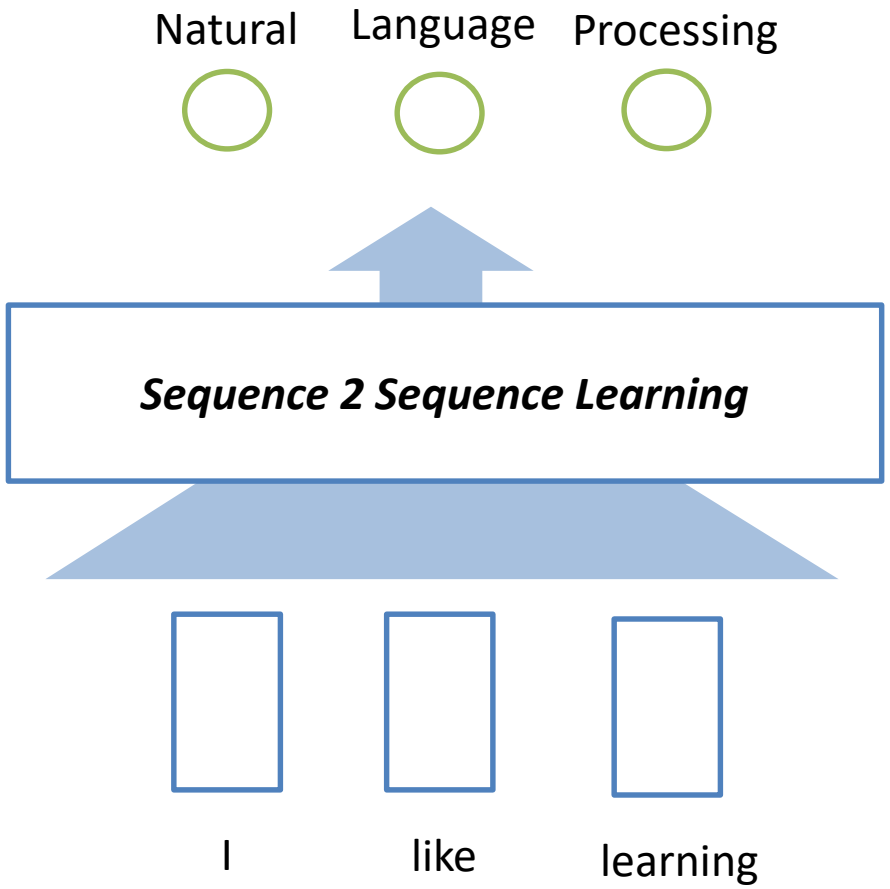
Seq2Seq – Sentence Completion



Sequence 2 Sequence Learning

I like learning | Natural Language Processing

Seq2Seq – Sentence Completion



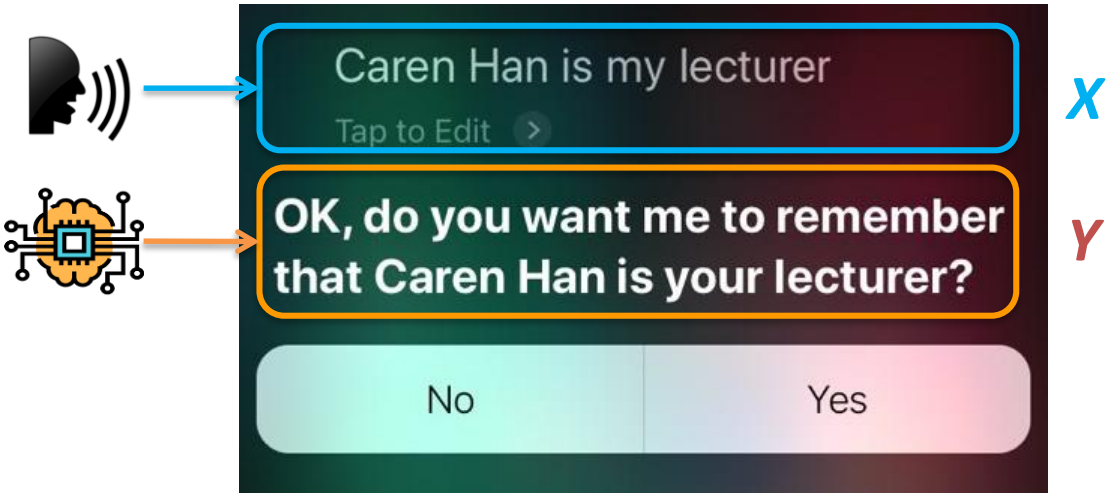
Output: Partial Sentence

Input: Partial Sentence

Sequence 2 Sequence Learning

Seq2Seq – Conversation Modelling

Conversation

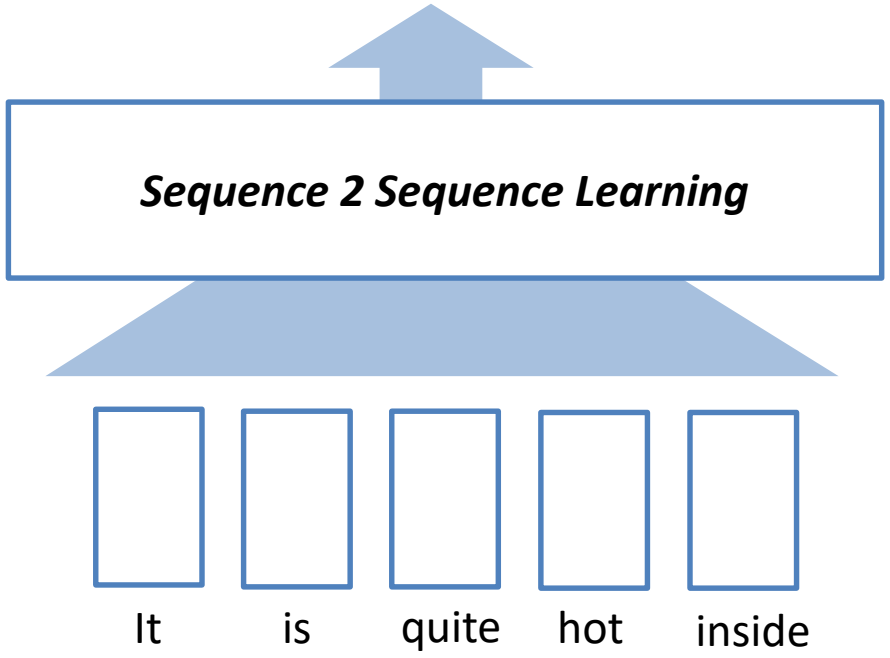


Sequence 2 Sequence Learning

Seq2Seq – Conversation Modelling

Okay. I will open windows for you

Output: Utterance



Input: Utterance

0 LECTURE PLAN

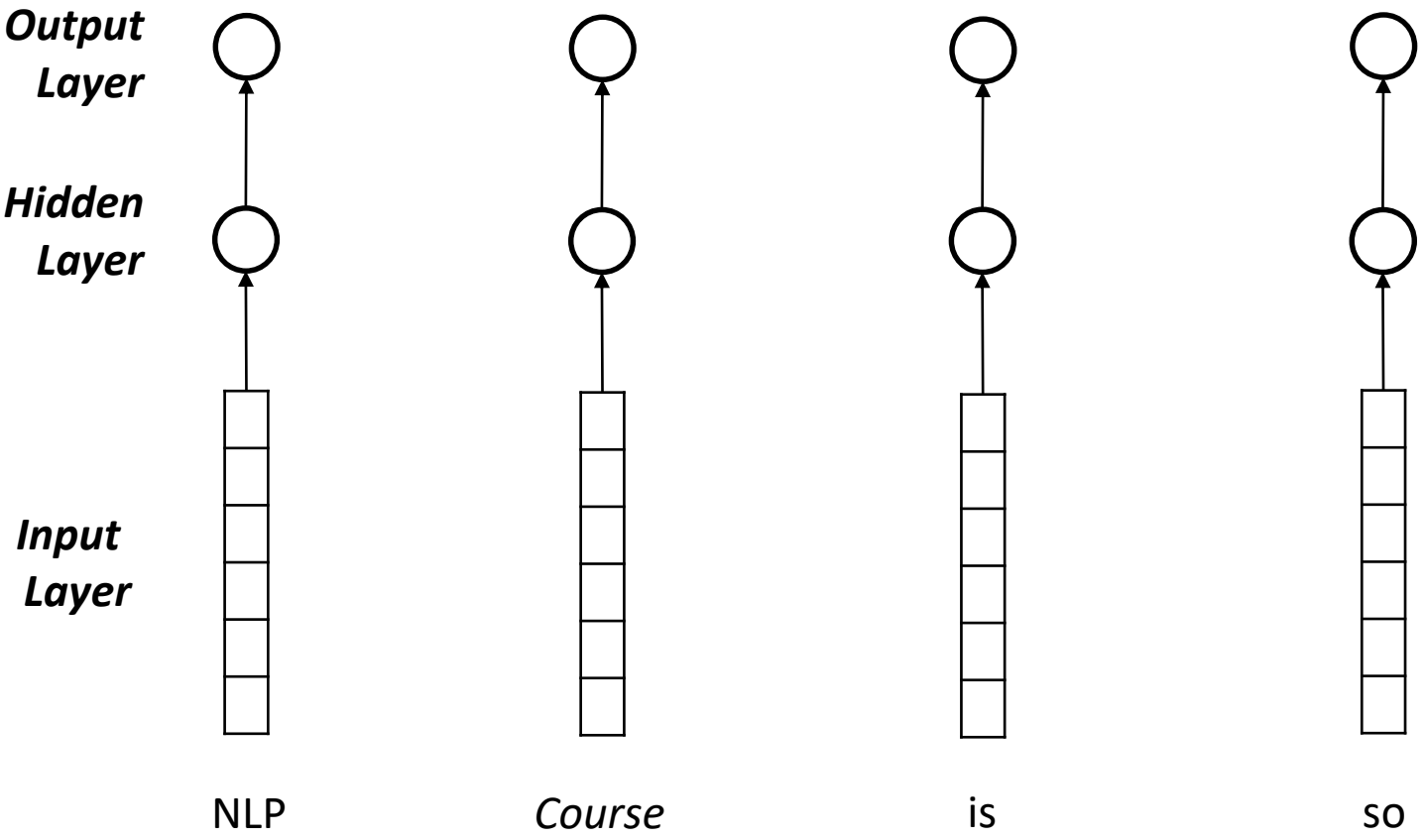
THE UNIVERSITY OF SYDNEY

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

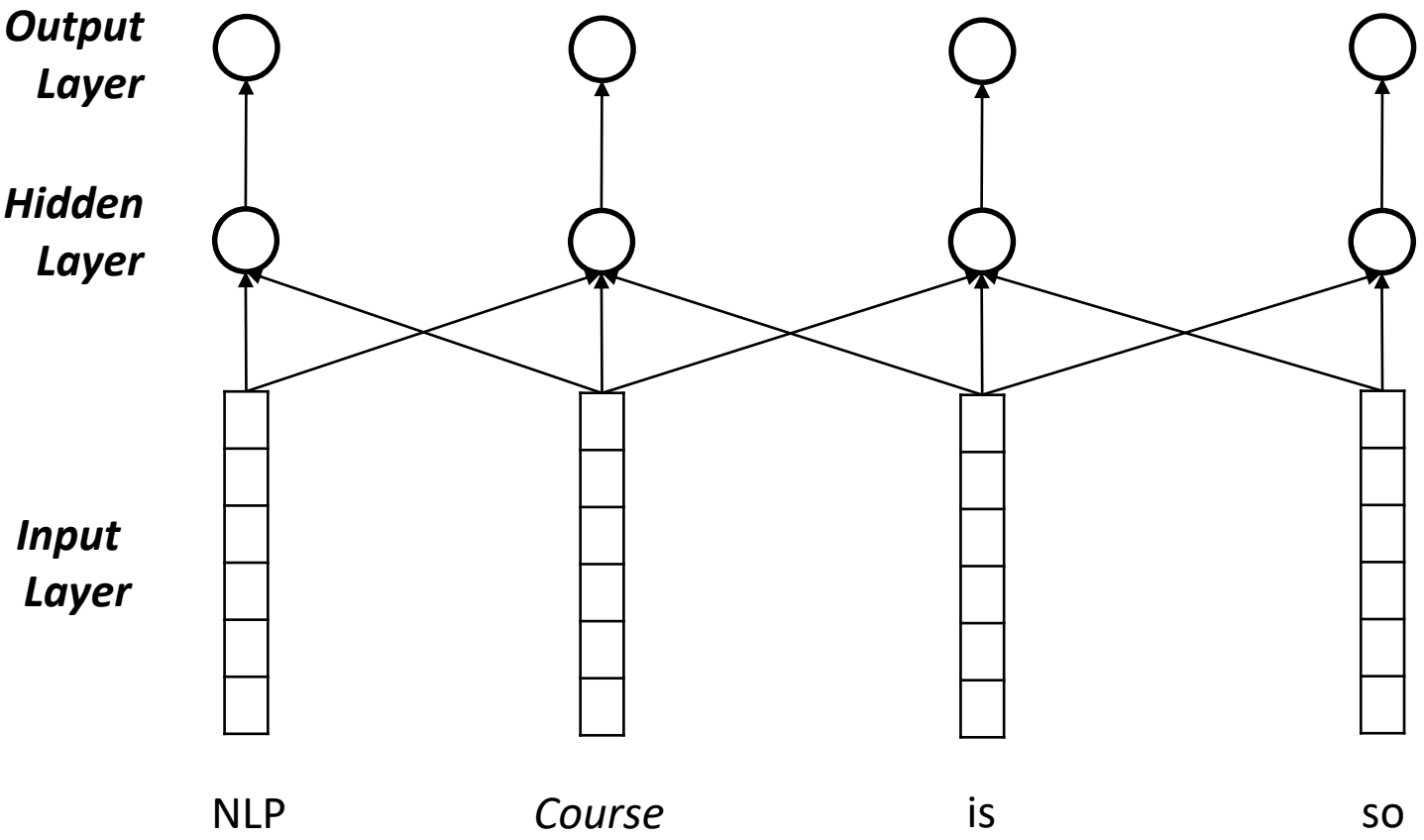
Seq2Seq with Deep Learning

Prediction



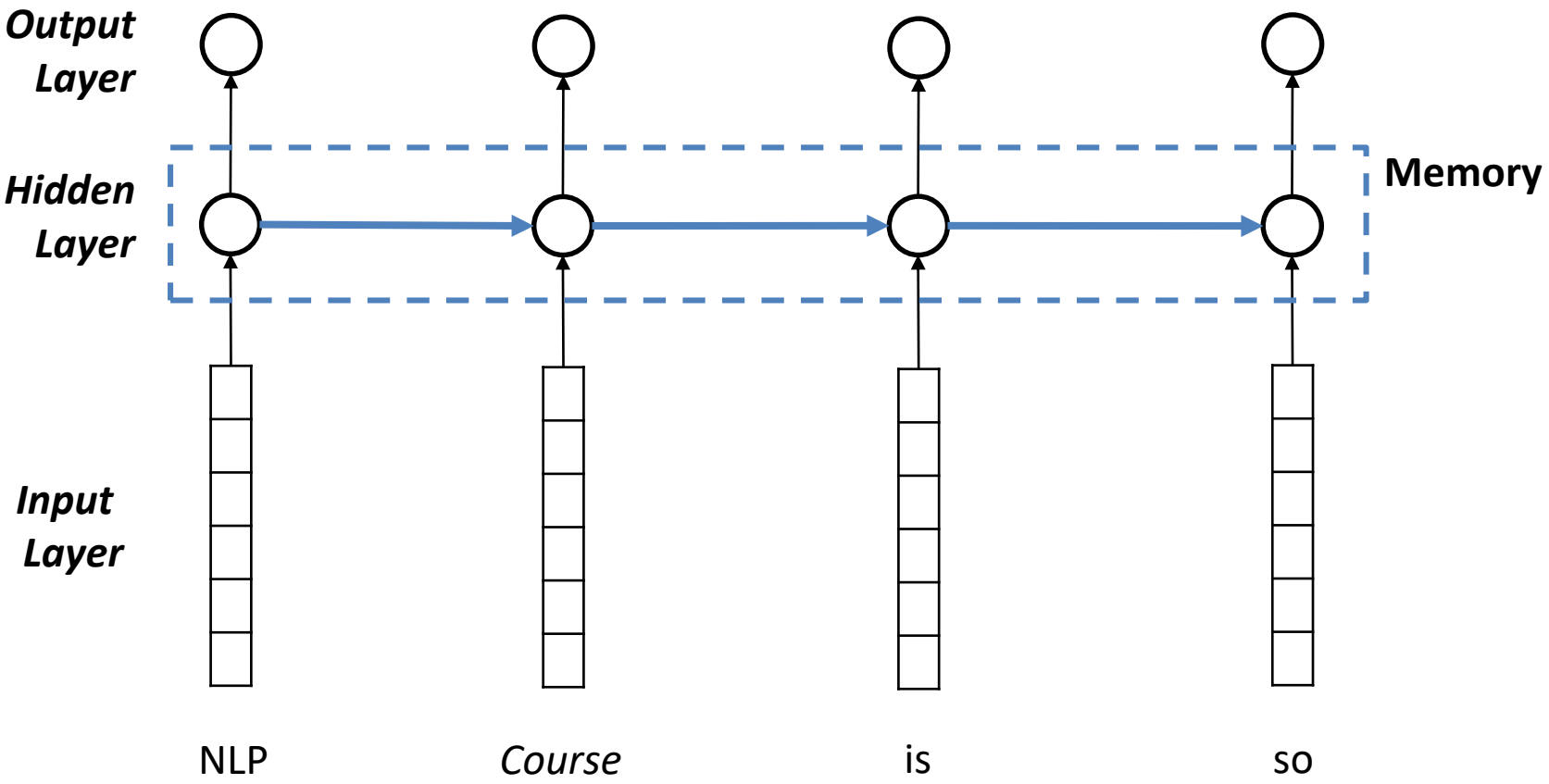
Seq2Seq with Deep Learning

Prediction + Convolution Idea



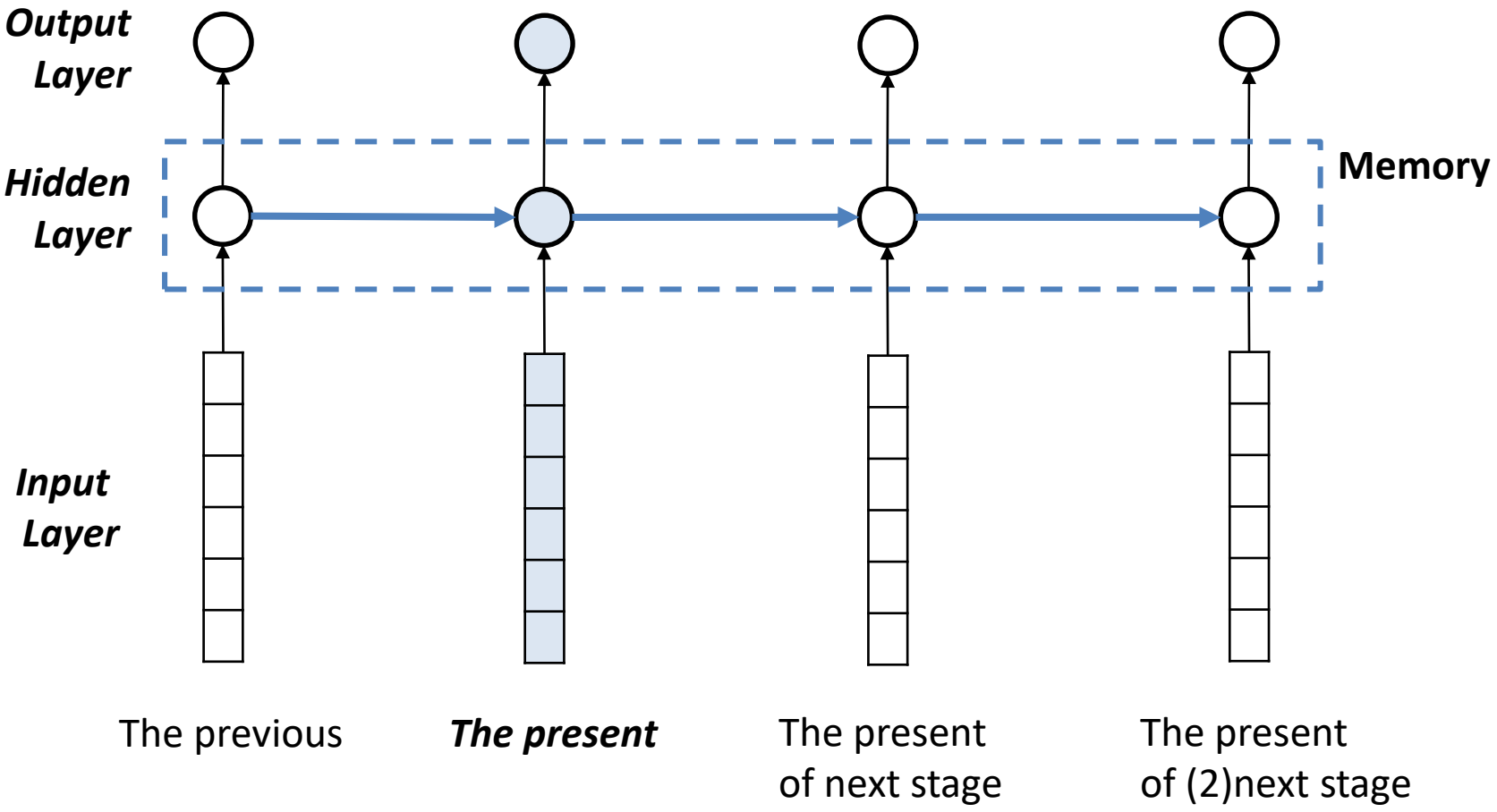
Seq2Seq with Deep Learning

Prediction + Memory = Sequence Modelling



Seq2Seq with Deep Learning

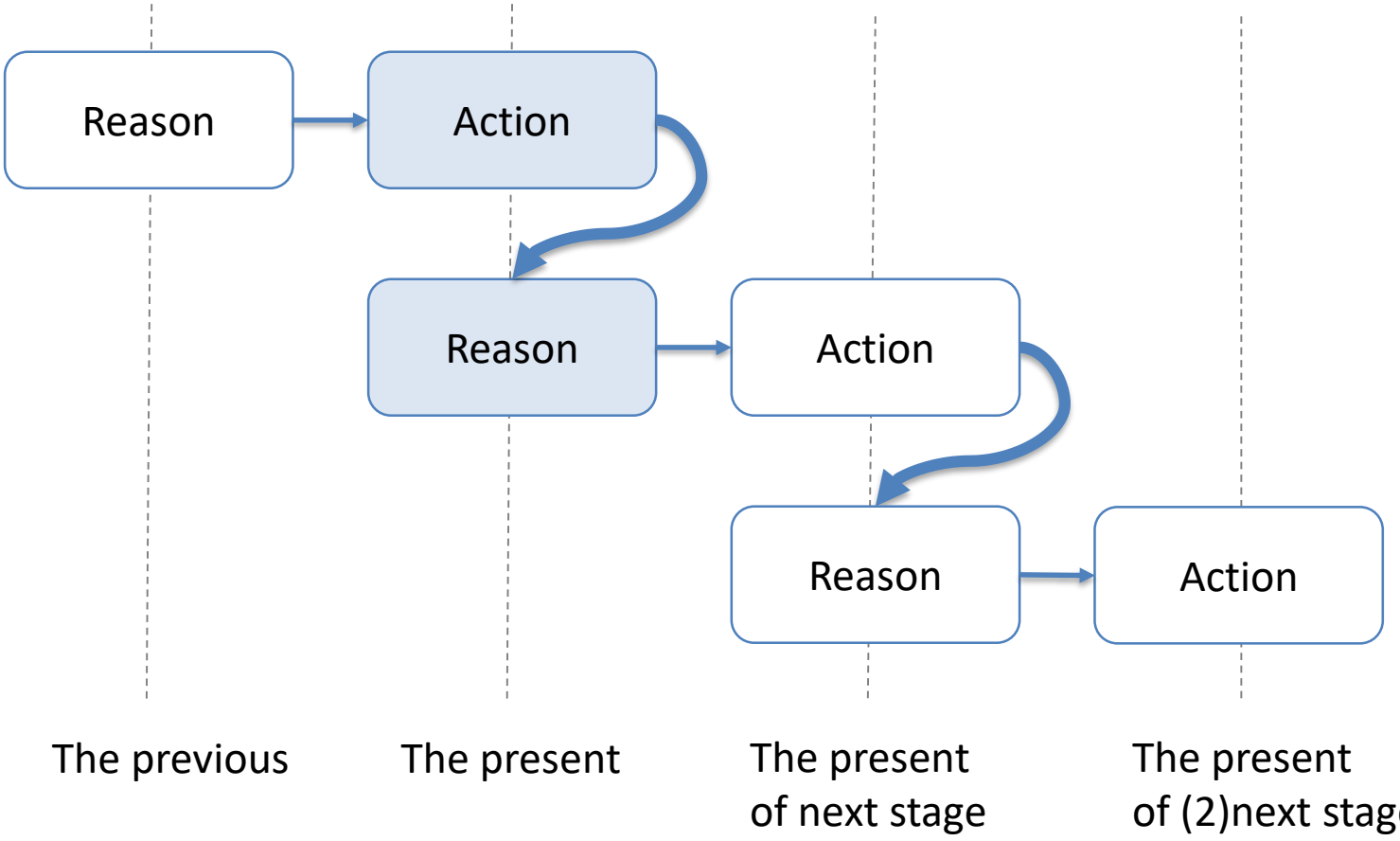
Prediction + Memory = Sequence Modelling



Seq2Seq with Deep Learning

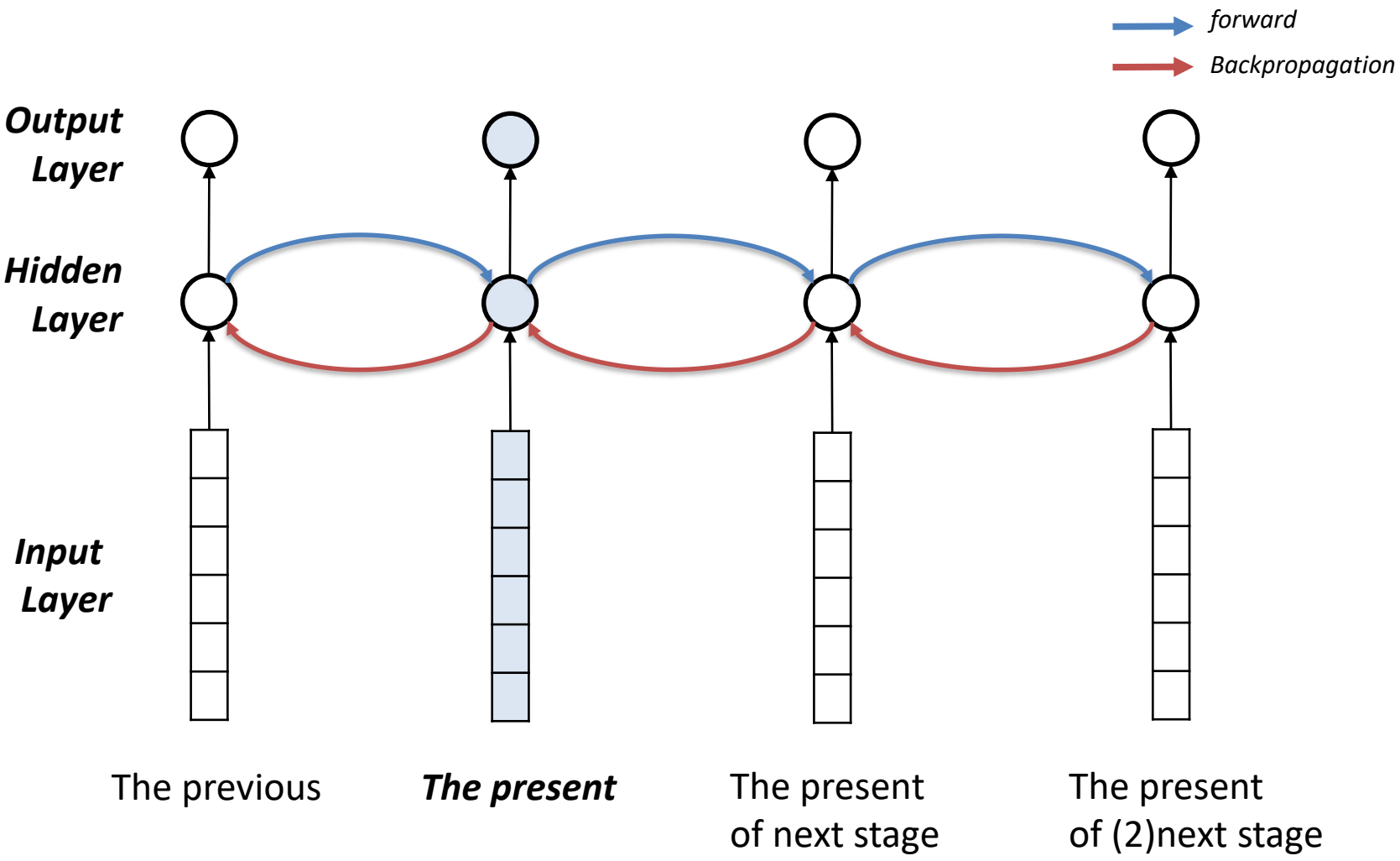
Neural Network + Memory

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



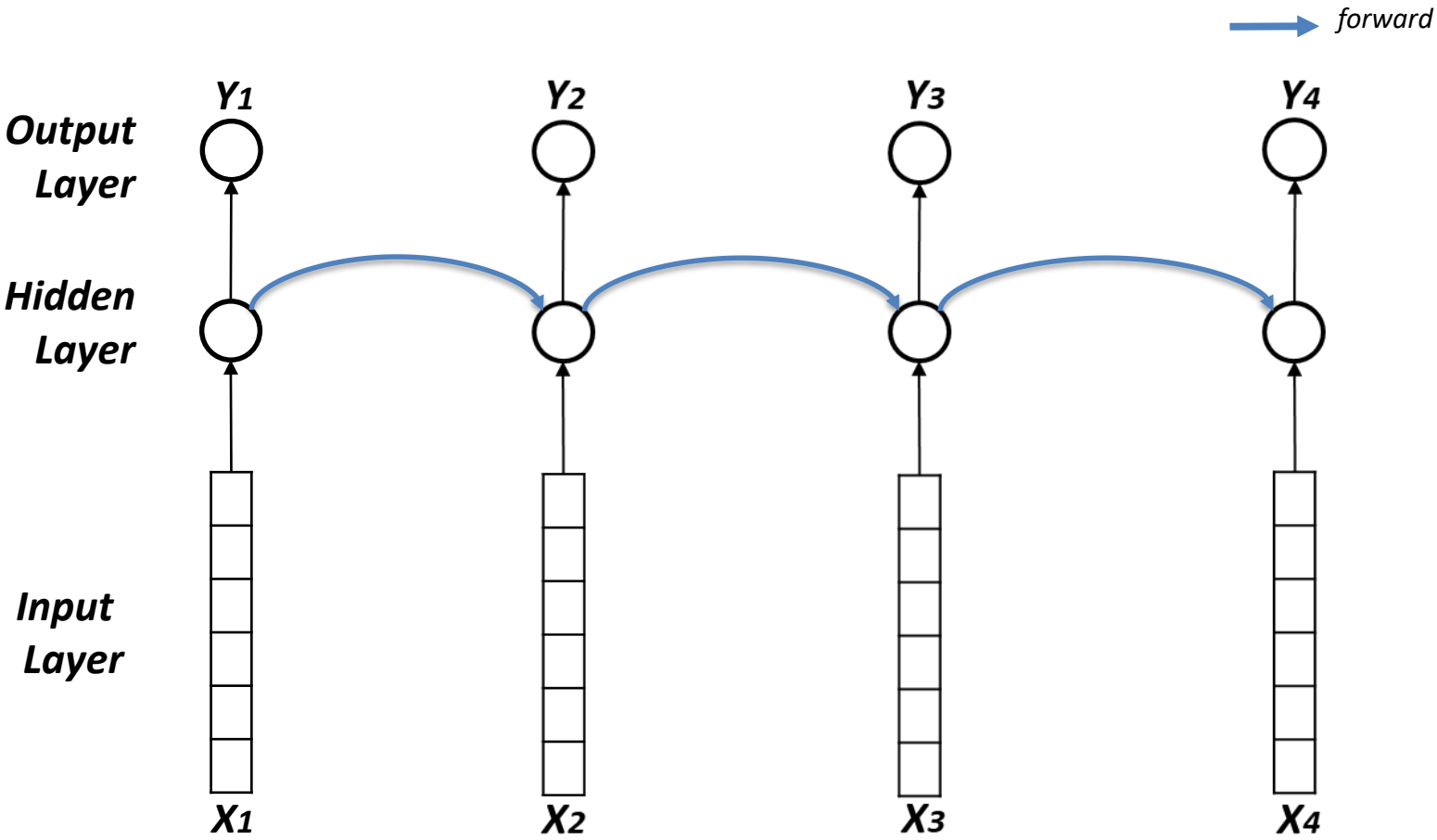
Seq2Seq with Deep Learning

Neural Network + Memory



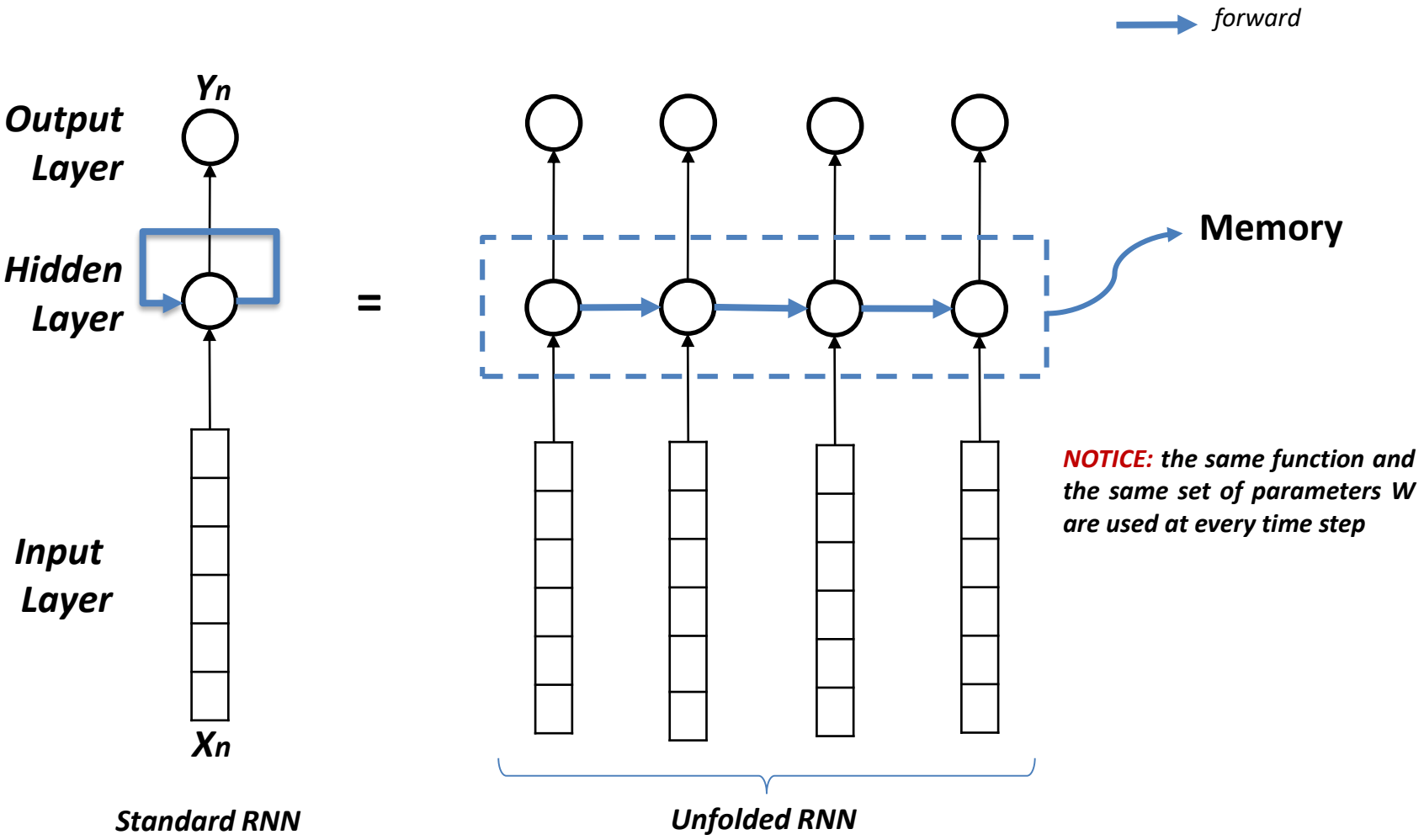
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



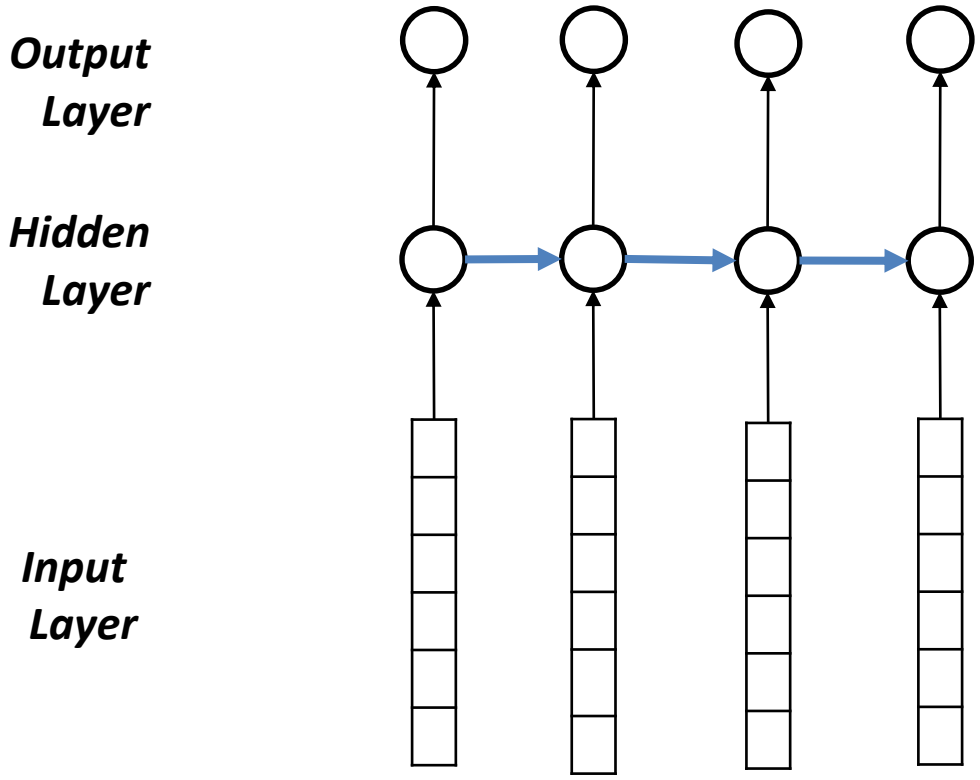
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network (RNN)



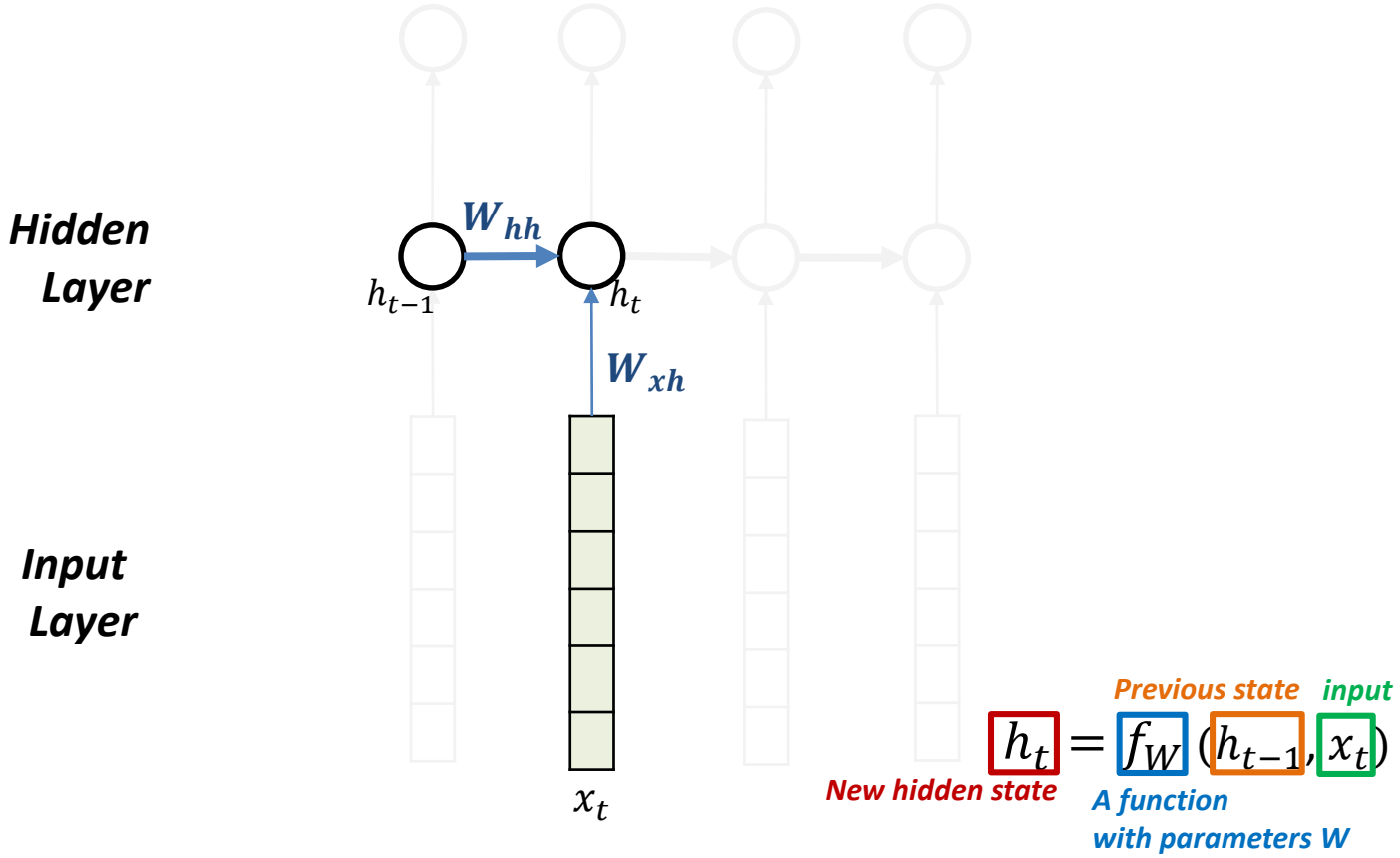
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network (RNN)



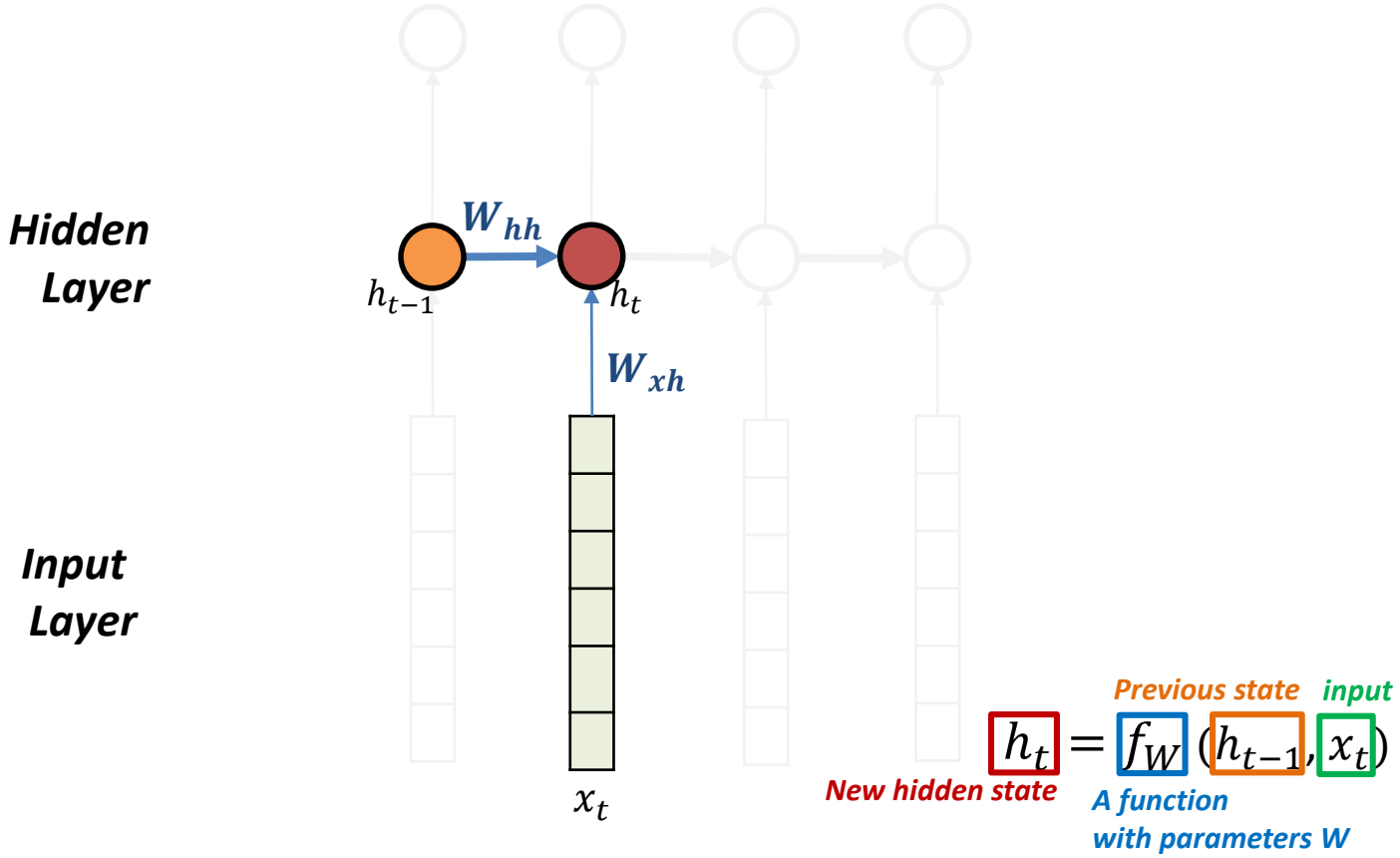
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



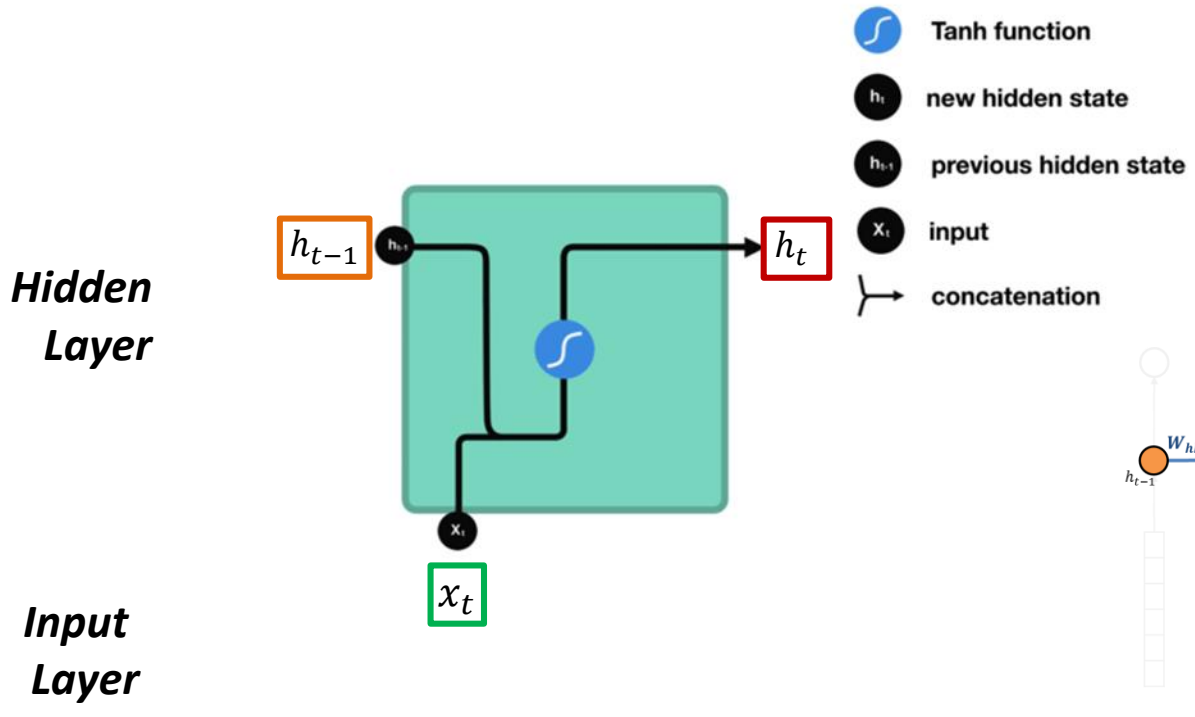
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

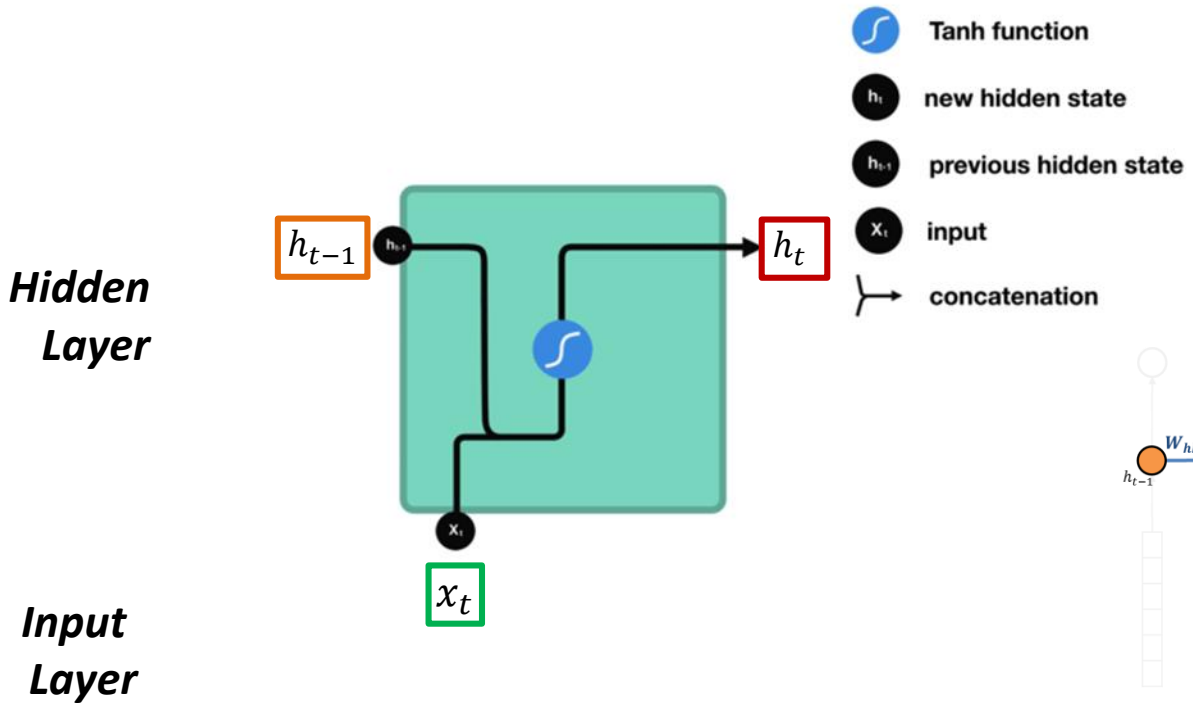


$$h_t = f_W(h_{t-1}, x_t)$$

New hidden state — *Previous state* — *input* — *A function with parameters W*

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



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

New hidden state

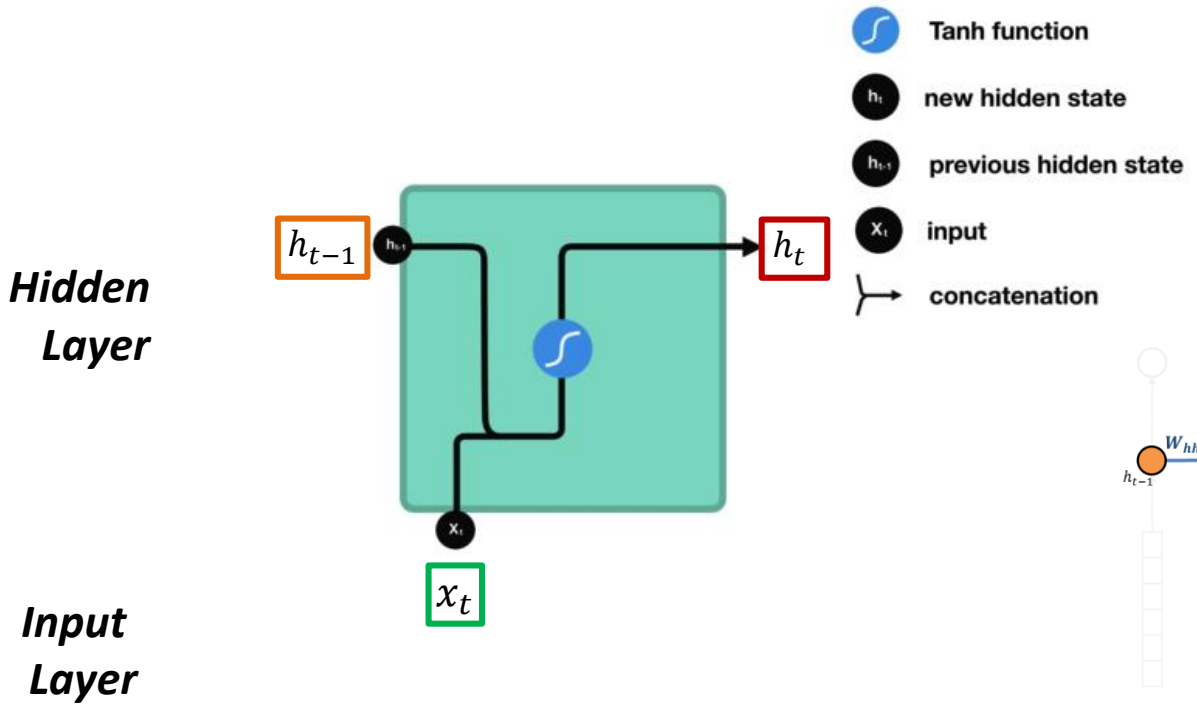
A function with parameters W

Previous state

input

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



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

New hidden state

A function with parameters W

Previous state

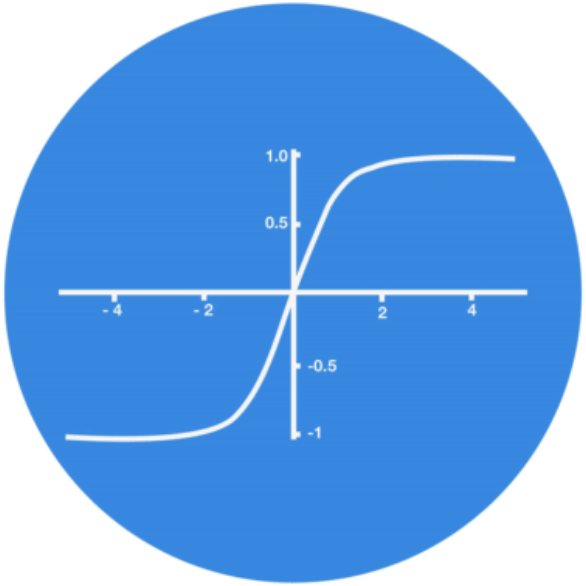
input

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.

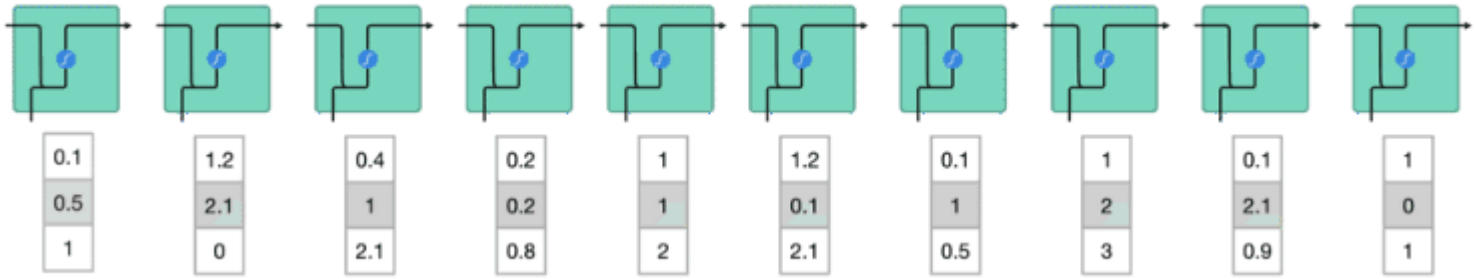
5
0.1
-0.5



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

With Sequence Input

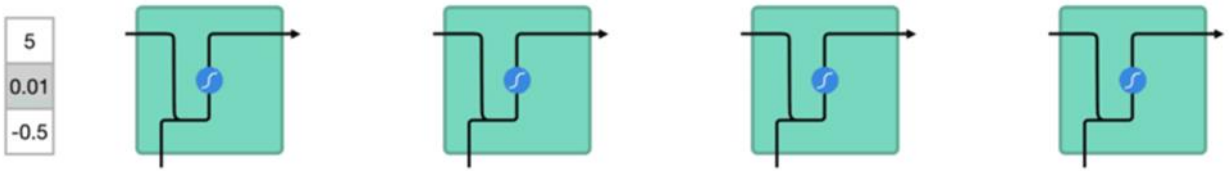


Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Q: Why do we need tanh function?

Vector Transformations without tanh

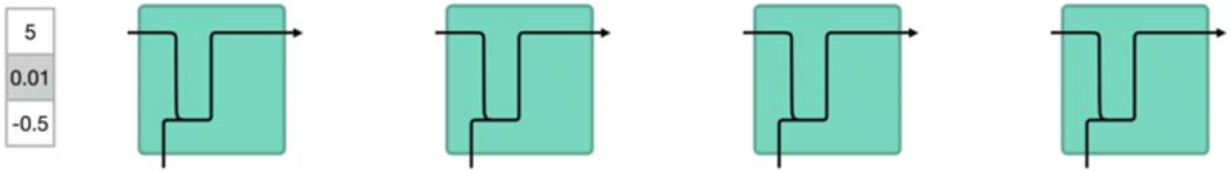


Vector Transformations with tanh

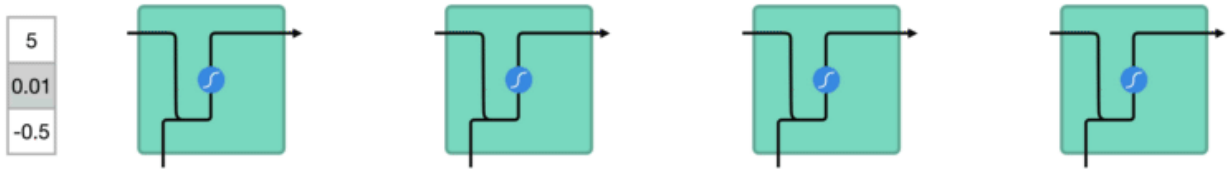
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Q: Why do we need tanh function?



Vector Transformations without tanh

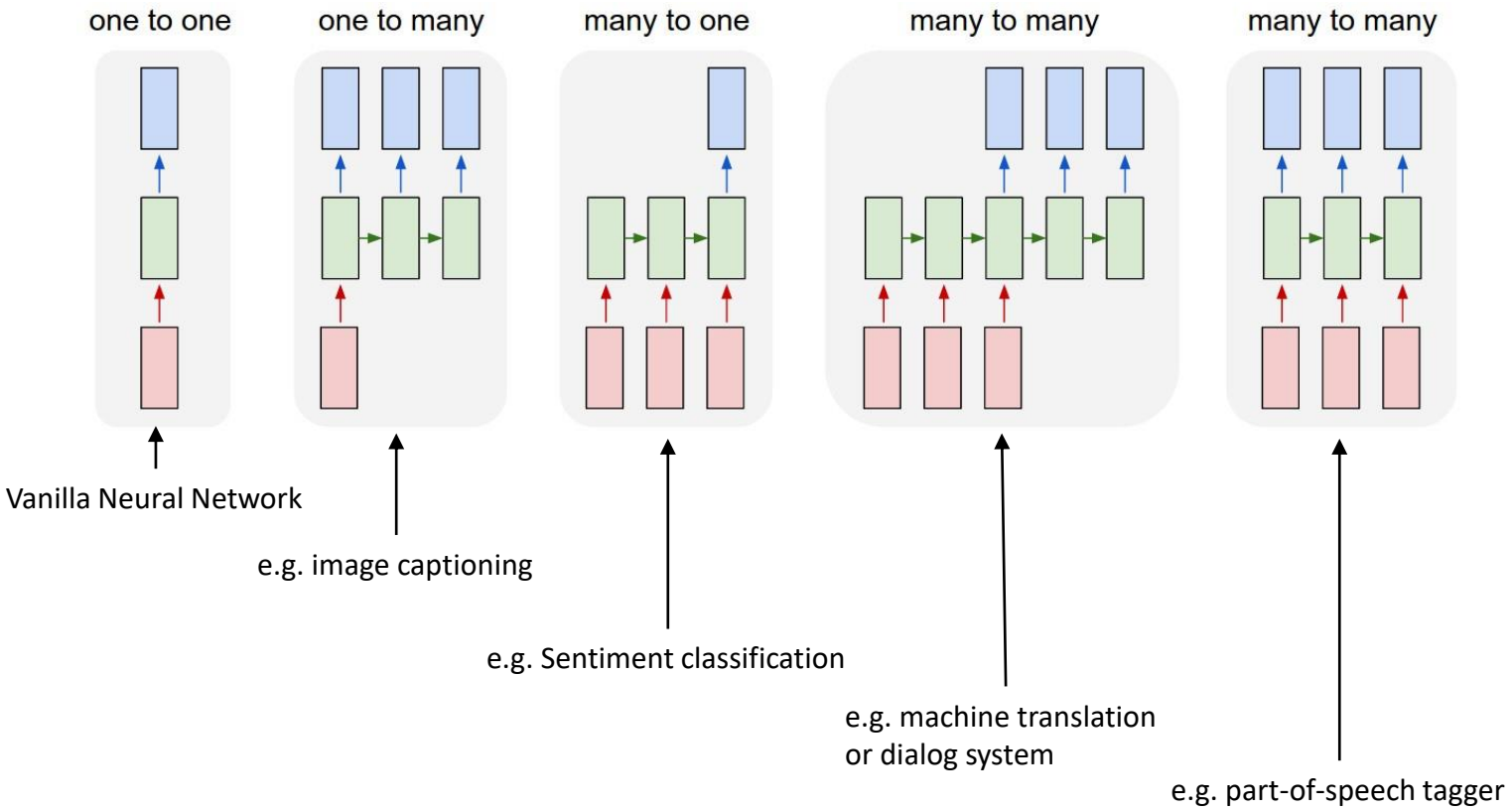


Vector Transformations with tanh

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

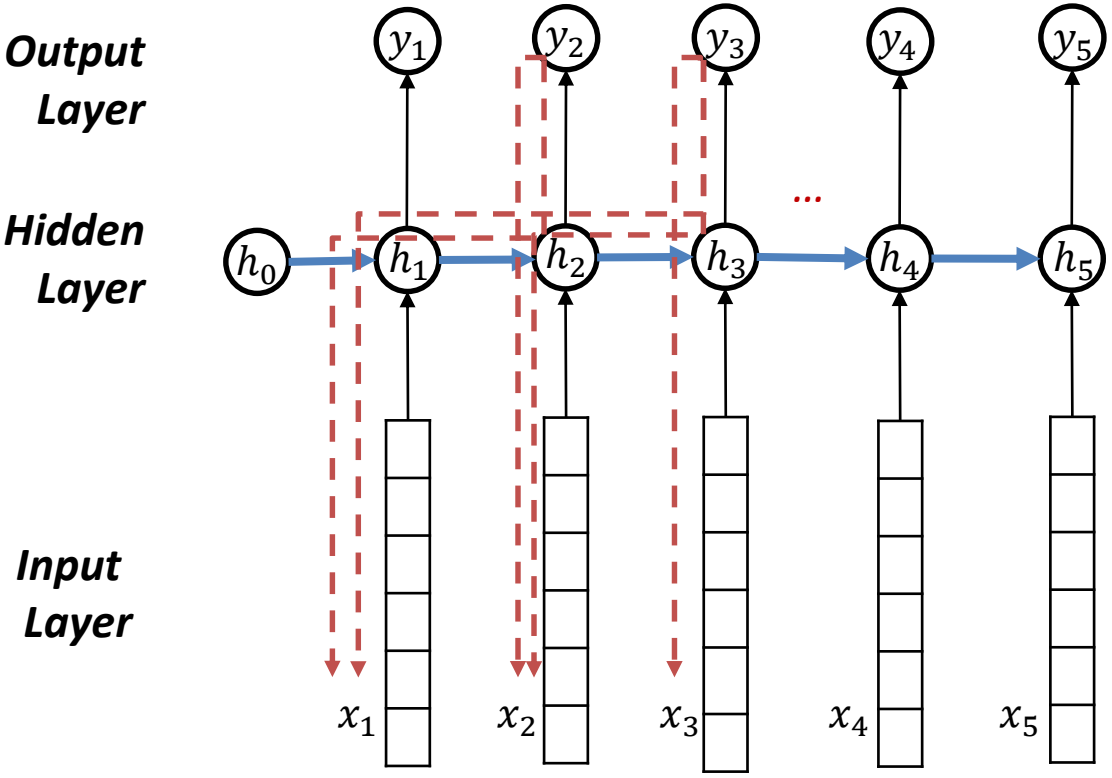
Several Variants of RNN



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Backpropagation through time



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

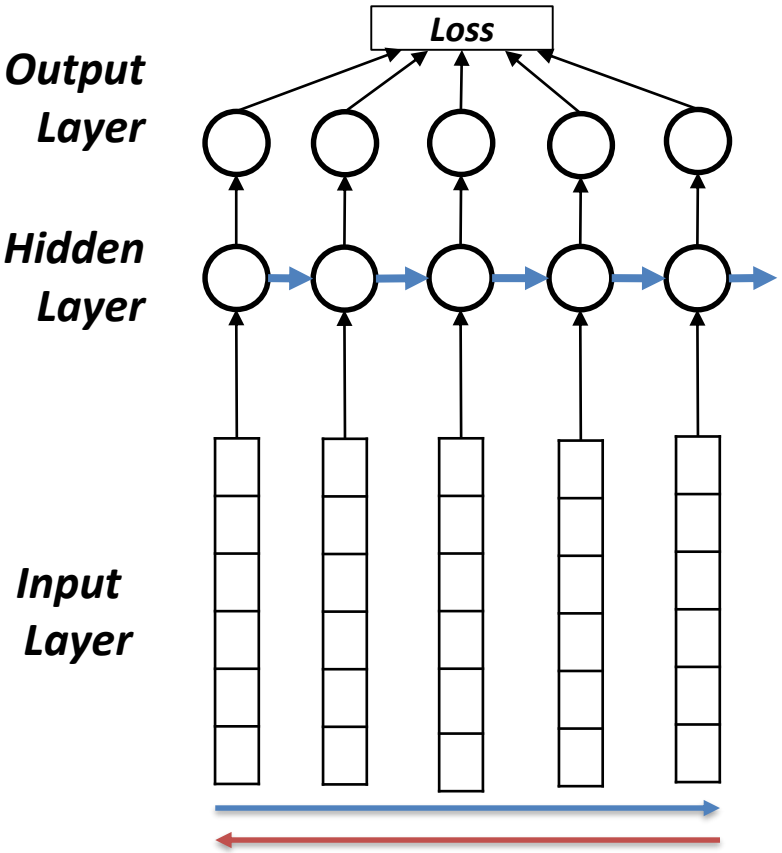
*Use **Truncate Backpropagation** through time*

- *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

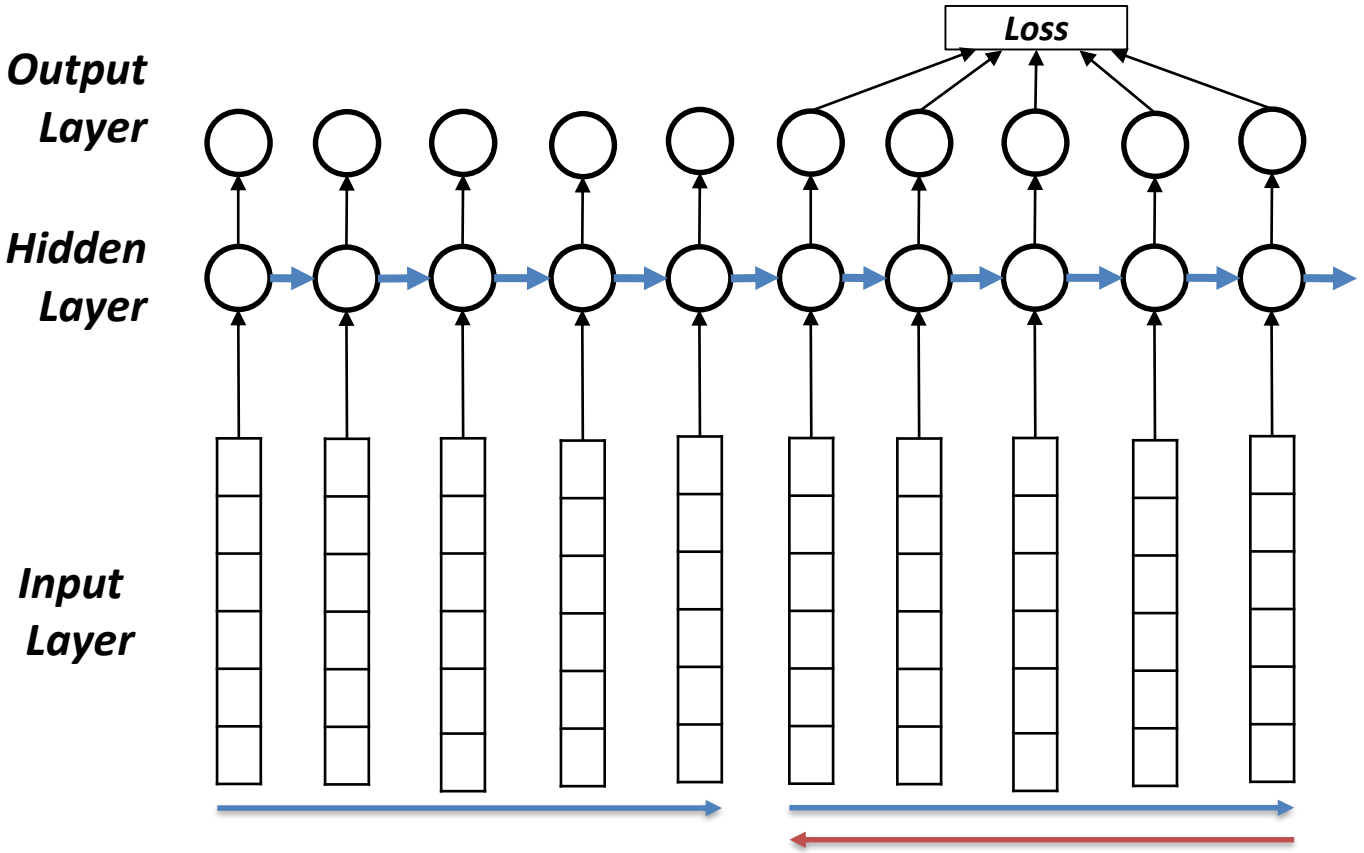


Run forward and backward through chunks of the sequence instead of whole sequence

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

Truncated Backpropagation through time

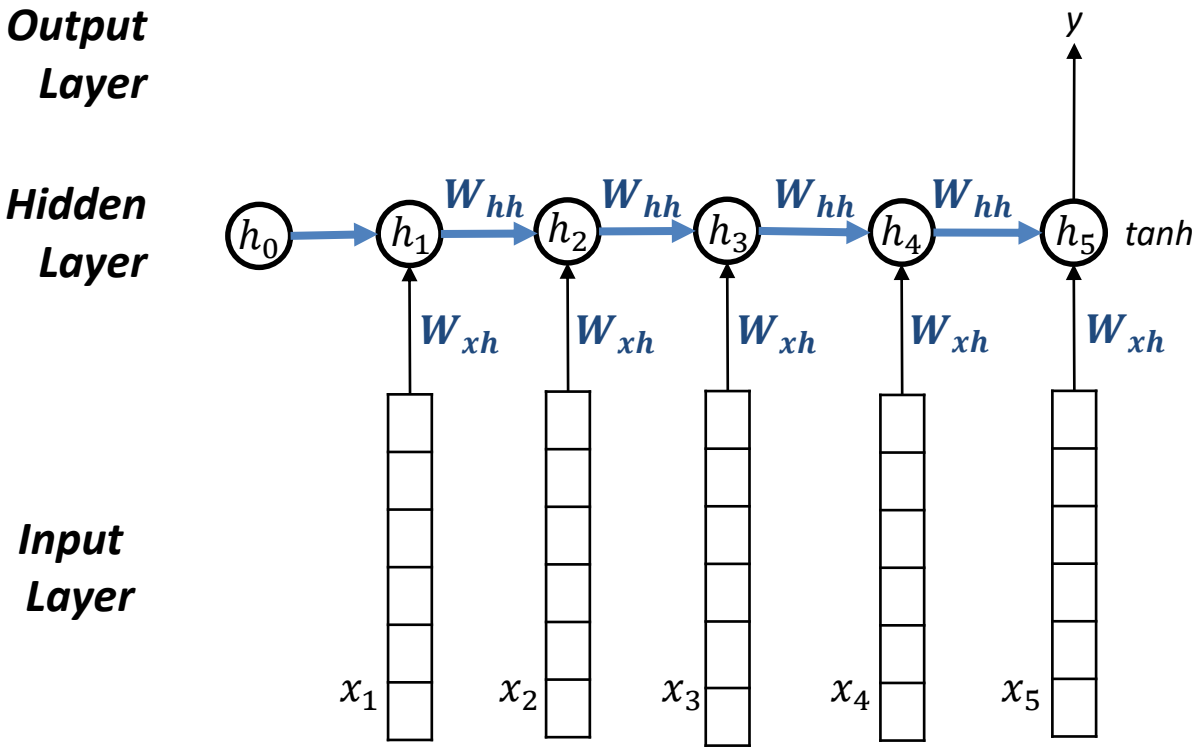


Carry hidden states forward in time forever, but only backpropagate for some smaller number of steps

Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

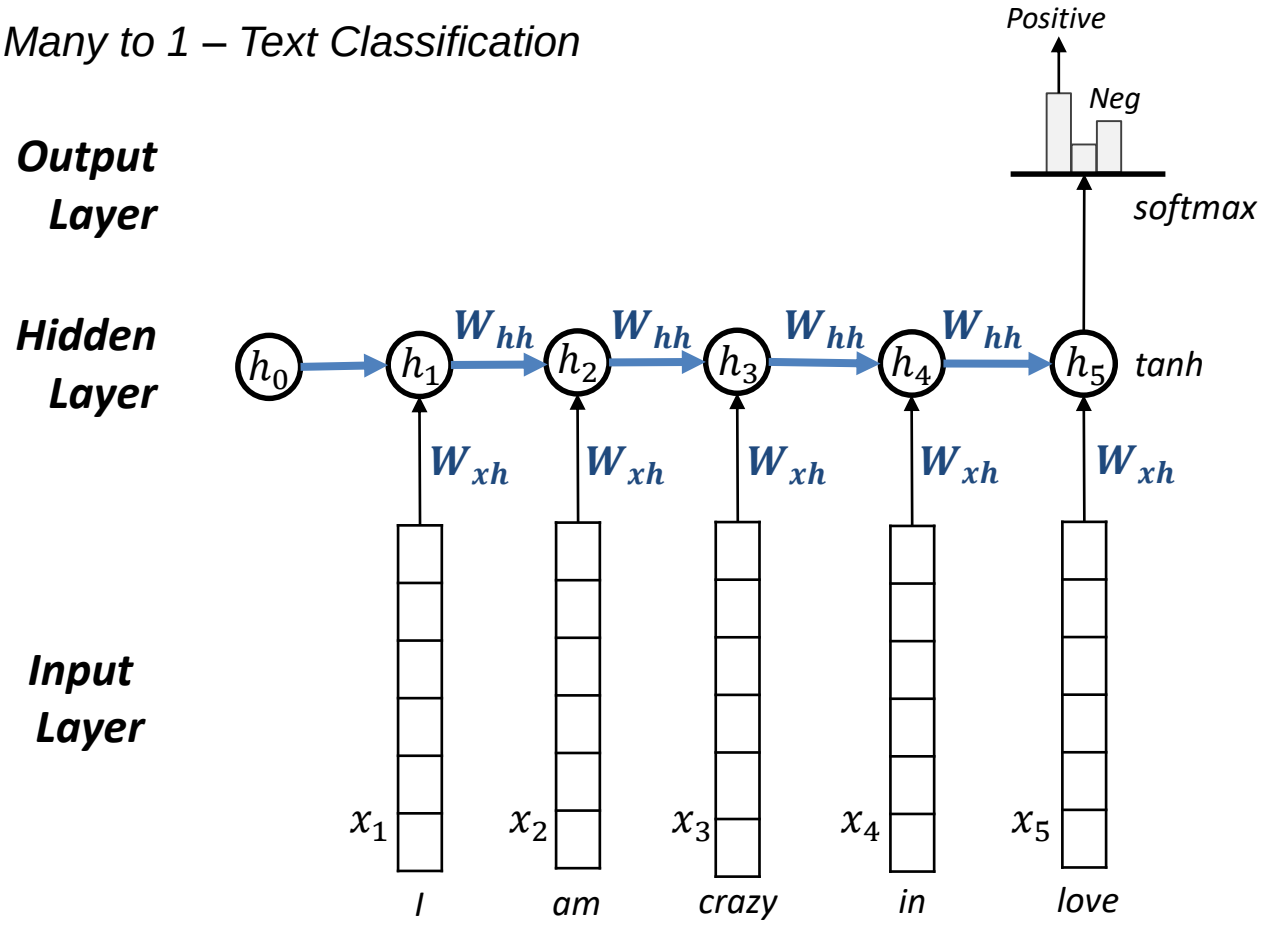
Many to 1



Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network

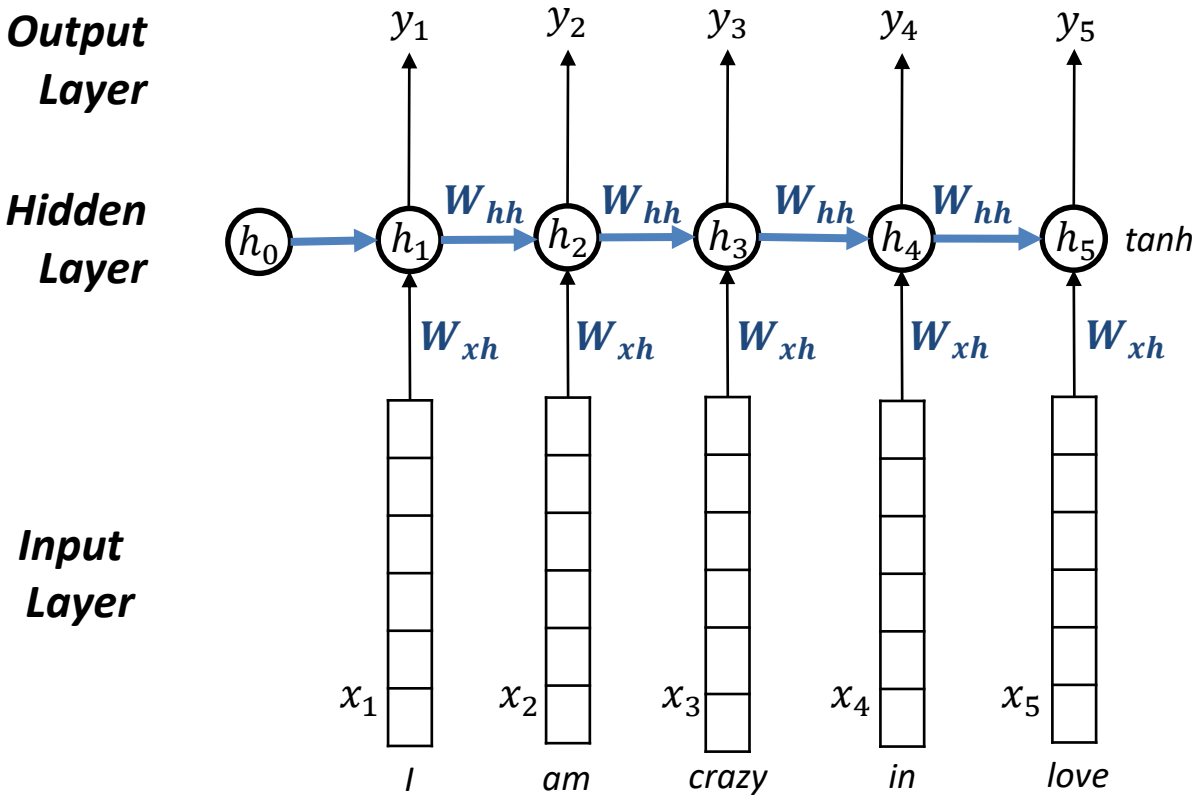
Many to 1 – Text Classification



Seq2Seq with Deep Learning

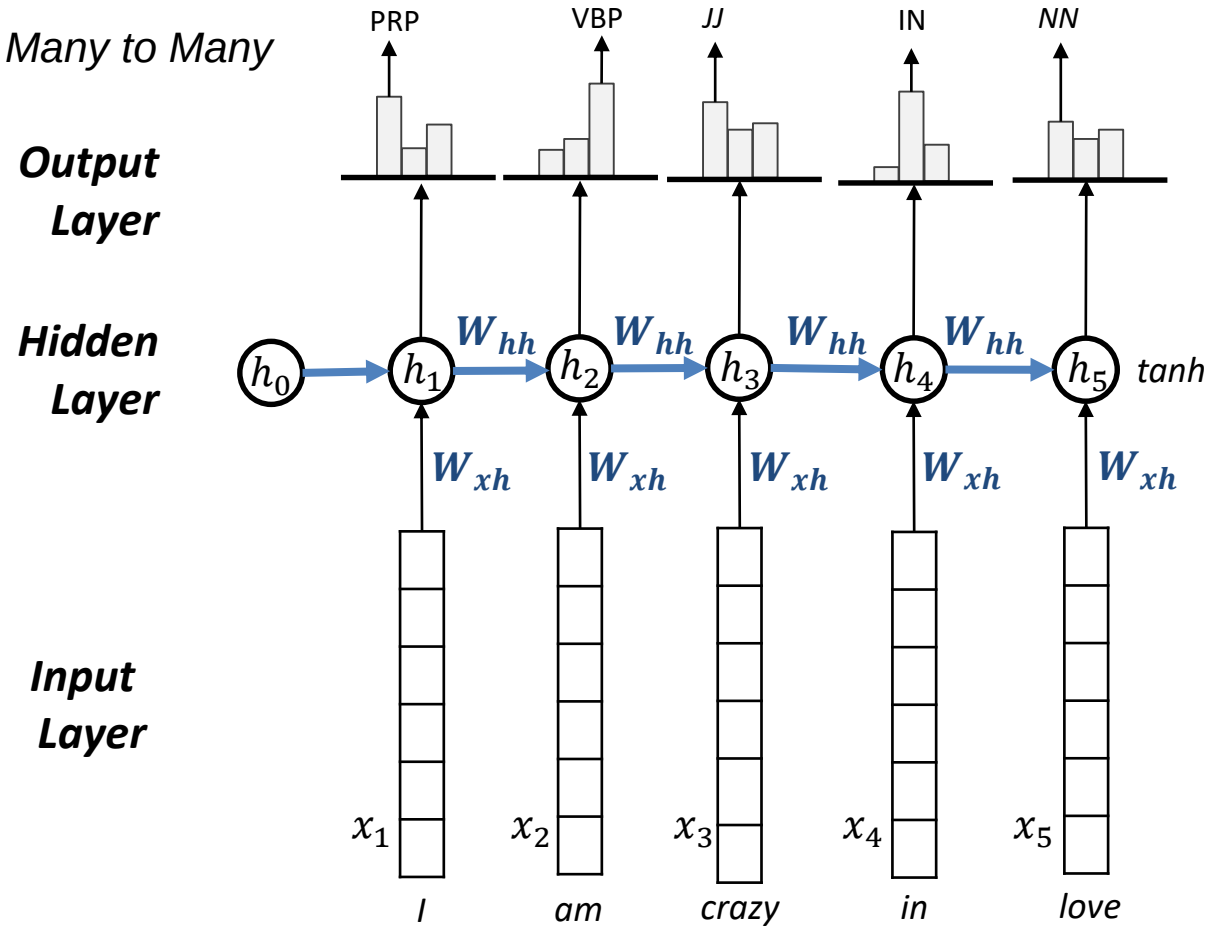
Neural Network + Memory = Recurrent Neural Network

Many to Many



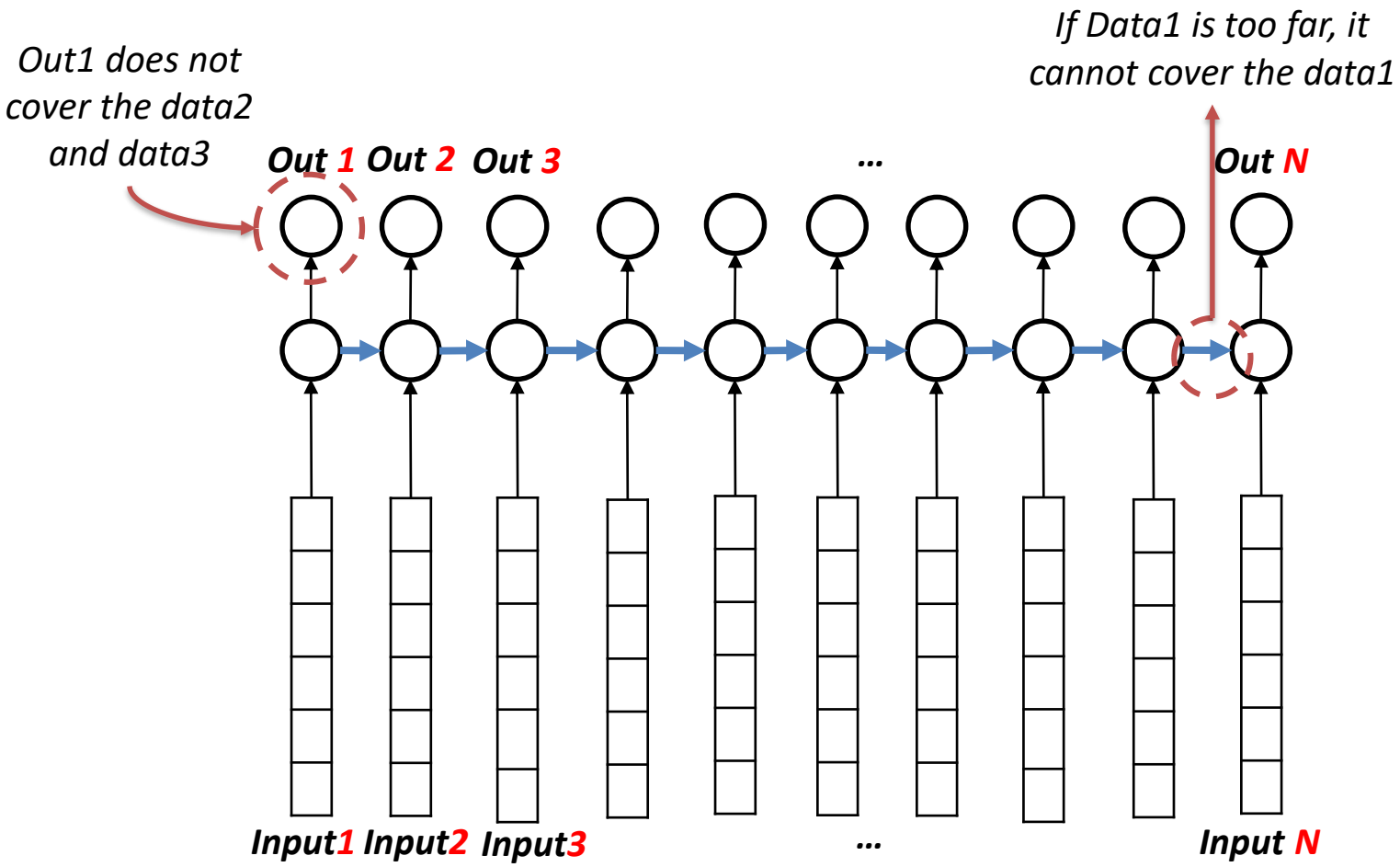
Seq2Seq with Deep Learning

Neural Network + Memory = Recurrent Neural Network



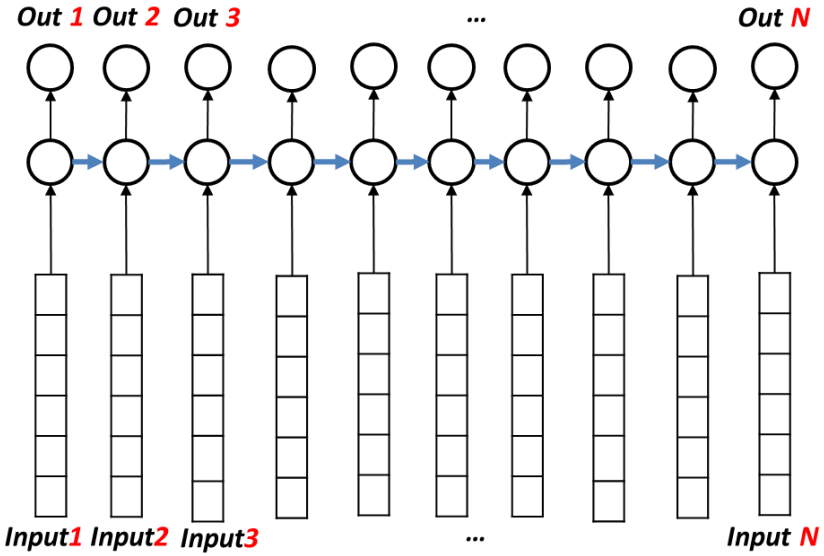
Seq2Seq with Deep Learning

Limitation of Vanilla RNN



Seq2Seq with Deep Learning

Limitation of Vanilla RNN

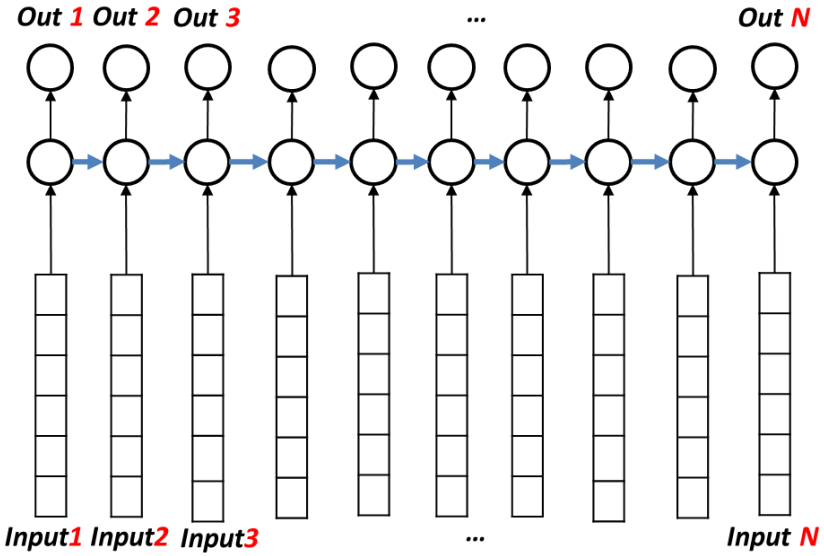


Limitation1: Vanishing Gradient Issue

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.

Seq2Seq with Deep Learning

Limitation of Vanilla RNN

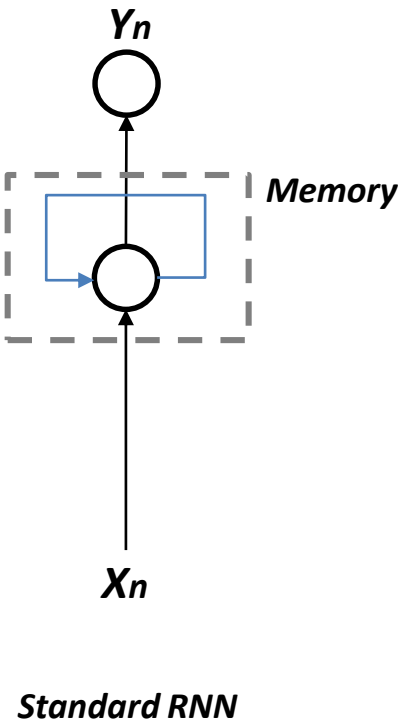


Limitation2: Exploding Gradient

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 in turn, 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.

Seq2Seq with Deep Learning

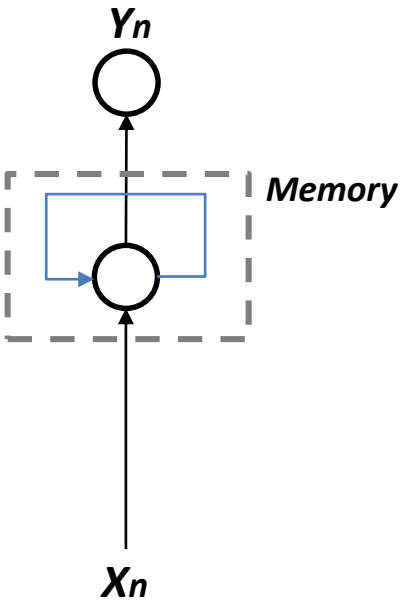
LSTM (Long Short-Term Memory) - Idea



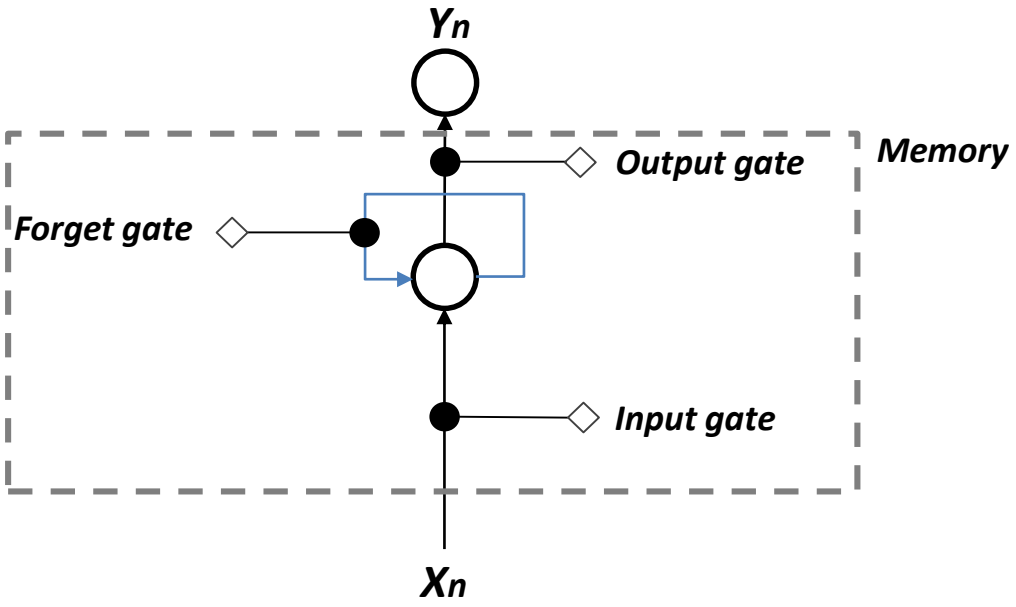
Standard RNN

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) - Idea



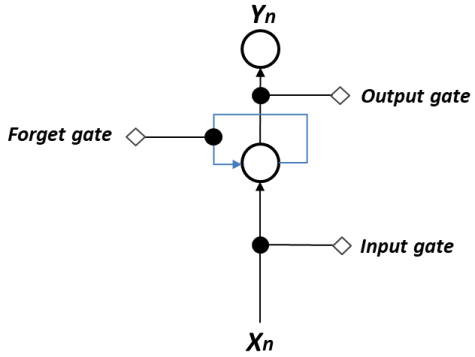
Standard RNN

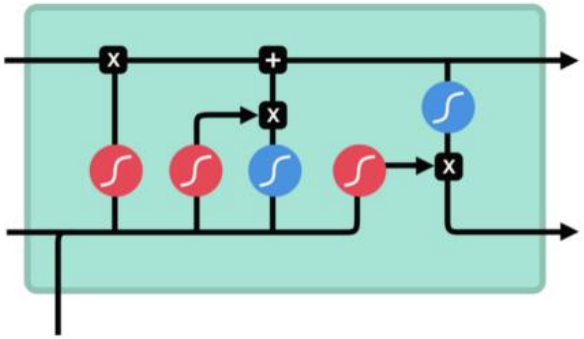


Long Short-Term Memory

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 SOTA in many sequence learning problems

Seq2Seq with Deep Learning

Sigmoid activation

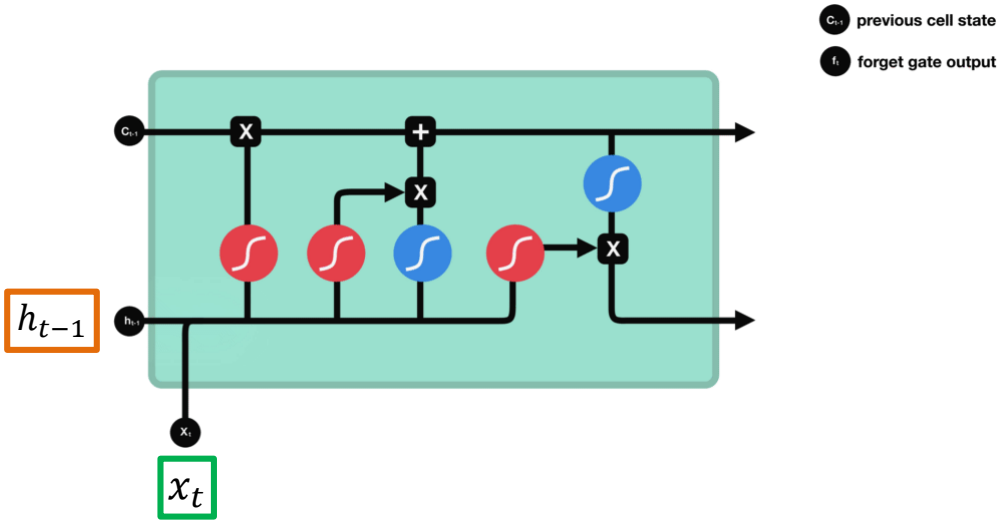
A sigmoid activation is similar to the tanh activation. Instead of squishing values between -1 and 1, it squishes values between 0 and 1.

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

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

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Forget Gate



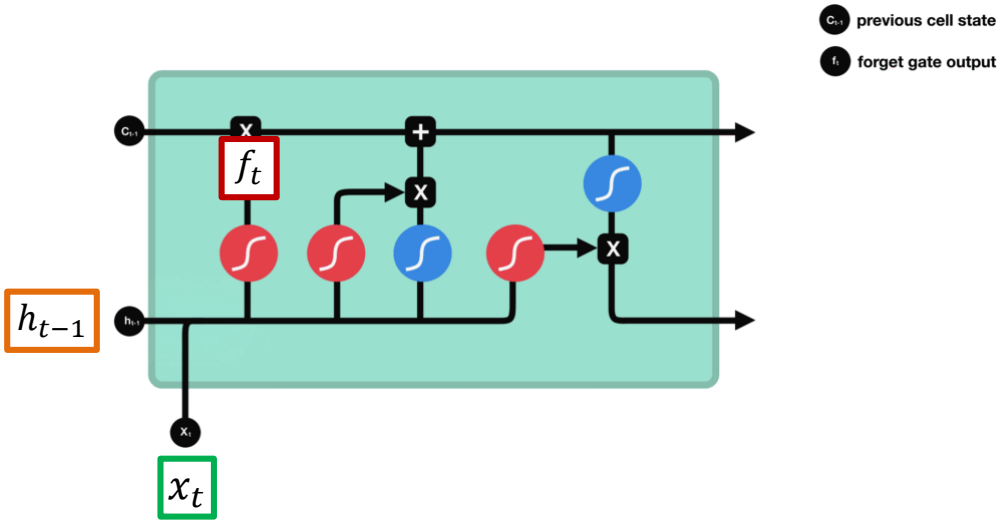
$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f)$$

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.

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Forget Gate



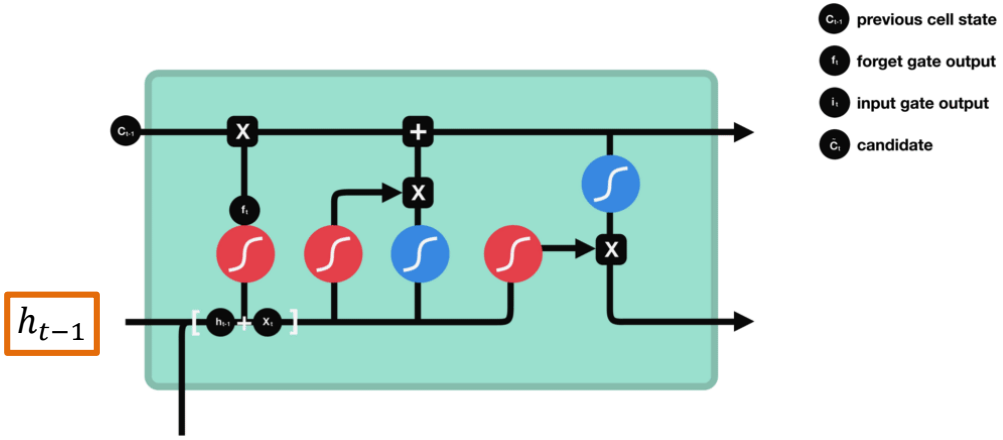
$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f)$$

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.

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Input Gate



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

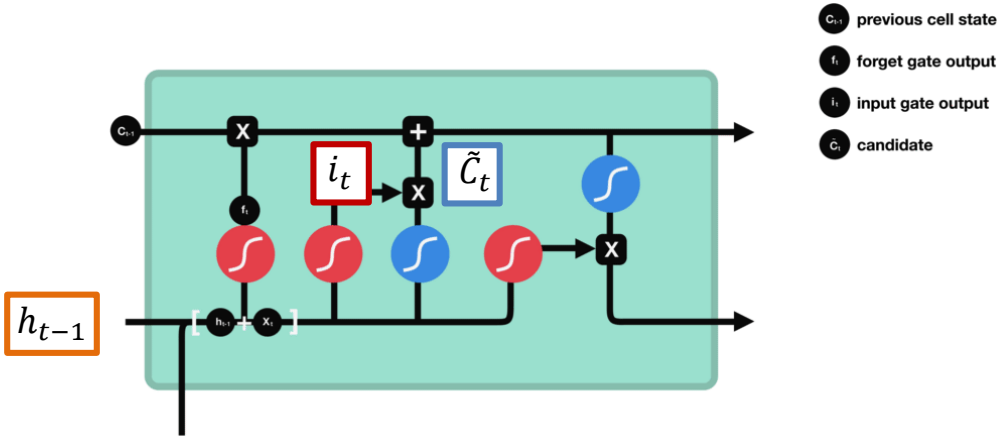
$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$$

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) – Input Gate



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

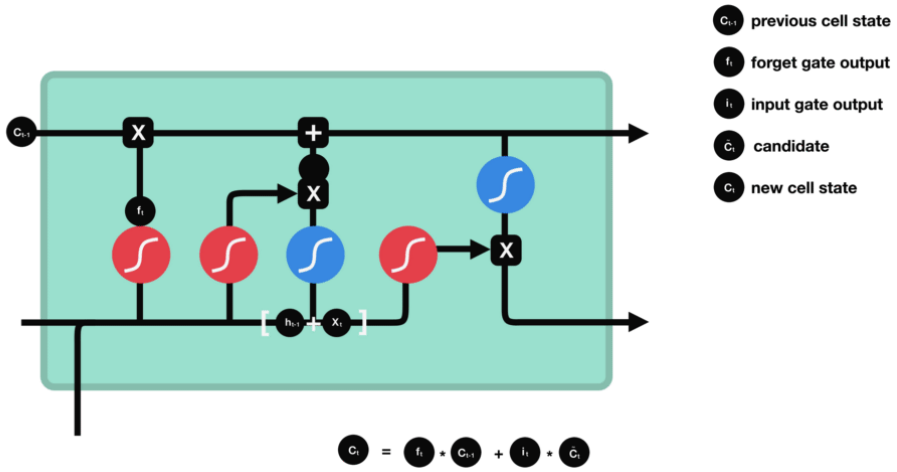
$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$$

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

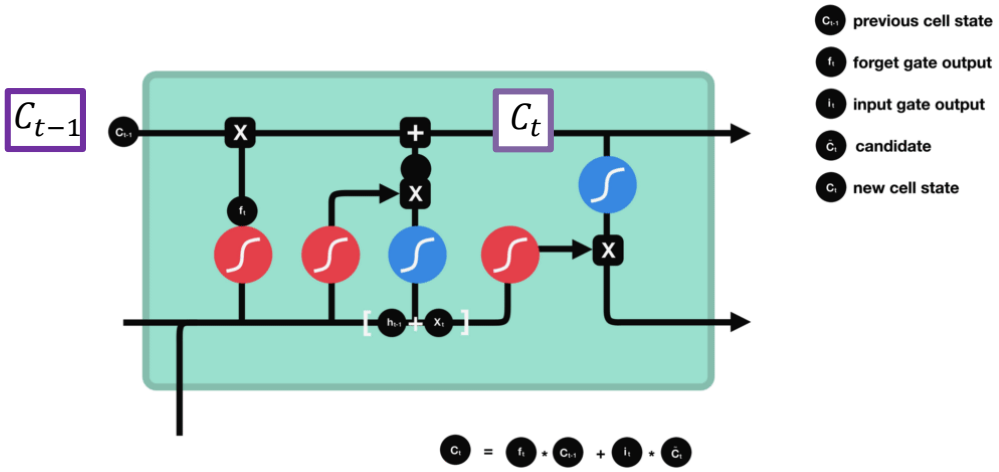


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- *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*

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Cell States

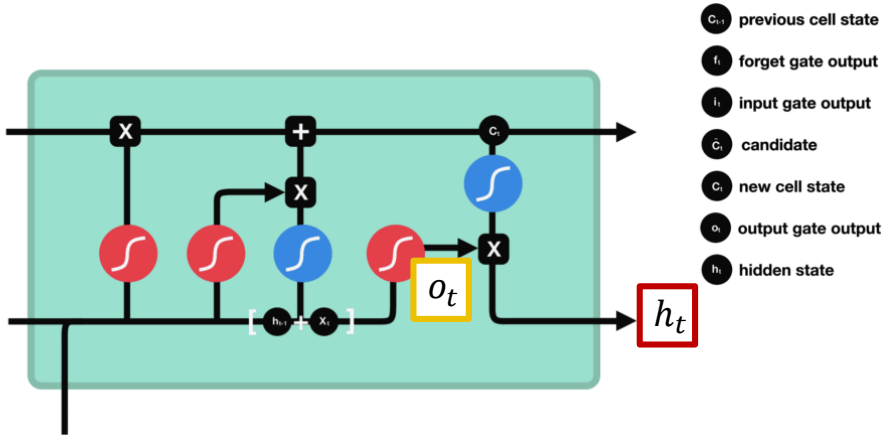


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- *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*

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) – Output Gate



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

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

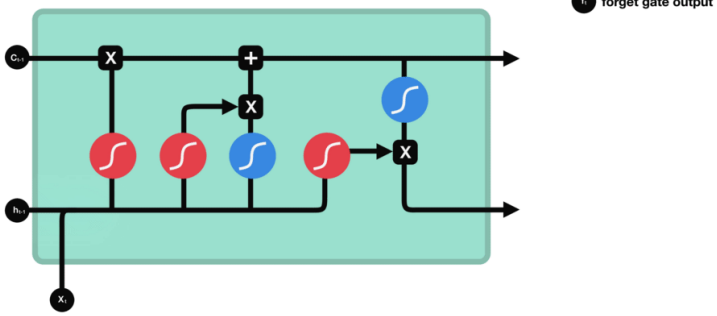
decides what the next hidden state should be.

- 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

Seq2Seq with Deep Learning

LSTM (Long Short-Term Memory) - Overall

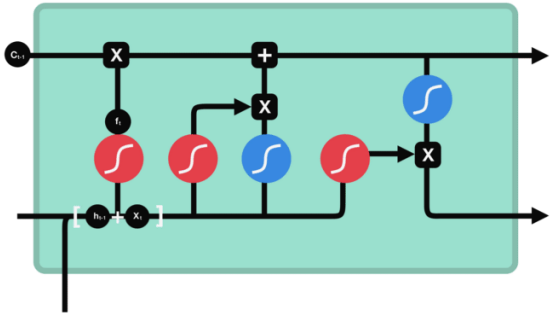
Forget gate



previous cell state

forget gate output

Input gate



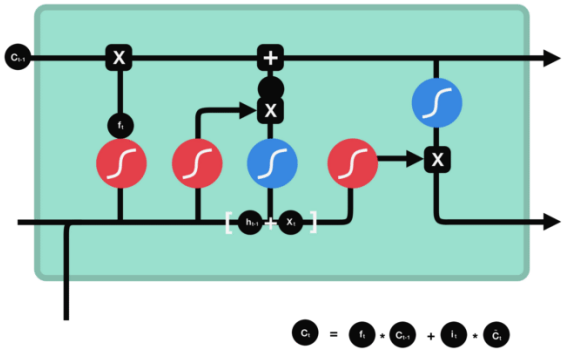
previous cell state

forget gate output

input gate output

candidate

Cell state



previous cell state

forget gate output

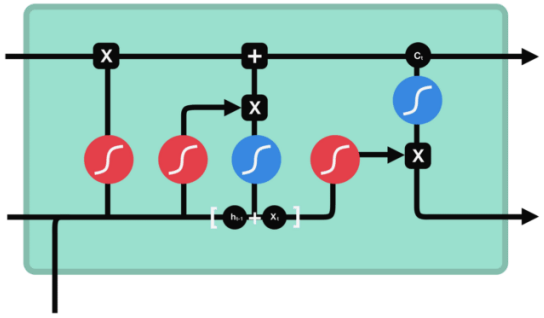
input gate output

candidate

new cell state

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

Output gate



previous cell state

forget gate output

input gate output

candidate

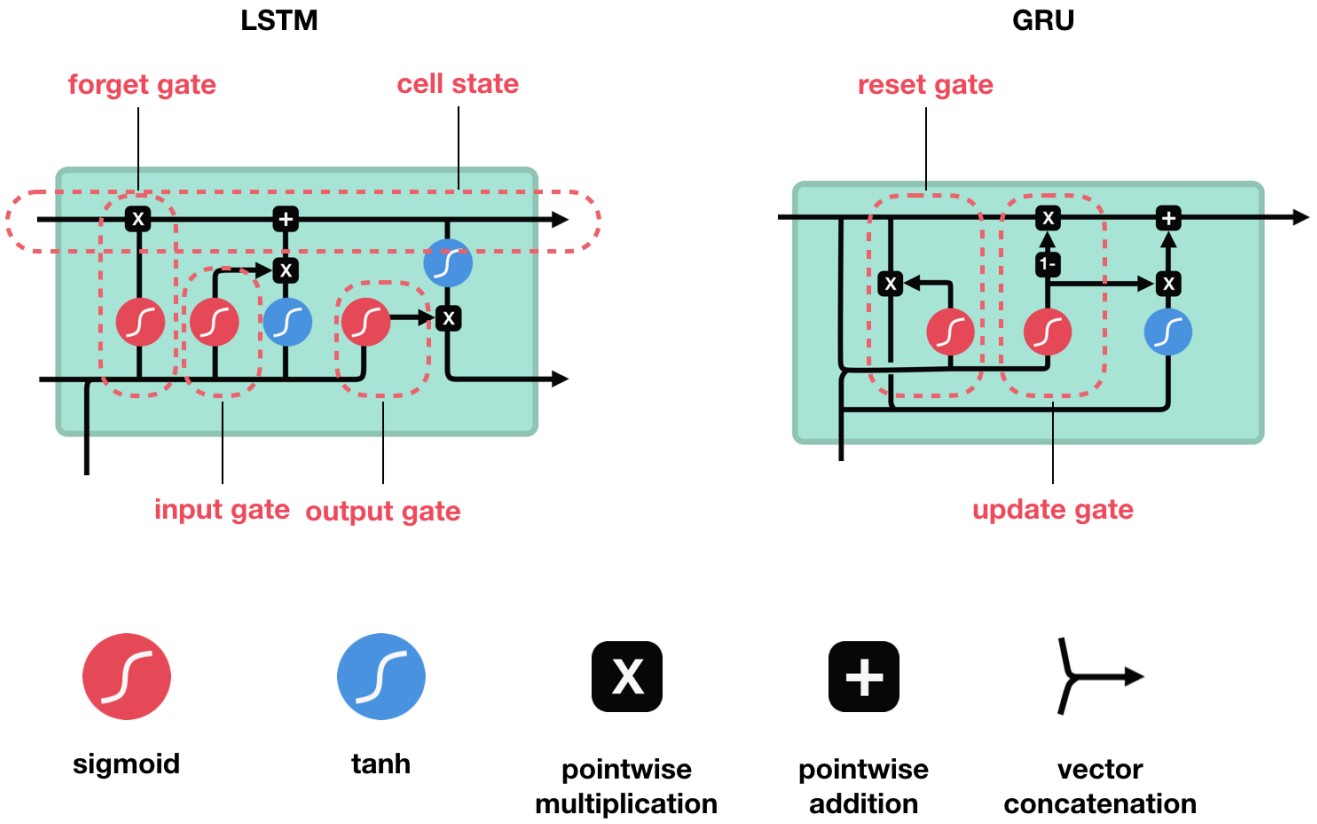
new cell state

output gate output

hidden state

Seq2Seq with Deep Learning

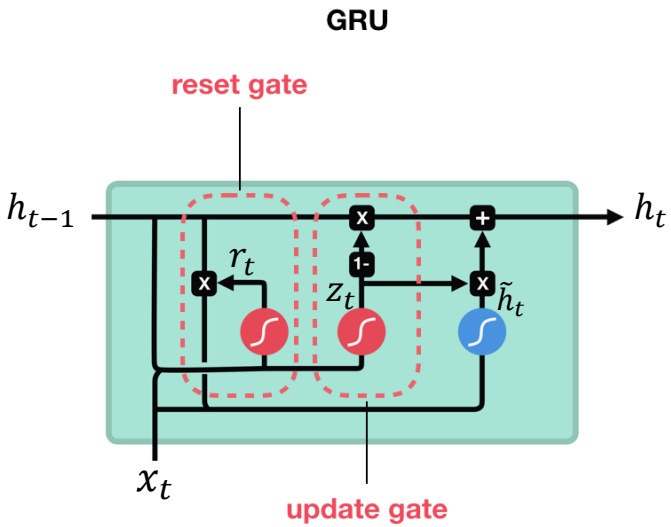
Gated Recurrent Unit



Seq2Seq with Deep Learning

Gated Recurrent Unit

- *GRU first computes an **update gate** based on **current input word vector** and **hidden state***
  - Acts similar to the forget and input gate of an LSTM. It decides what information to throw away and what new information to add
- *reset gate is another gate is used to decide how much past information to forget*
  - 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*



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

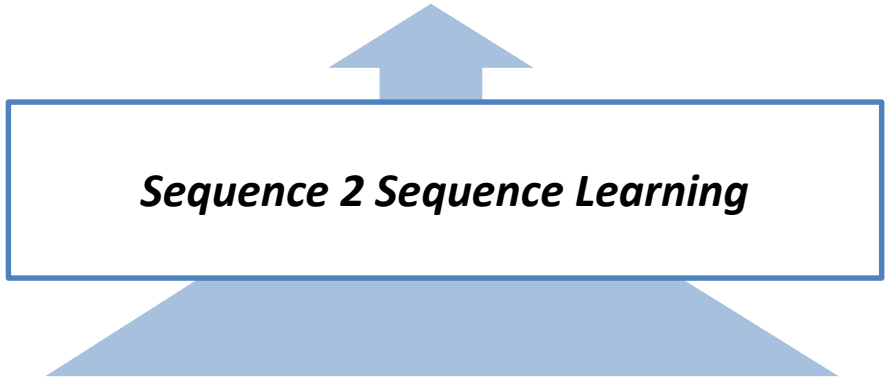
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Seq2Seq Modelling

Seq2Seq – PoS tagger

ADV VERB DET NOUN NOUN

Output: Part of Speech

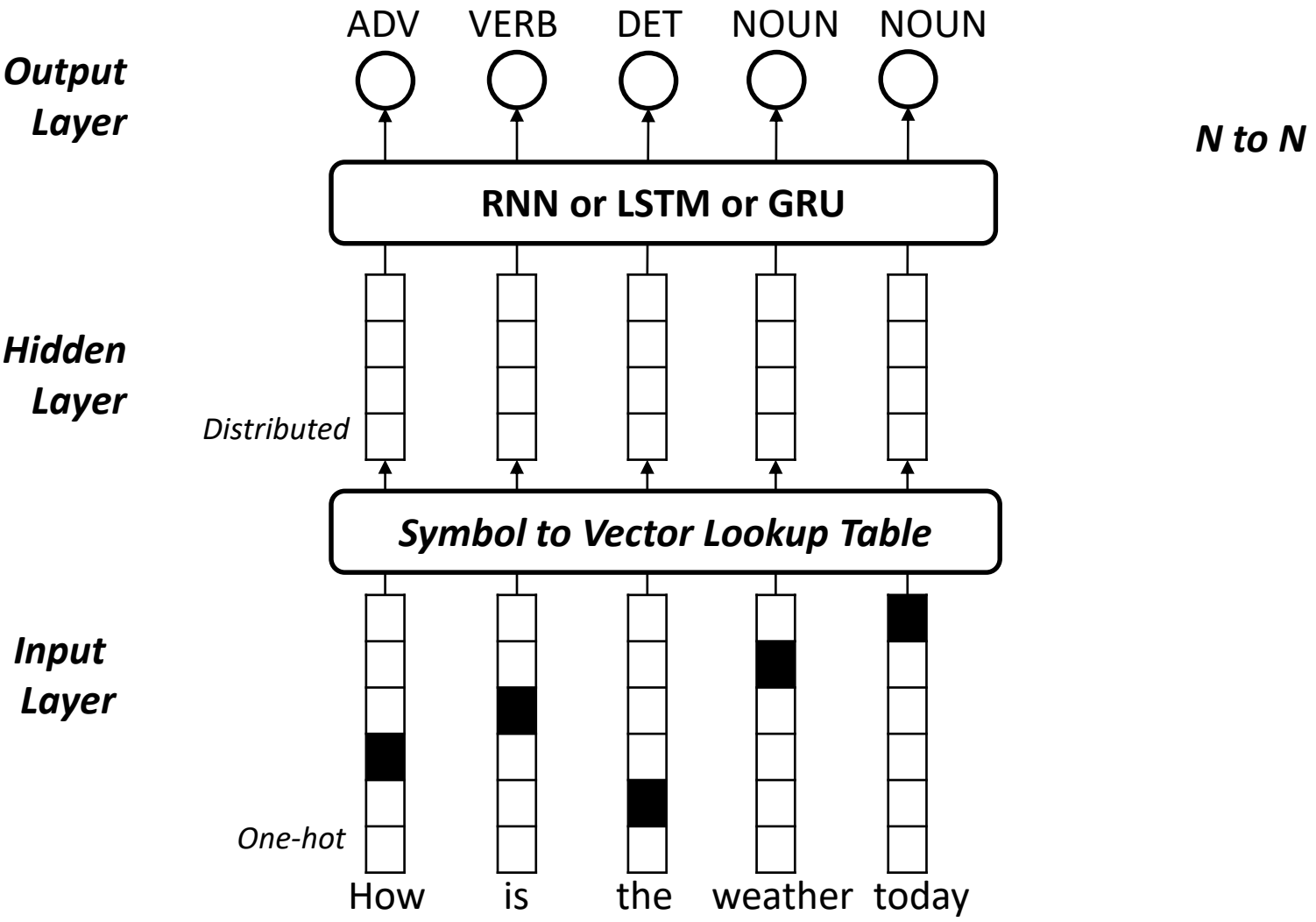


How is the weather today

Input: Text

Seq2Seq Modelling

Sequence Modelling for POS Tagging



0 LECTURE PLAN

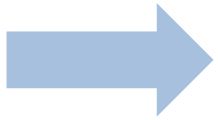
THE UNIVERSITY OF SYDNEY

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

Data Transformation for Deep Learning NLP

ImageNet: Image Classification



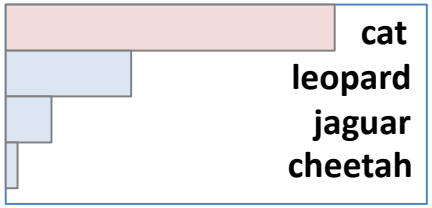
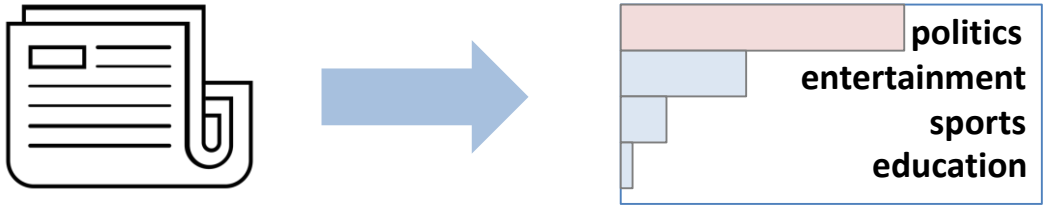


Image Pixel

Data Transformation for Deep Learning NLP

Topic Classification



News Articles

Data Transformation for Deep Learning NLP

Visual Question Answering

Data Transformation for Deep Learning NLP

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

Where is he sitting

Submit

Predicted top-5 answers with confidence:

couch	71.669%
chair	21.119%
sofa	2.730%
living room	1.376%
room	1.276%

Data Transformation for Deep Learning NLP

Visual Question Answering

Why is he surprised

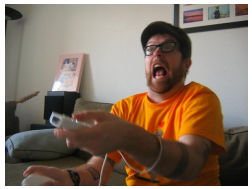
Submit

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%

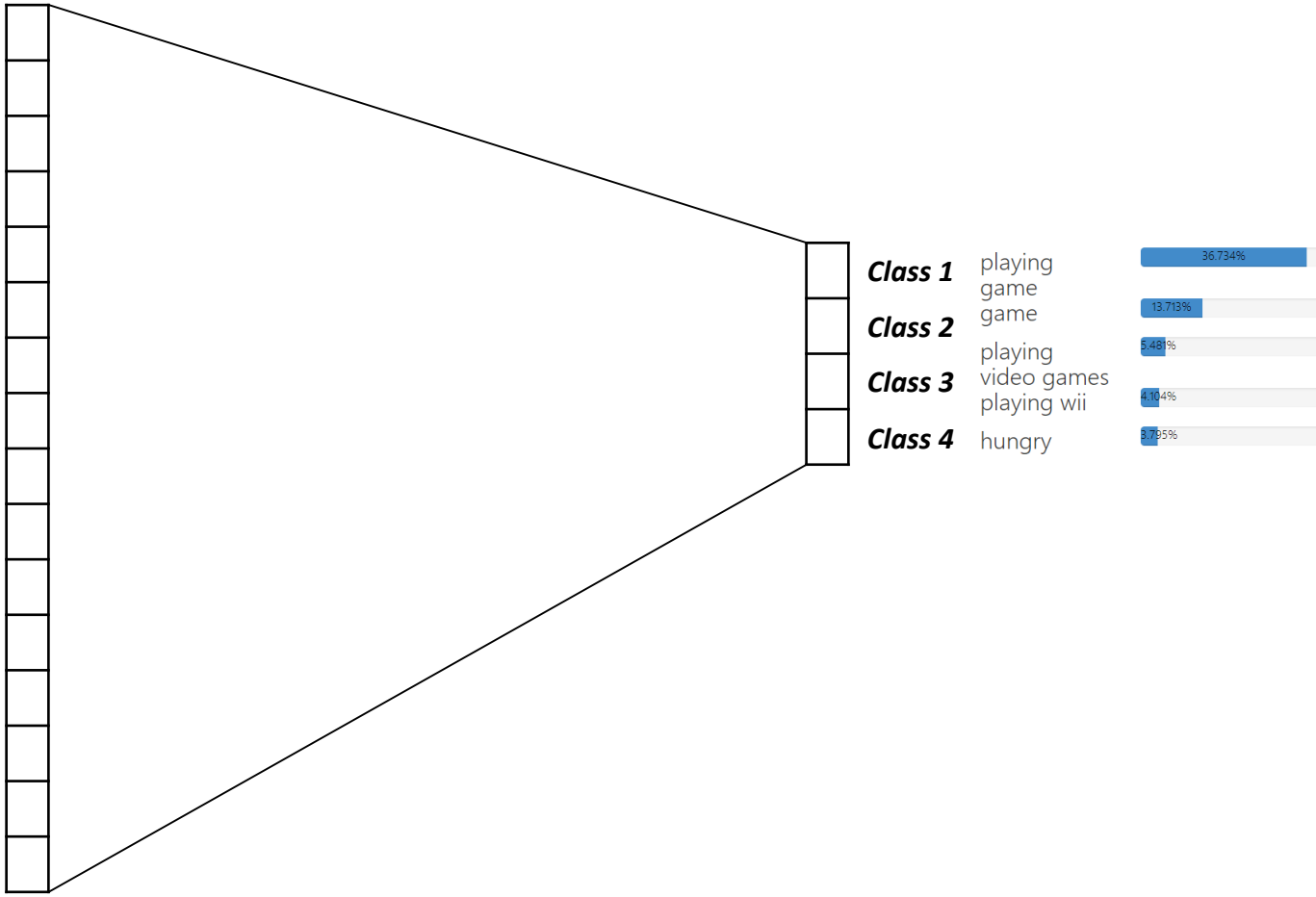
Data Transformation for Deep Learning NLP

Classification Formulation



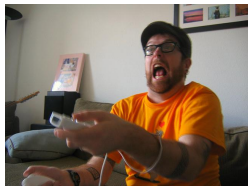
Input

Why is he surprised



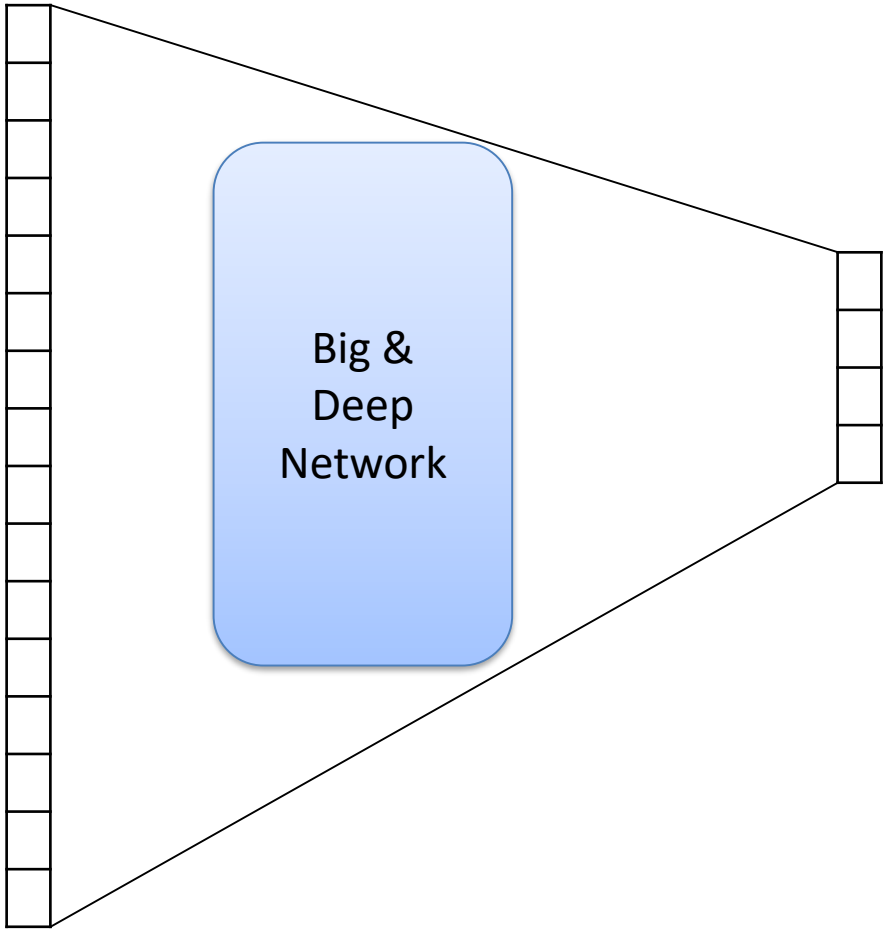
Data Transformation for Deep Learning NLP

Classification Formulation



Input

Why is he surprised



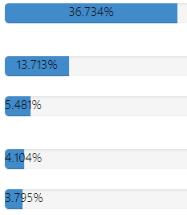
Class 1 playing game

Class 2 game

Class 3 playing video games

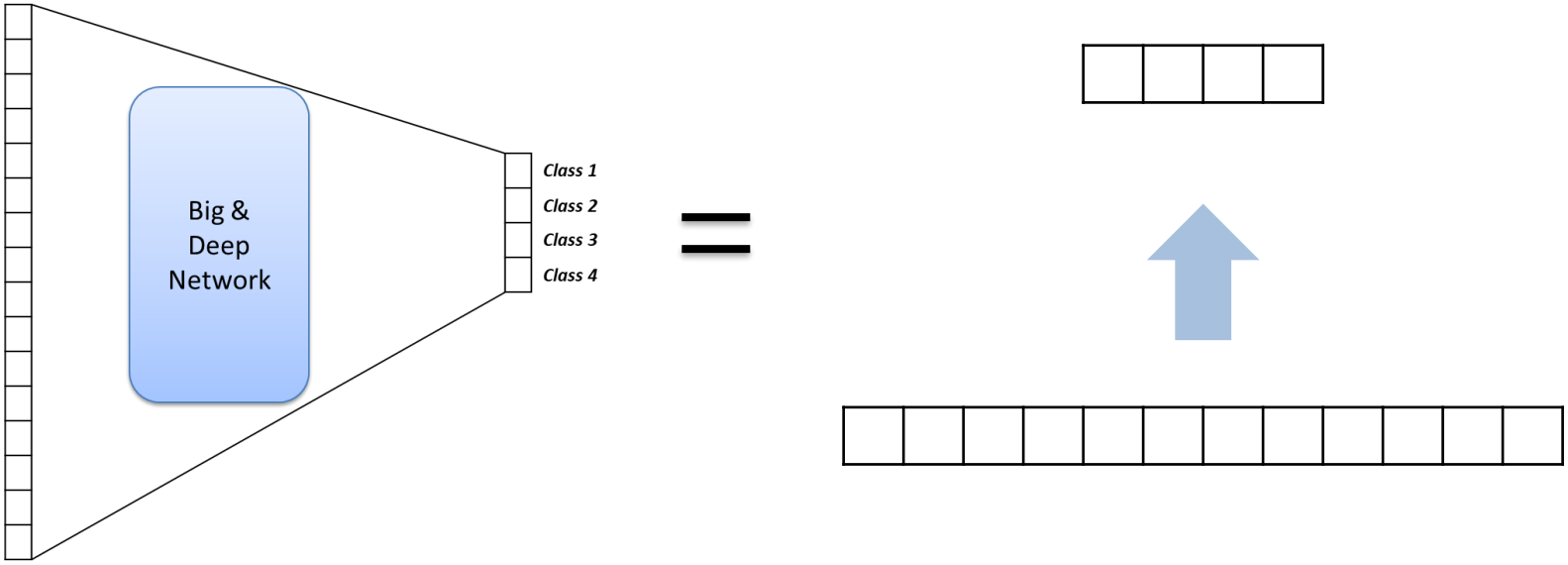
Class 4 playing wii

Class 4 hungry



Data Transformation for Deep Learning NLP

Classification



Data Transformation for Deep Learning NLP

Graphical Notation for Data

Data X
10
2
8
2
15
3
5
1
5

=

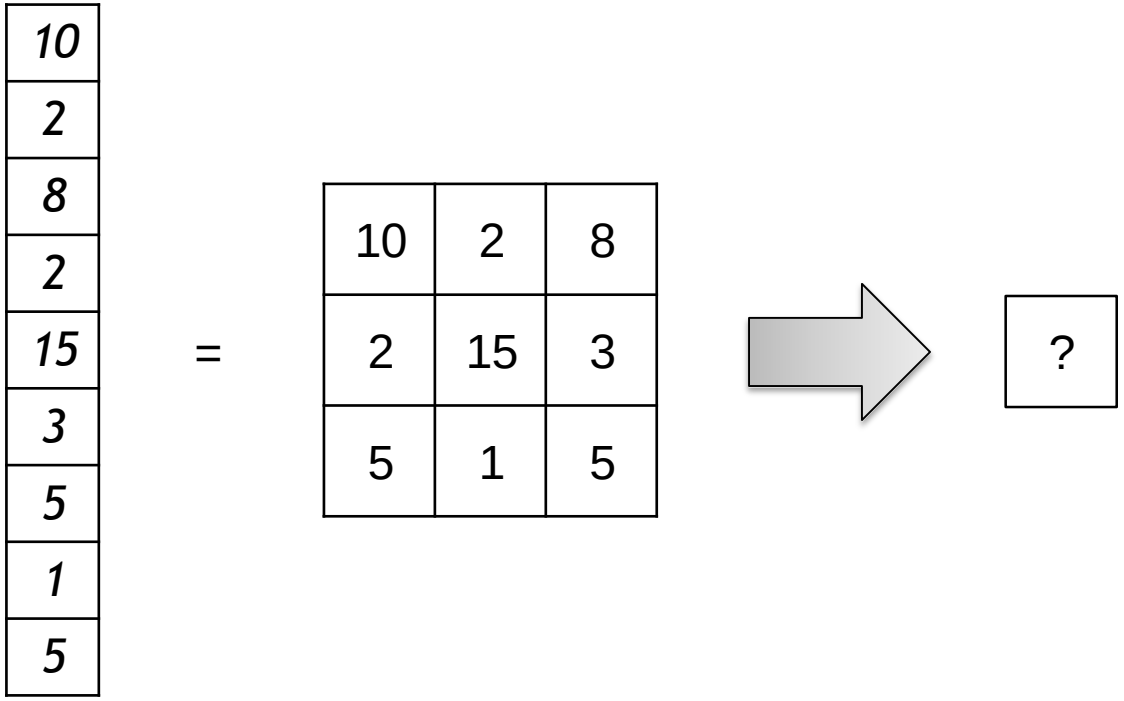
Data X		
10	2	8
2	15	3
5	1	5

Assume this is 9-dimension data, not 3x3

**** I draw in 3x3 in order to visualize it better in the lecture note***

Data Transformation for Deep Learning NLP

V to 1

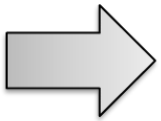


Data Transformation for Deep Learning NLP

V to 1 – Simple Method

center one

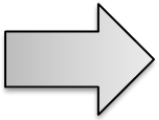
10	2	8
2	15	3
5	1	5



15

average

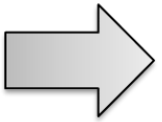
10	2	8
2	15	3
5	1	5



5.6

median

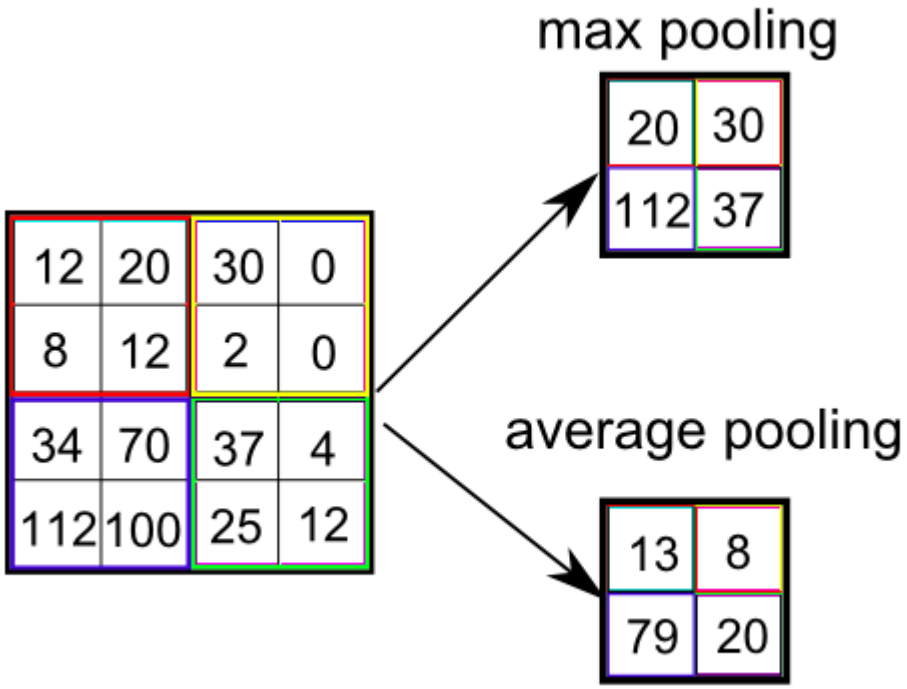
10	2	8
2	15	3
5	1	5



5

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

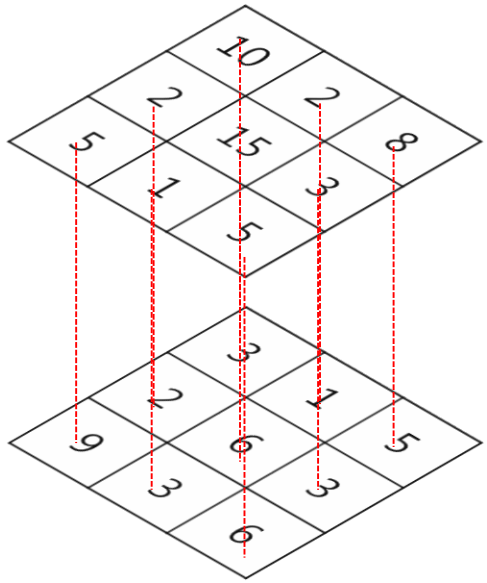
10	2	8
2	15	3
5	1	5

Value

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

Weight

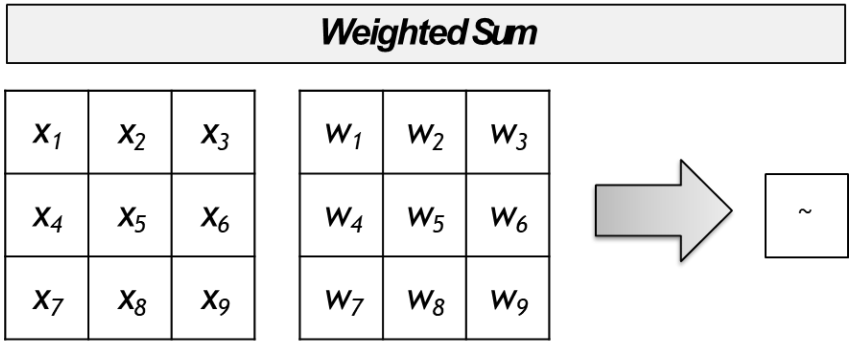
→ 6.65



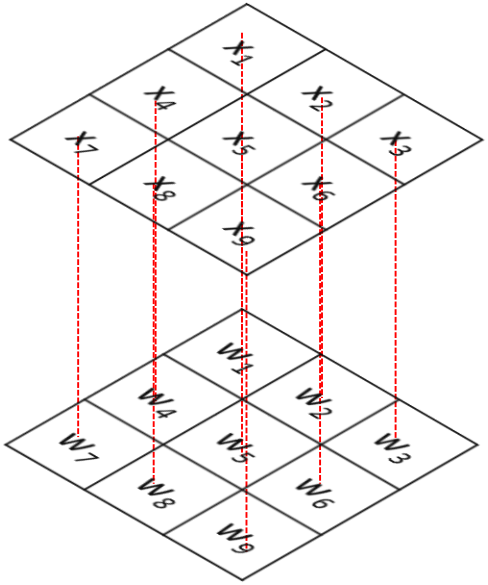
Element-wise multiplication

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

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

X

[9x1] matrix

w_1
w_2
w_3
w_4
w_5
w_6
w_7
w_8
w_9

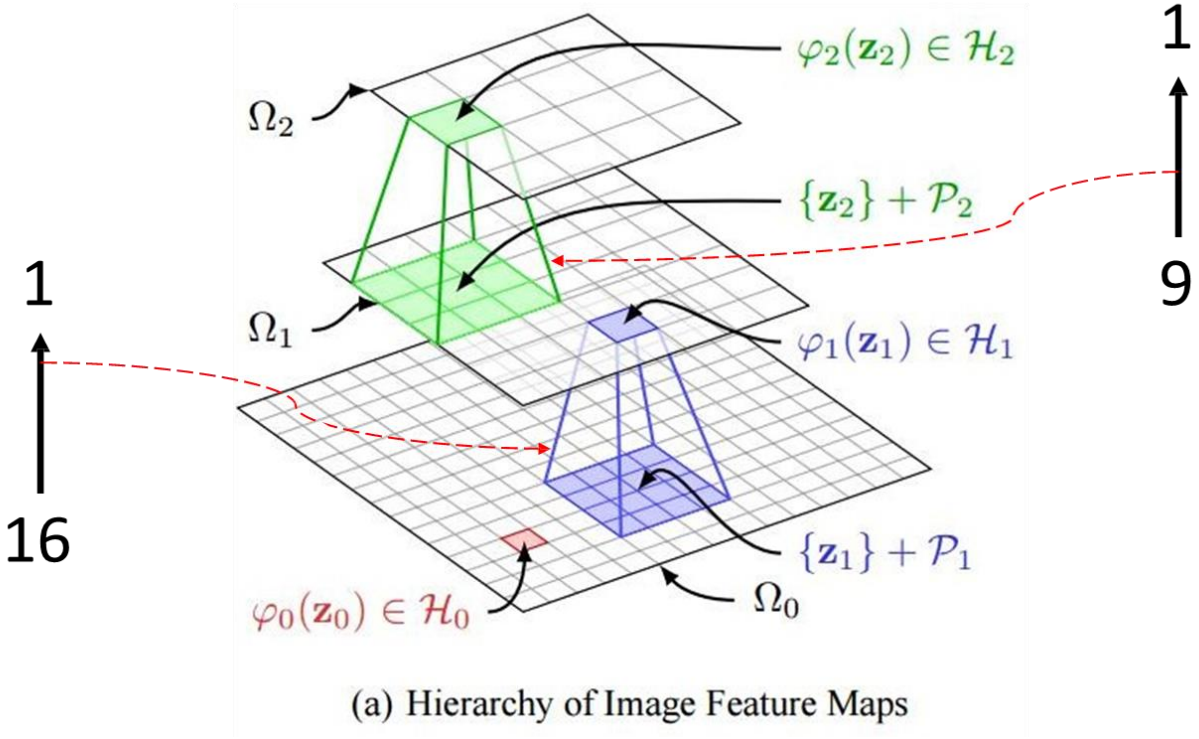
=

[1x1] matrix

$$\sum_i^9 x_i * w_i$$

Data Transformation for Deep Learning NLP

Convolution Neural Network (1)

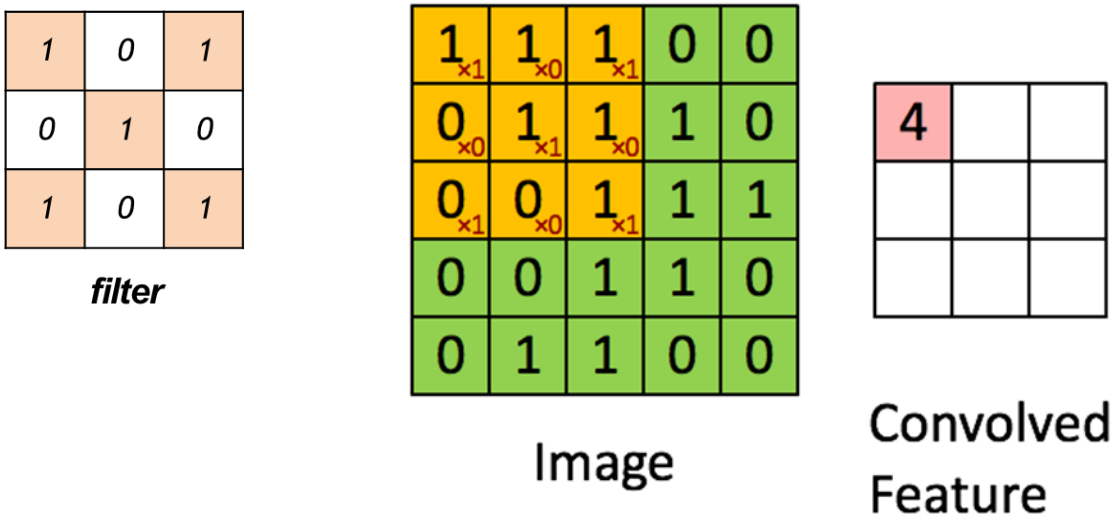


Data Abstraction

Mairal et al., Convolutional Kernel Networks, 2014

Data Transformation for Deep Learning NLP

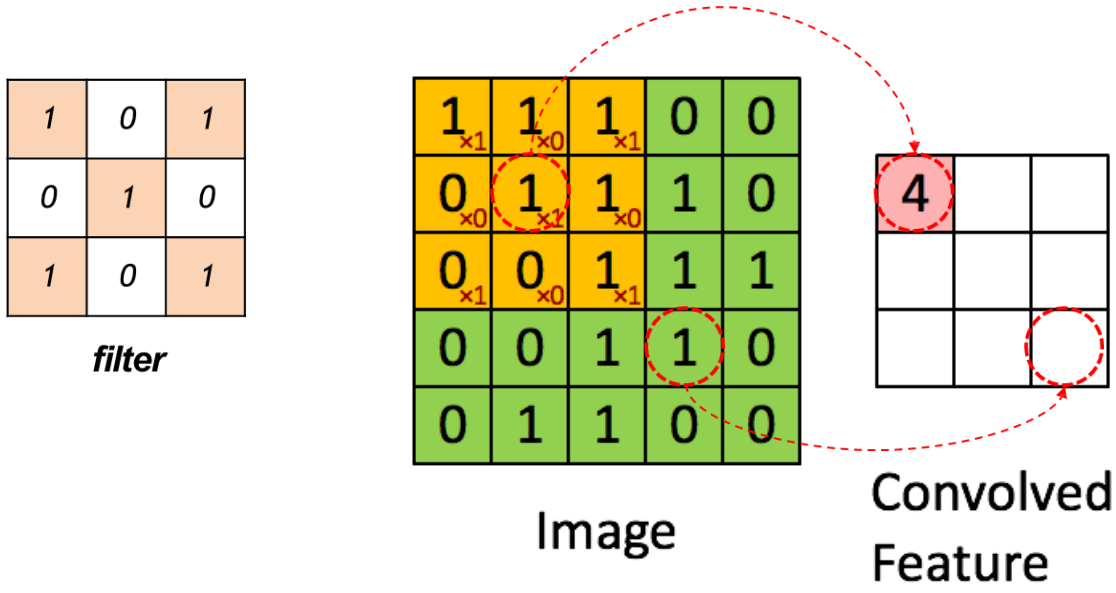
Convolution Neural Network (2)



Mairal et al., Convolutional Kernel Networks, 2014

Data Transformation for Deep Learning NLP

Convolution Neural Network (2)



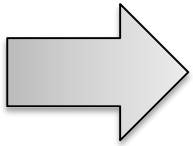
Mairal et al., Convolutional Kernel Networks, 2014

Data Transformation for Deep Learning NLP

V to V'

V= 9

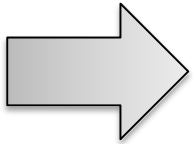
10	2	8
2	15	3
5	1	5



V'= 2

?
?

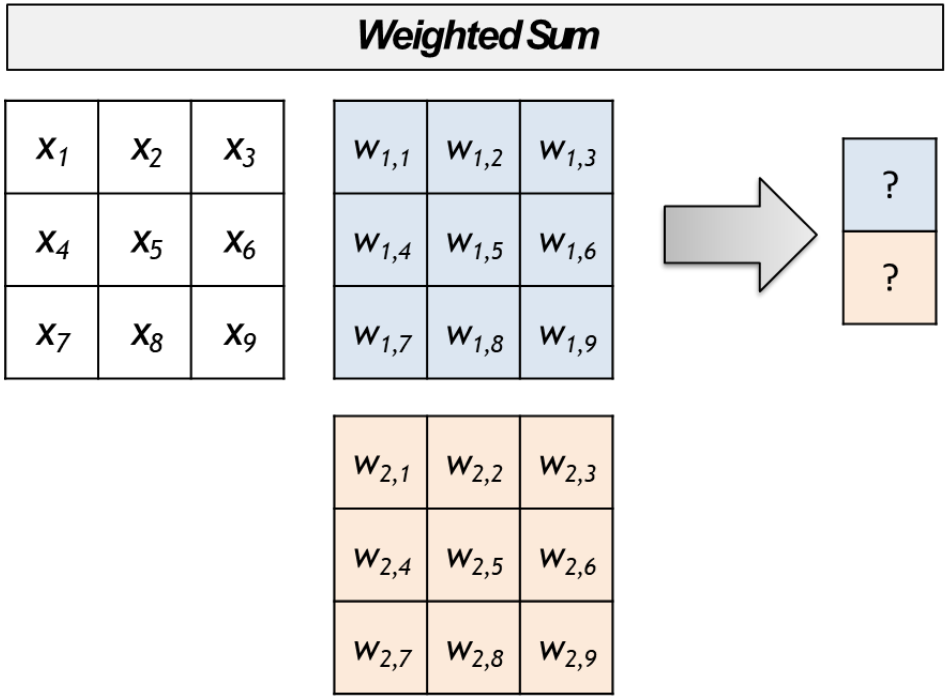
10	*2*	*8*	*2*	*15*	*3*	*5*	*1*	*5*



?	*?*

Data Transformation for Deep Learning NLP

V to V' – generalized method



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

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

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

=

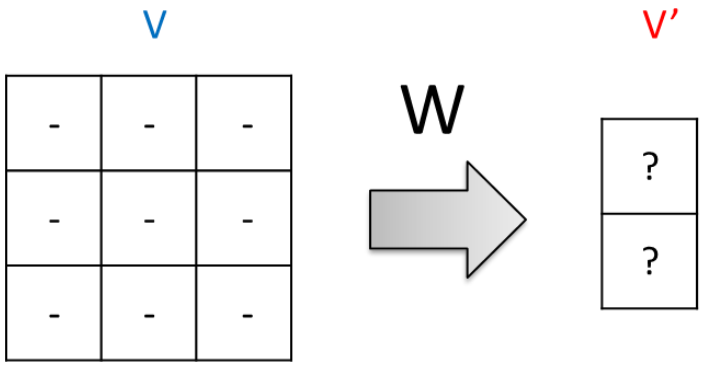
[1x2] matrix

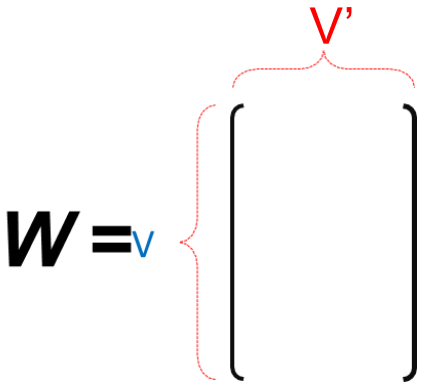
$$\left[\sum_i^9 x_i * w_{1,i}, \sum_i^9 x_i * w_{2,i} \right]$$

Fully Connected Network

Data Transformation for Deep Learning NLP

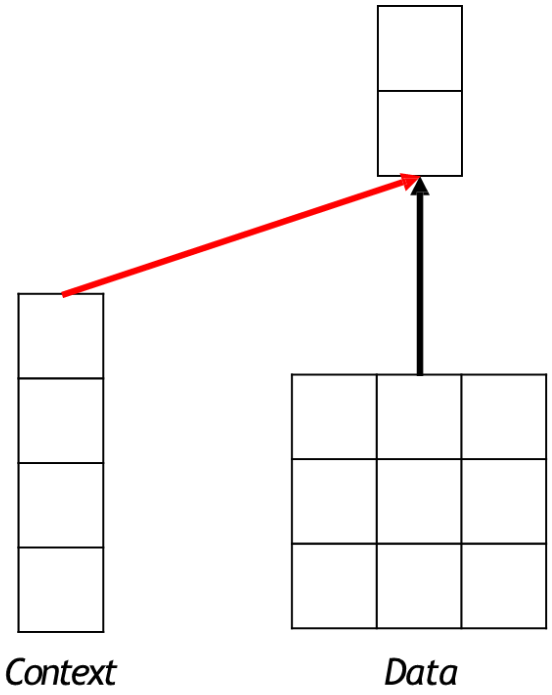
V to V' – Projection Notation





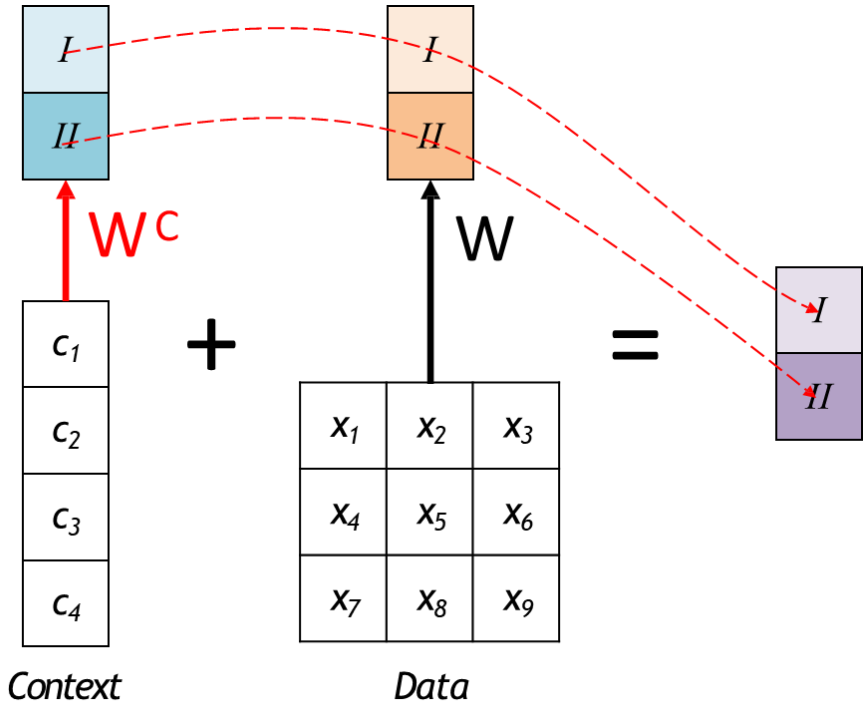
Data Transformation for Deep Learning NLP

V to V' – Projection with Context (1)



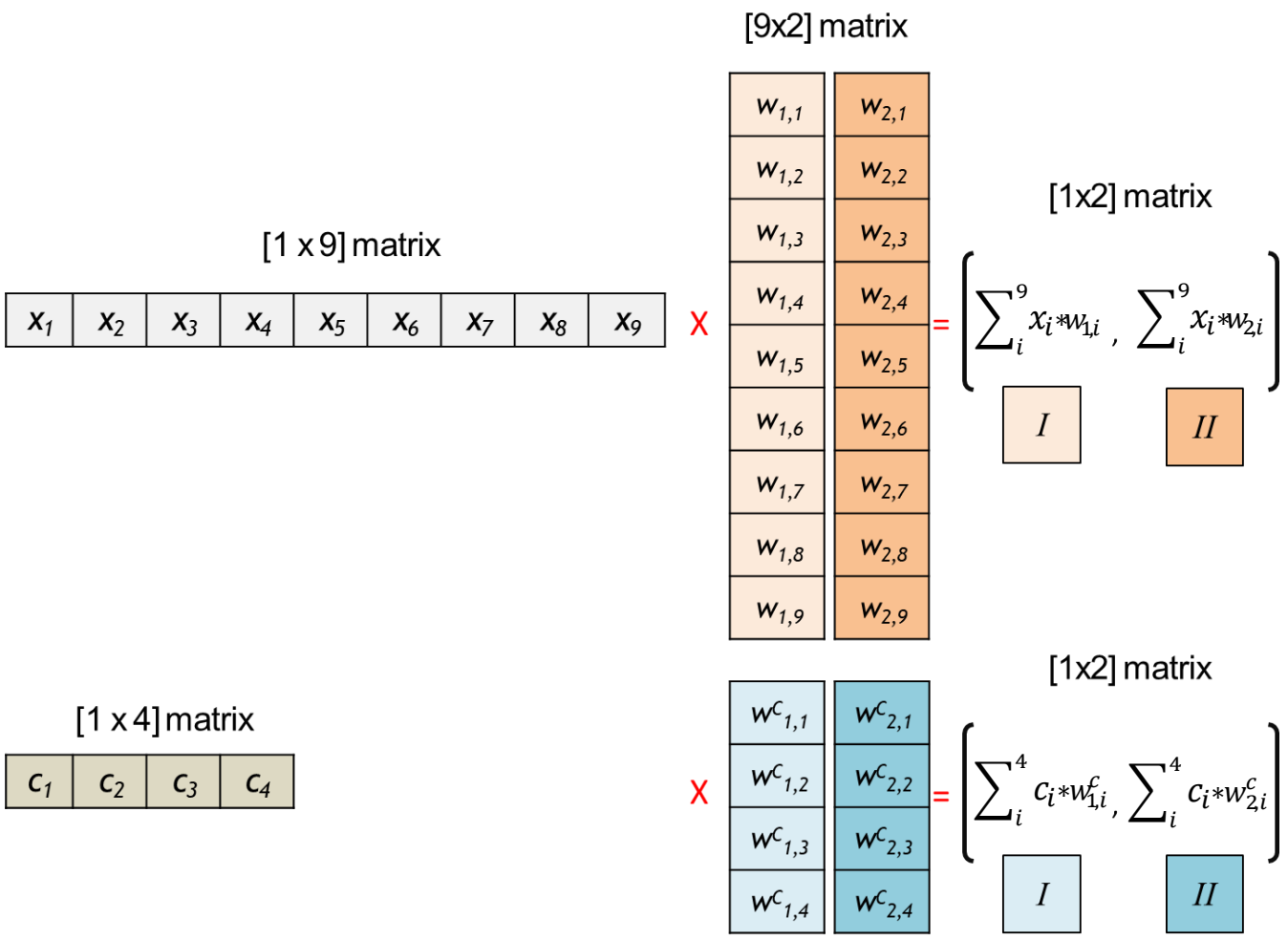
Data Transformation for Deep Learning NLP

V to V' – Projection with Context (2)



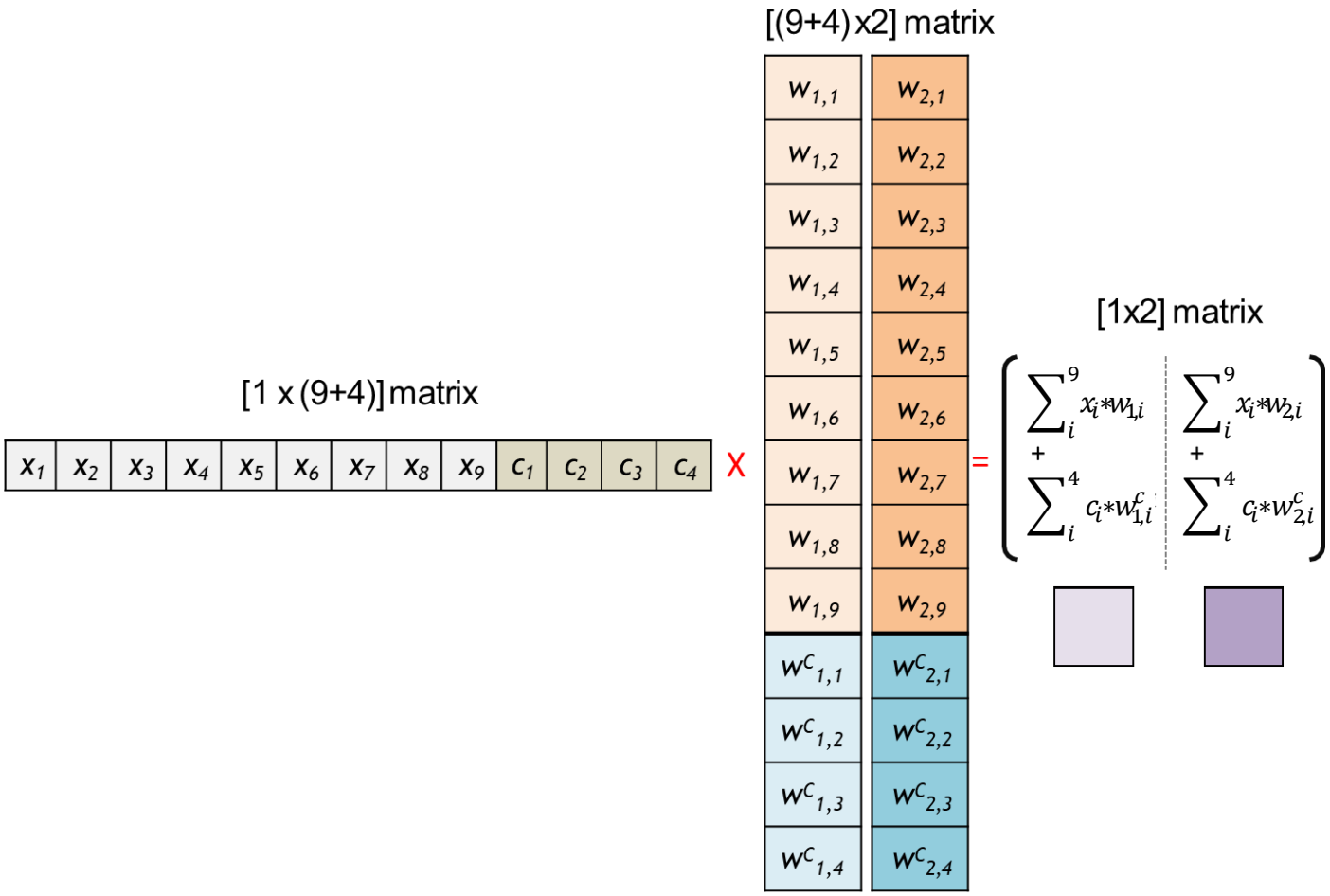
Data Transformation for Deep Learning NLP

V to V' with Context - Linear Algebra



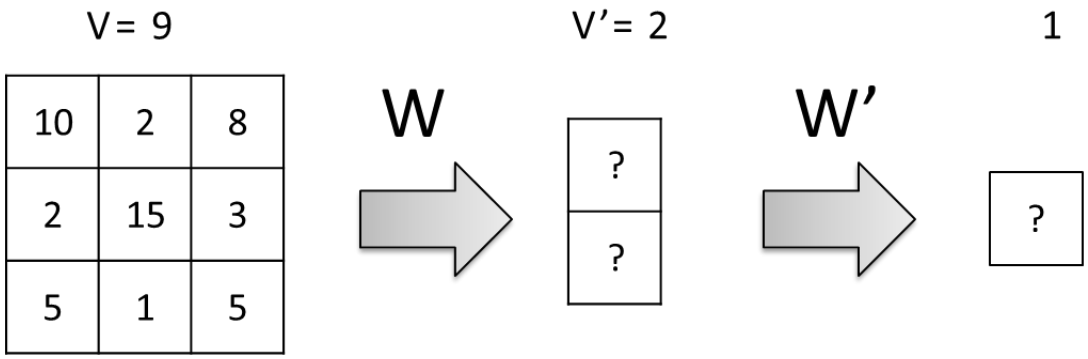
Data Transformation for Deep Learning NLP

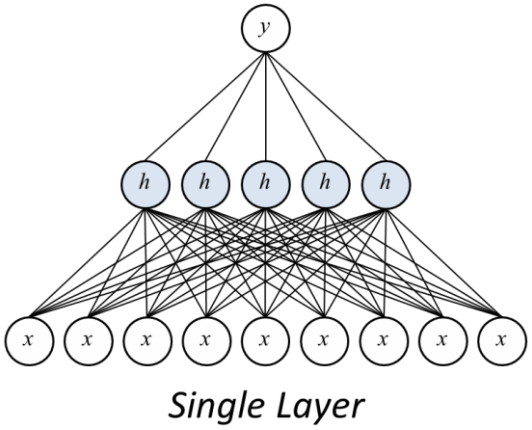
V to V' with Context - Linear Algebra (Simplified)



Data Transformation for Deep Learning NLP

V → V' → 1

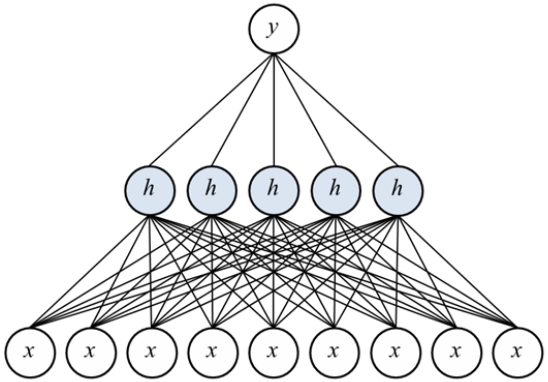




Single Layer

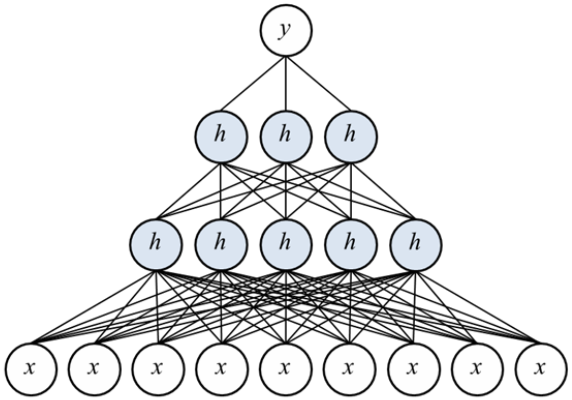
Data Transformation for Deep Learning NLP

V → V' → 1



Single Layer

V →V' →1



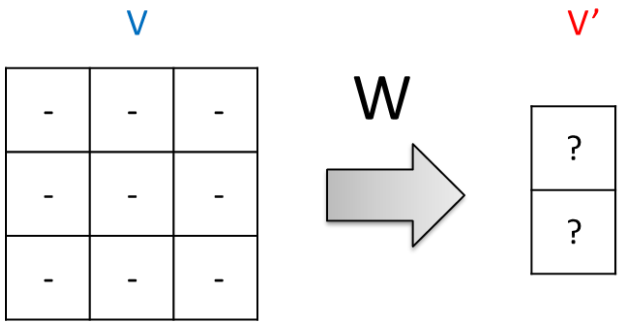
Multilayer

V →V' →V'' →1

Data Transformation for Deep Learning NLP

Seq2Seq Encoding

Single Item Summarisation



Multiple Item Summarisation

?

Data Transformation for Deep Learning NLP

Multiple Item Summarisation

10	2	8
2	15	3
5	1	5

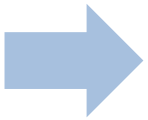
Data 1

13	4	8
4	5	2
1	45	31

Data 2

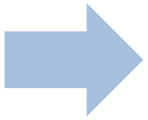
6	3	4
1	7	1
3	4	0

Data 3



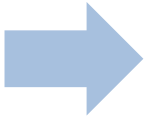
?	?	?
?	?	?
?	?	?

V



?
?

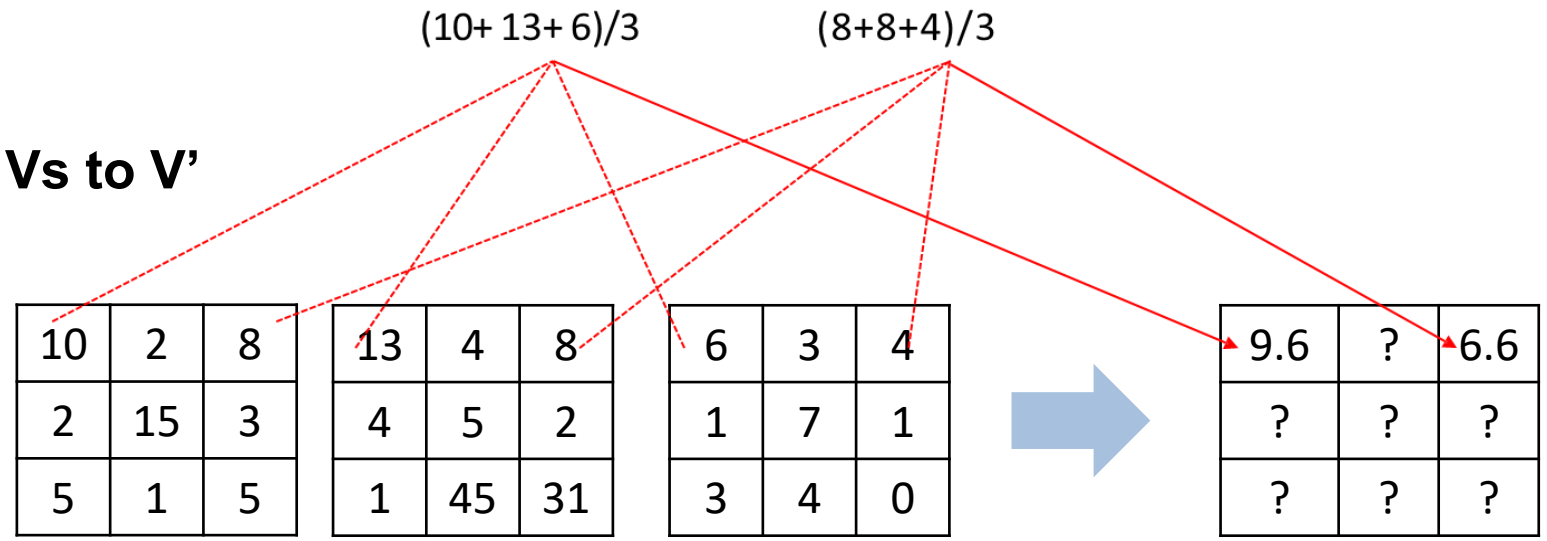
V'



?

1

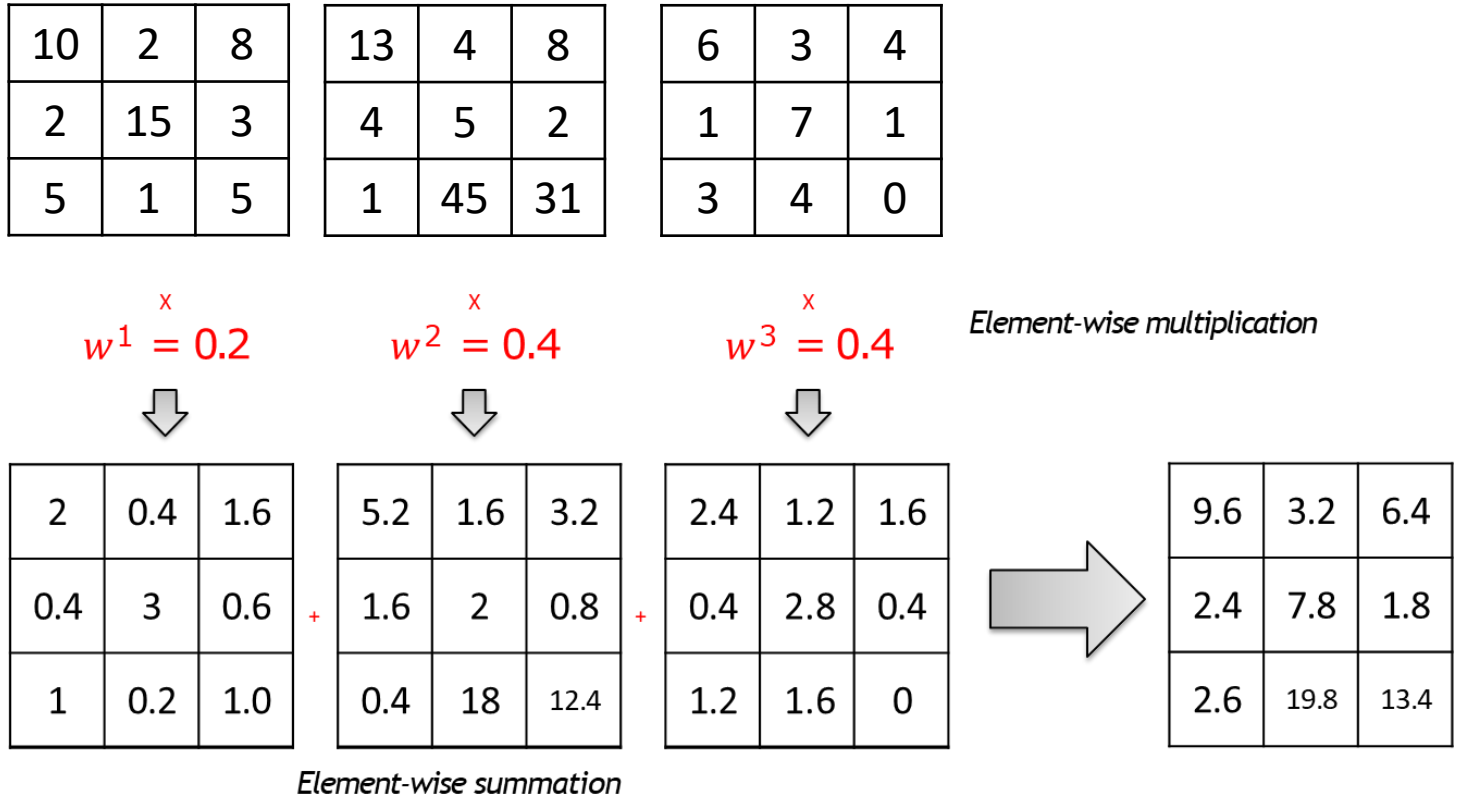
Data Transformation for Deep Learning NLP



Element-wise Average

Data Transformation for Deep Learning NLP

Vs to V'

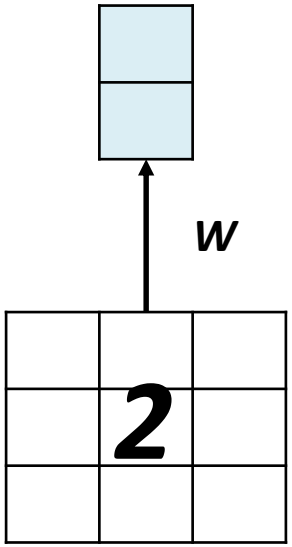


Data Transformation for Deep Learning NLP

Temporal Summarisation

Data Transformation for Deep Learning NLP

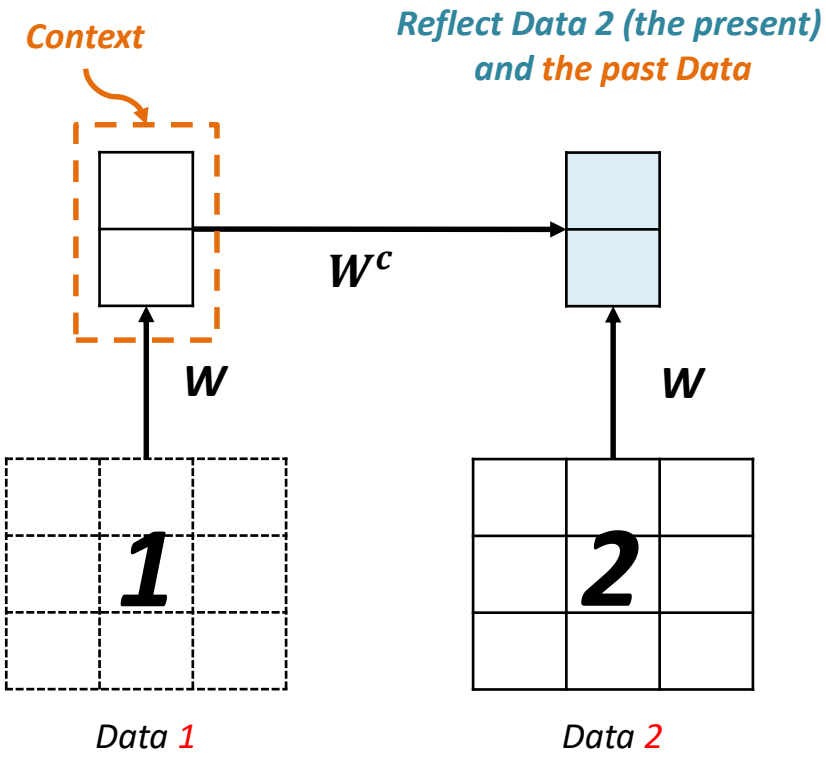
Vs → V's → V'



Data 2

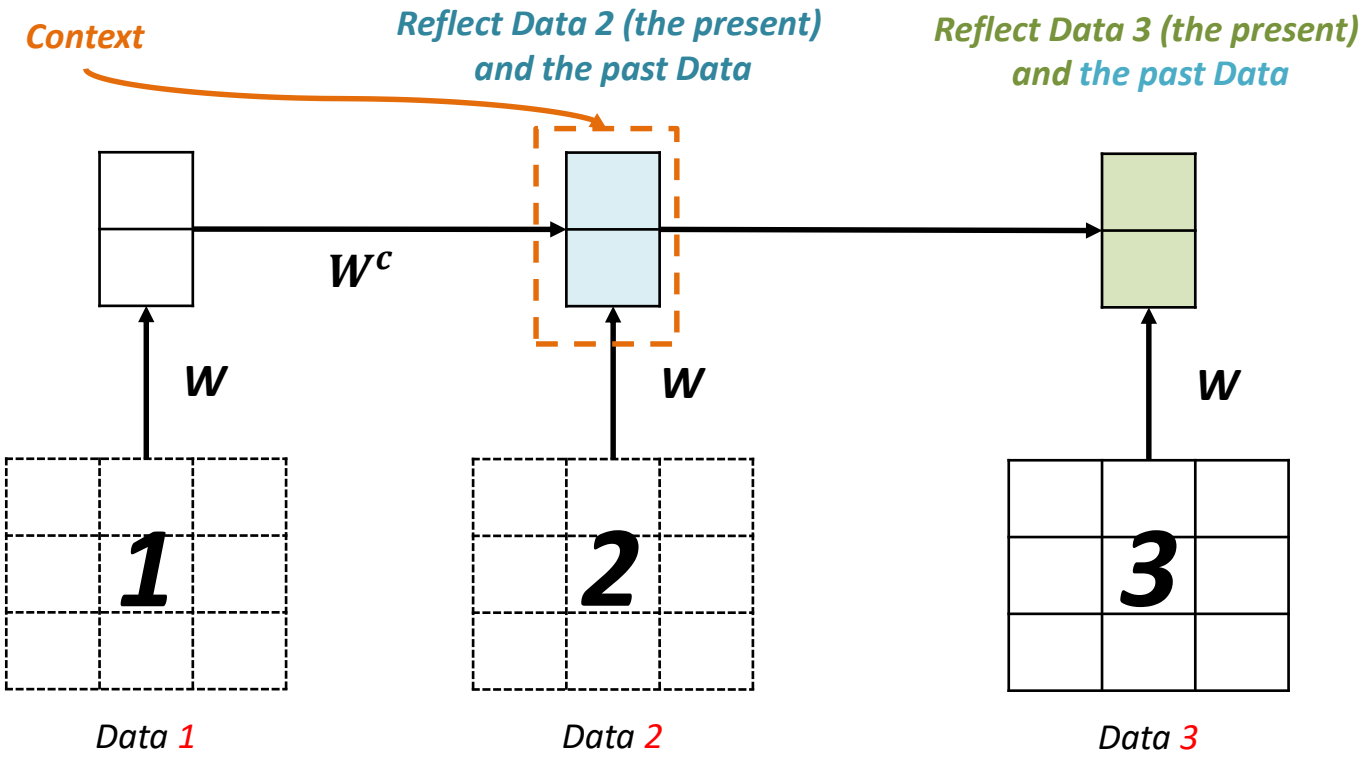
Data Transformation for Deep Learning NLP

Vs → V's → V'



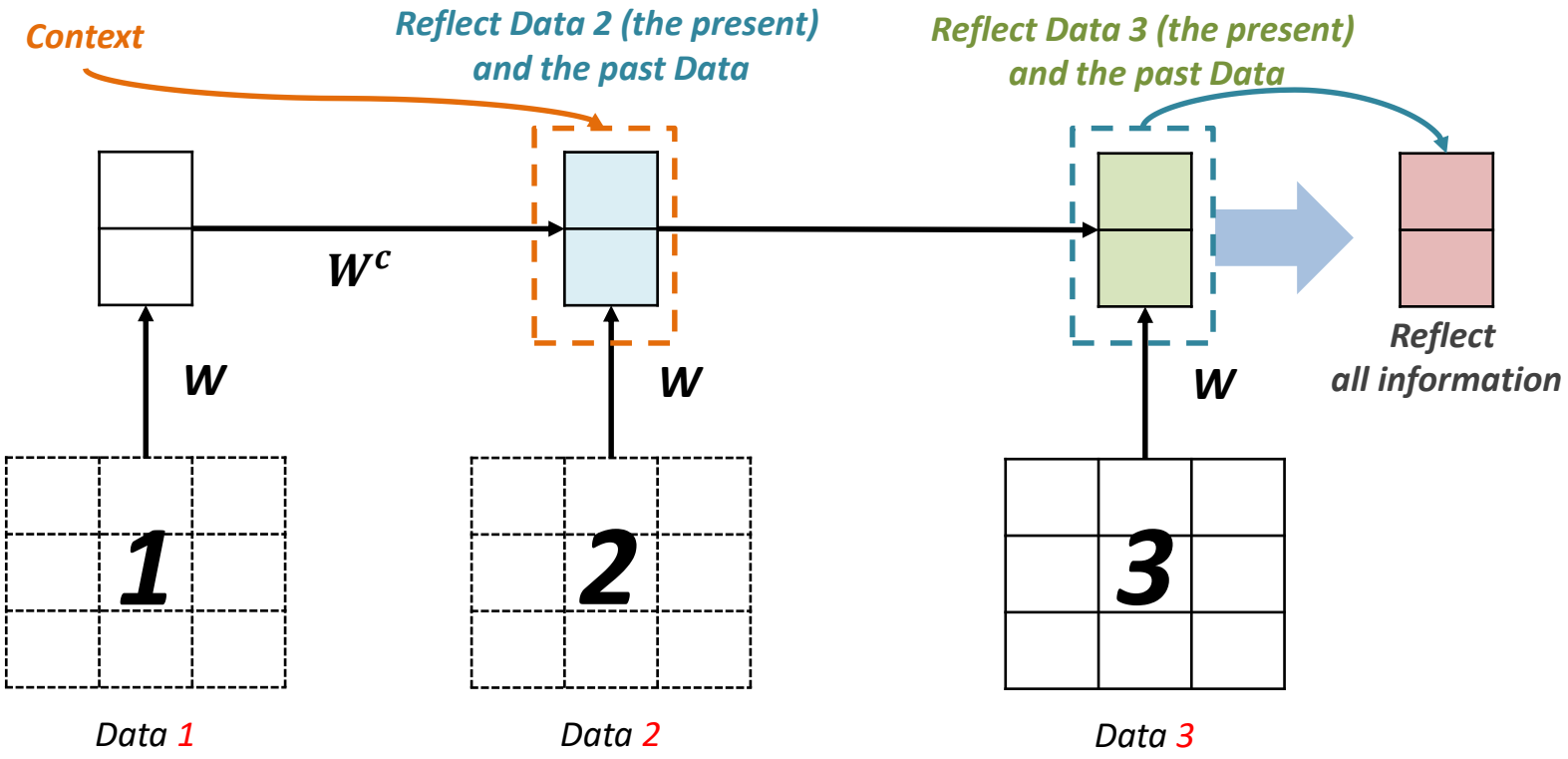
Data Transformation for Deep Learning NLP

Vs → V's → V'



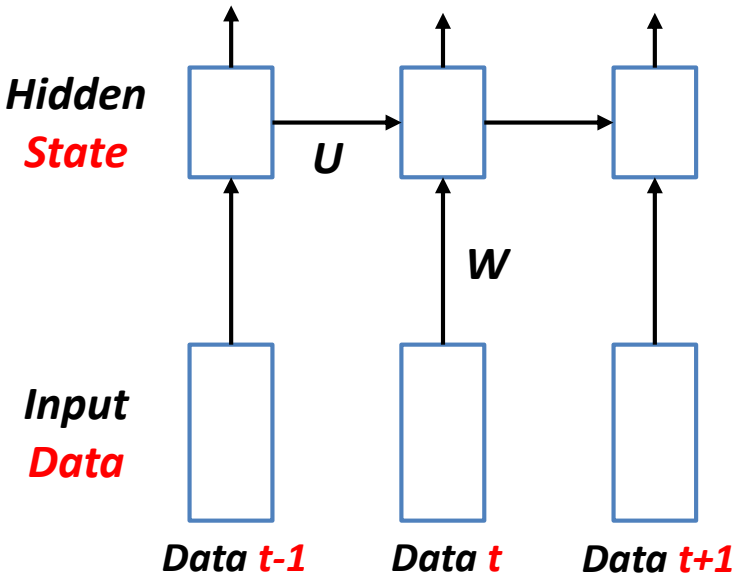
Data Transformation for Deep Learning NLP

Vs → V's → V'

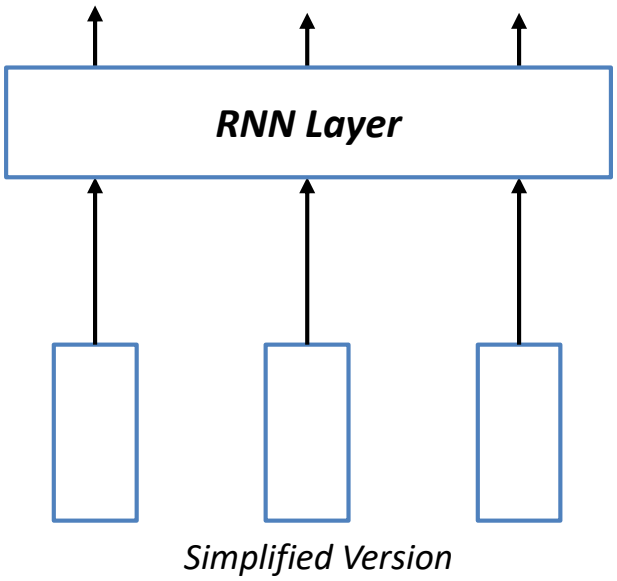


Data Transformation for Deep Learning NLP

Graphical Notation



Vs → V's

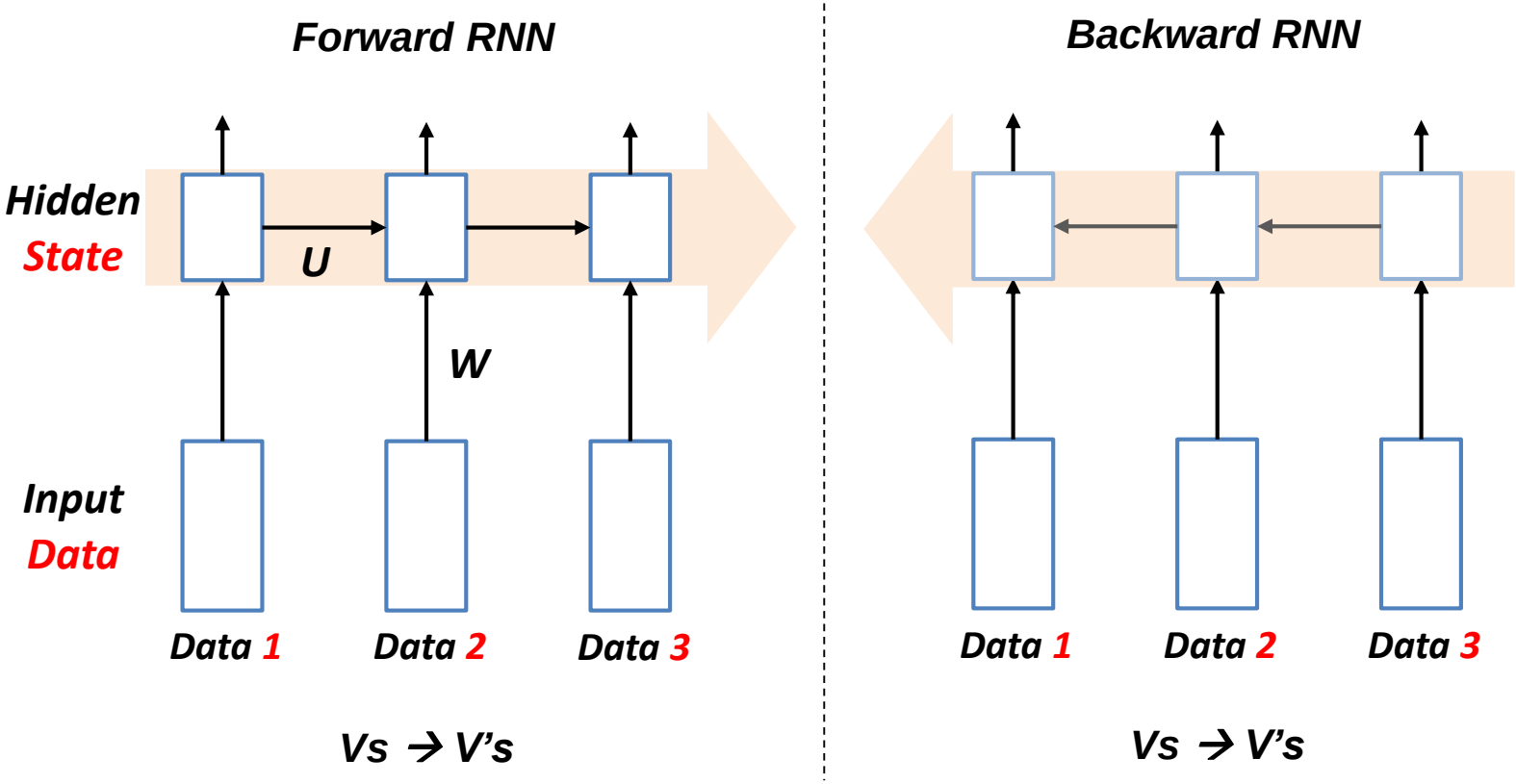


Simplified Version

Vs → V's

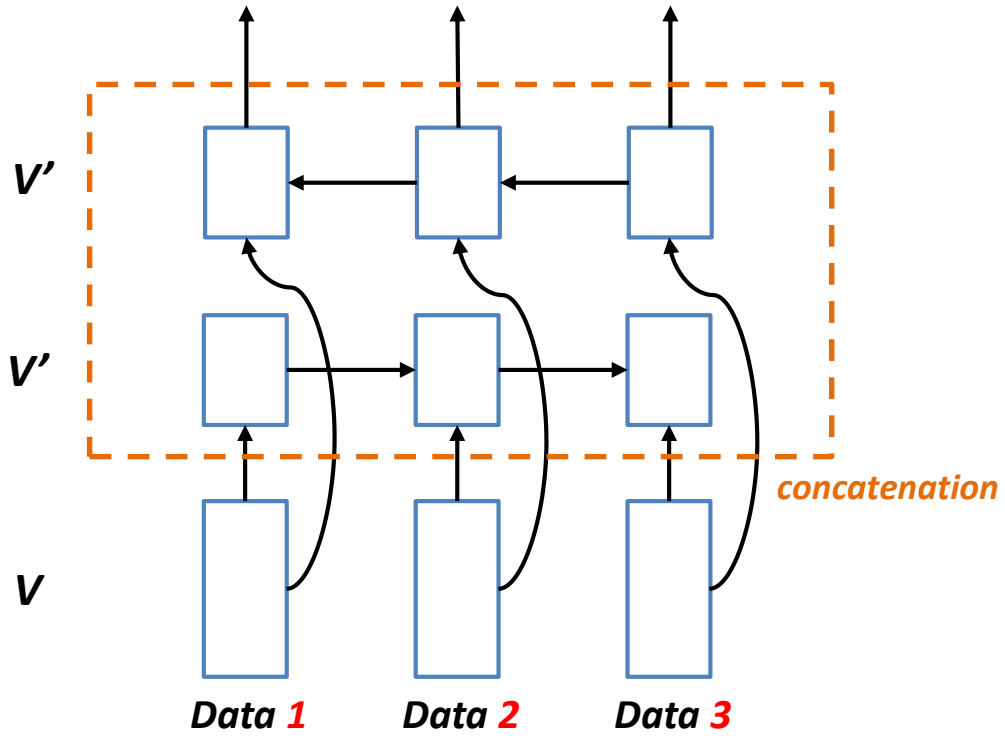
Data Transformation for Deep Learning NLP

Forward/Backward RNN



Data Transformation for Deep Learning NLP

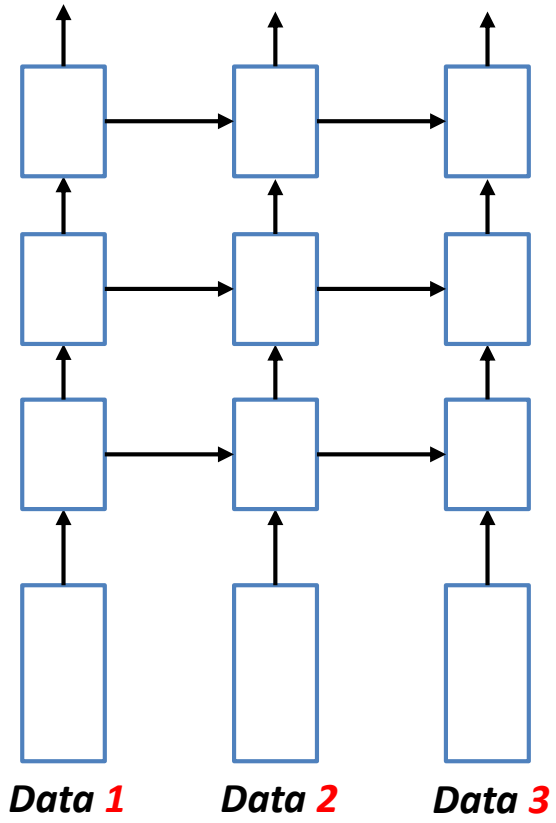
Bidirectional RNN



*Vs → (2*V')s*

Data Transformation for Deep Learning NLP

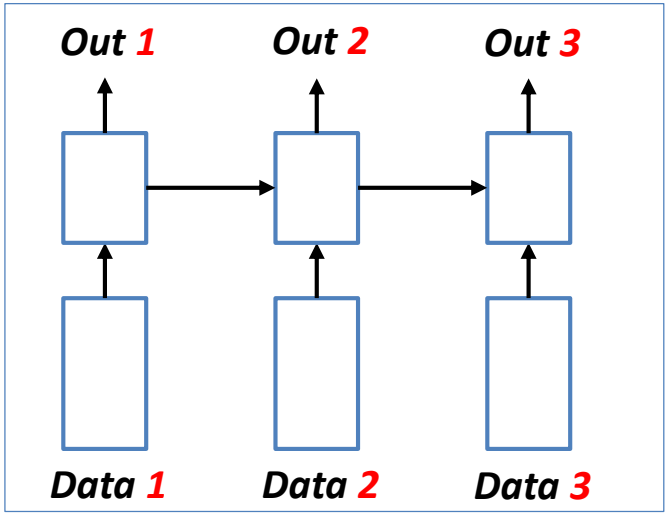
Stacking RNN



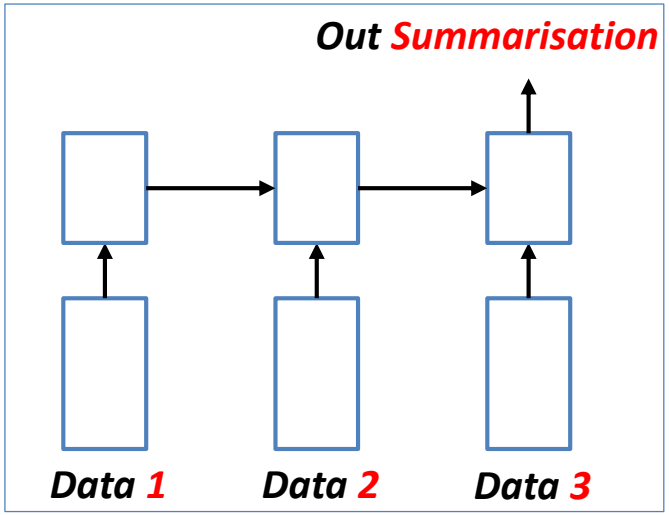
Vs → V's

Data Transformation for Deep Learning NLP

RNN: Input and Output



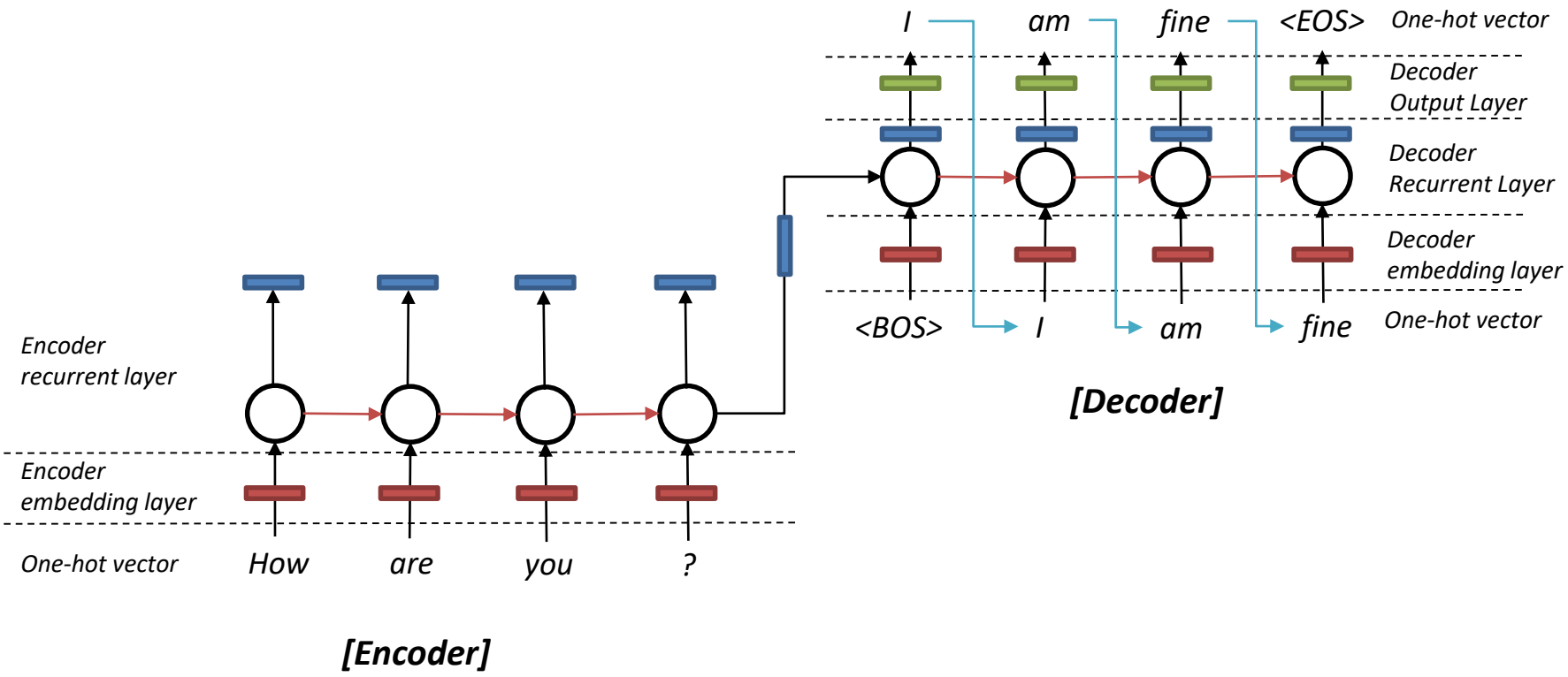
- ✓ *Vs → V's*
- ✓ *Len(Vs) → Len(V's)*



- ✓ *Vs → 1*

Data Transformation for Deep Learning NLP

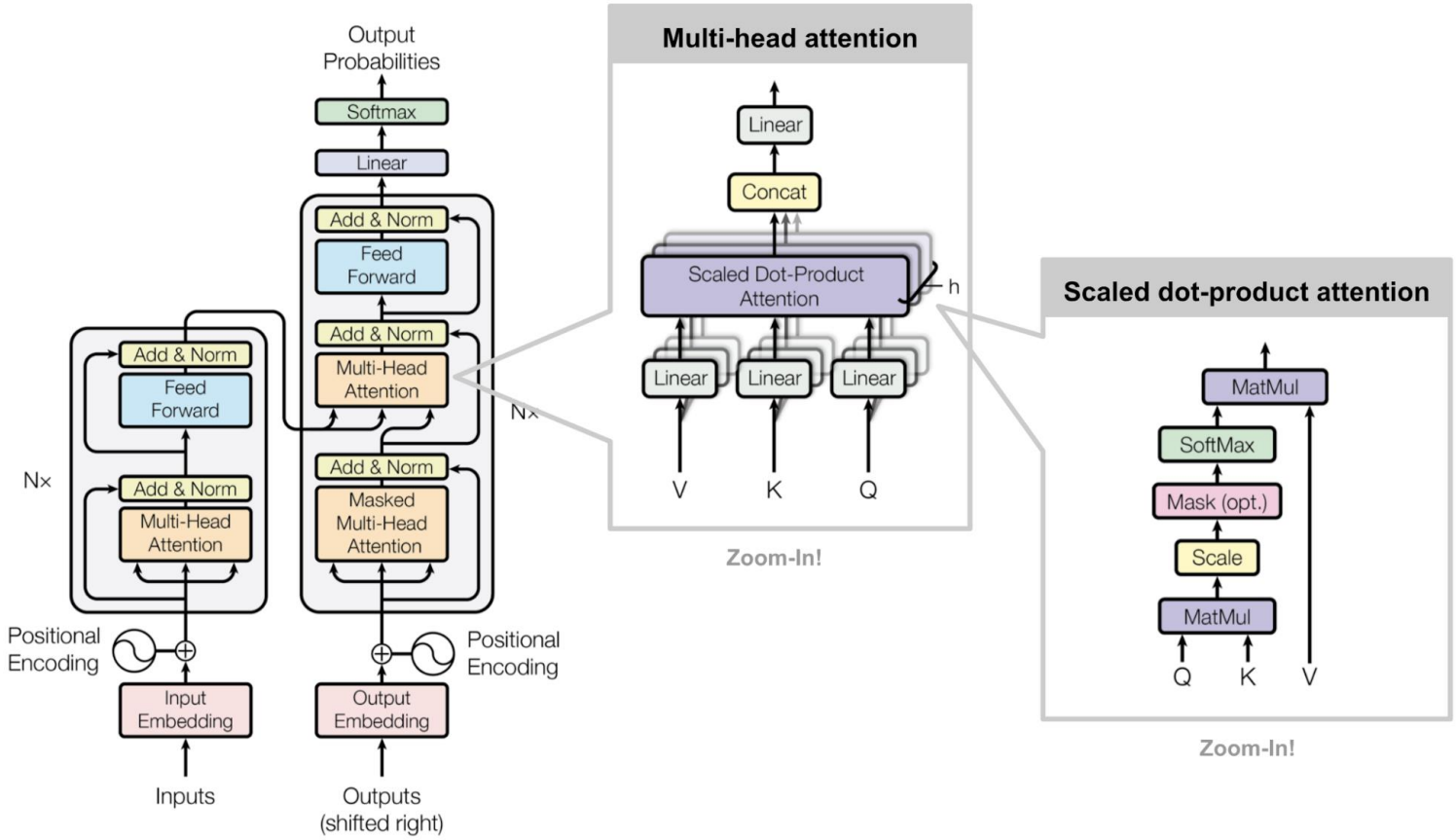
Seq2Seq Encoding and Decoding- Dialog System



Learn Encoder Decoder in Week 8

Data Transformation for Deep Learning NLP

I don’t like Recurrent models!! - Global model



Learn Encoder Decoder in Week 11 and 12

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

5 Next Week Preview

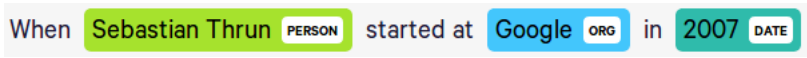
The purpose of Natural Language Processing: Overview

Application

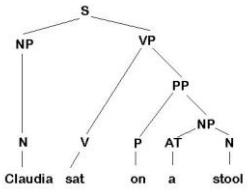
- ***Understanding***
  - ***Sentiment Analysis***
  - ***Topic Classification***
  - ***Topic Modelling***
  - ***....***
- ***Dialog***
- ***Translation***
- ***Searching***
  - ***Search***
  - ***....***

NLP Stack

- ***Entity Extraction*** — When Sebastian Thrun …



- ***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