

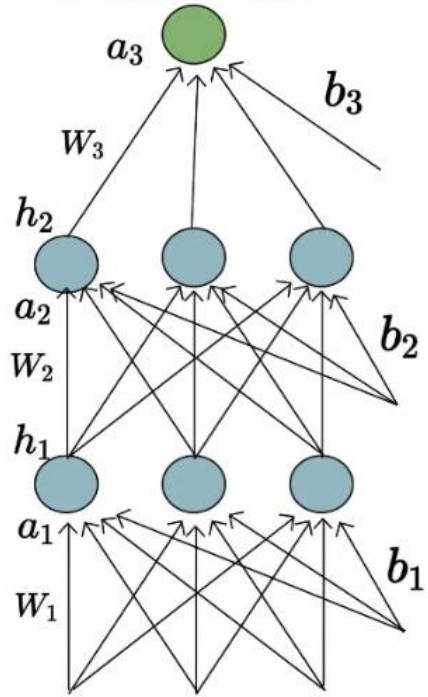
Deep Learning : Sequence Learning Problems



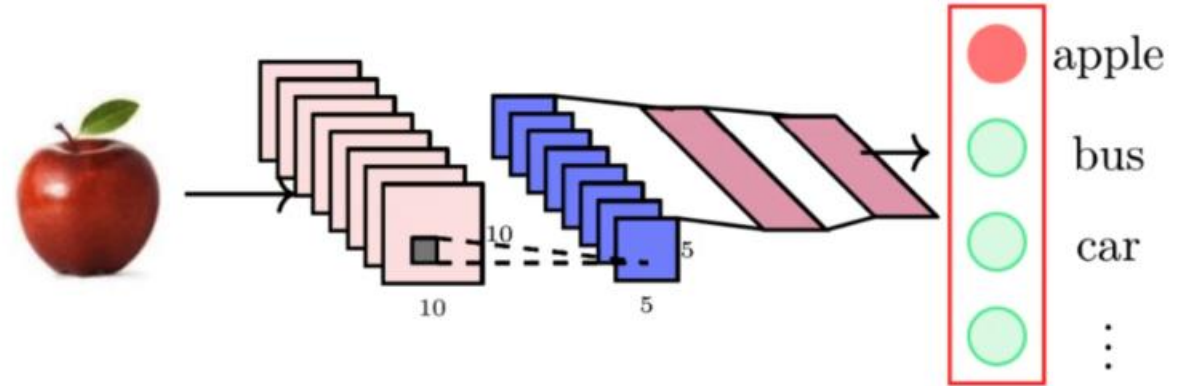
राष्ट्रीय प्रौद्योगिकी संस्थान सिक्किम
NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM

Course Instructor:
Dr. Bam Bahadur Sinha
Assistant Professor
Computer Science & Engineering
National Institute of Technology
Sikkim

Health Risk

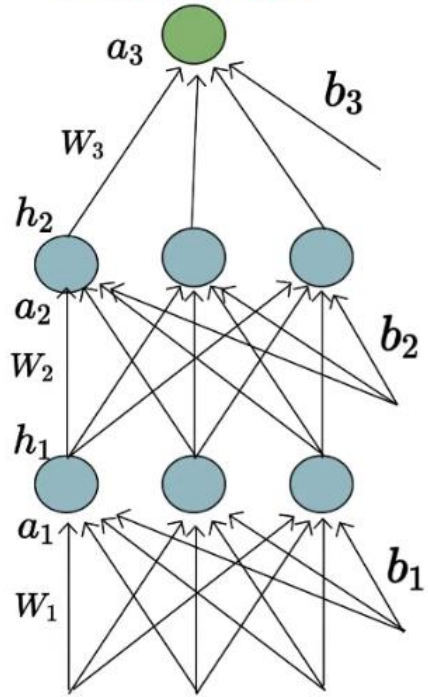


x_1 height weight sugar bp ECG

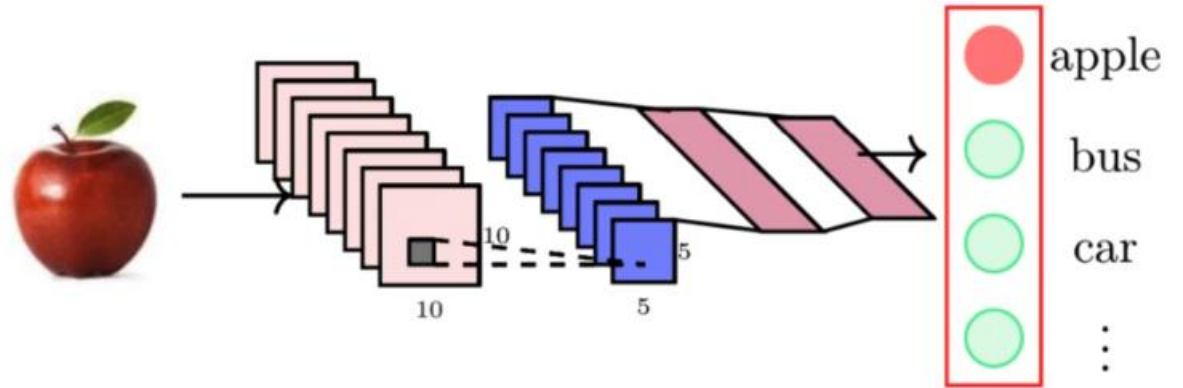


What kind of networks we have seen so far?

Health Risk

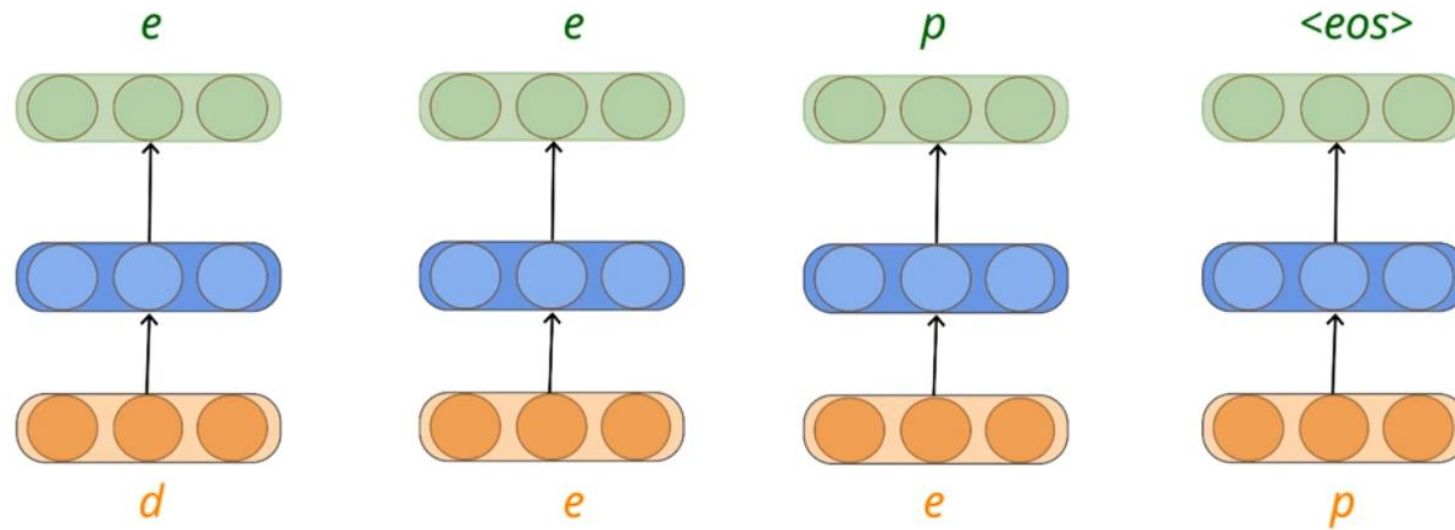


x_1 height weight sugar bp ECG



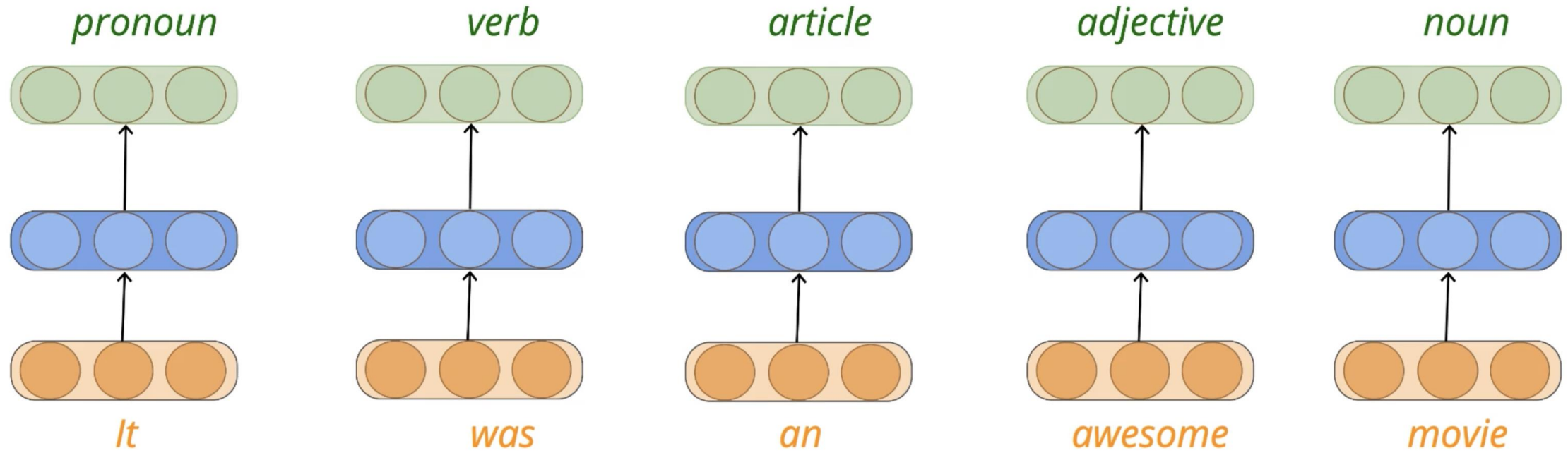
- ✓ Outputs are independent of previous inputs
- ✓ Input is of a fixed length

What kind of networks we have seen so far?

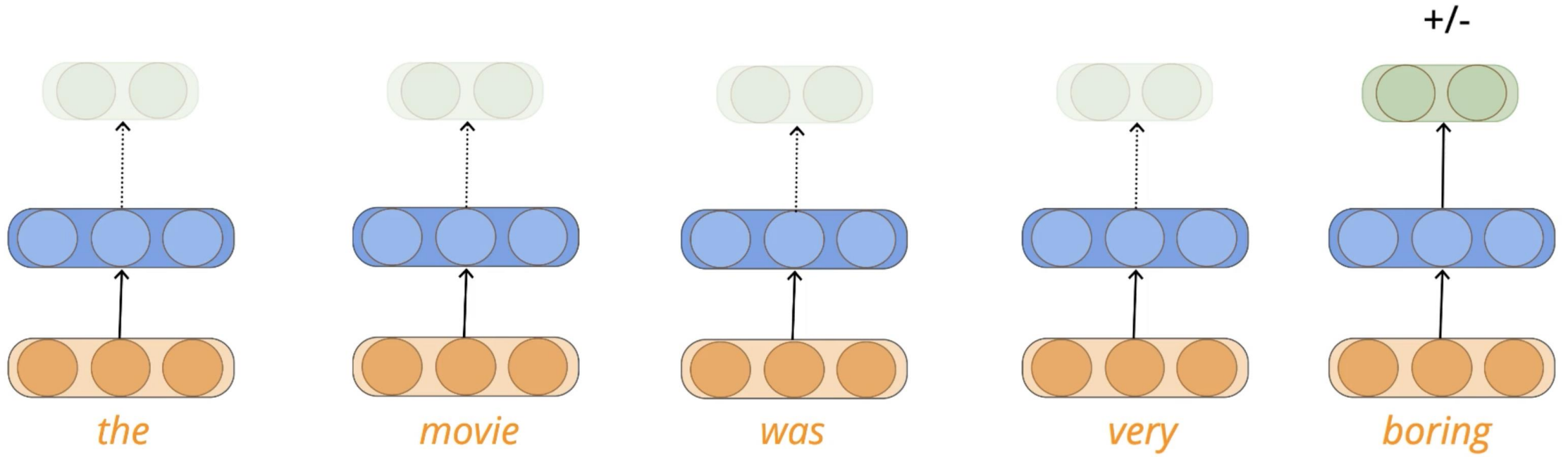


- ✓ Outputs depend on previous inputs also
- ✓ The length of the input is not fixed

Sequence Learning Problems
Eg: Autotyping, Auto text completion



Sequence of Words



Sentiment Classification

(Do we need to produce an output at every time step?)

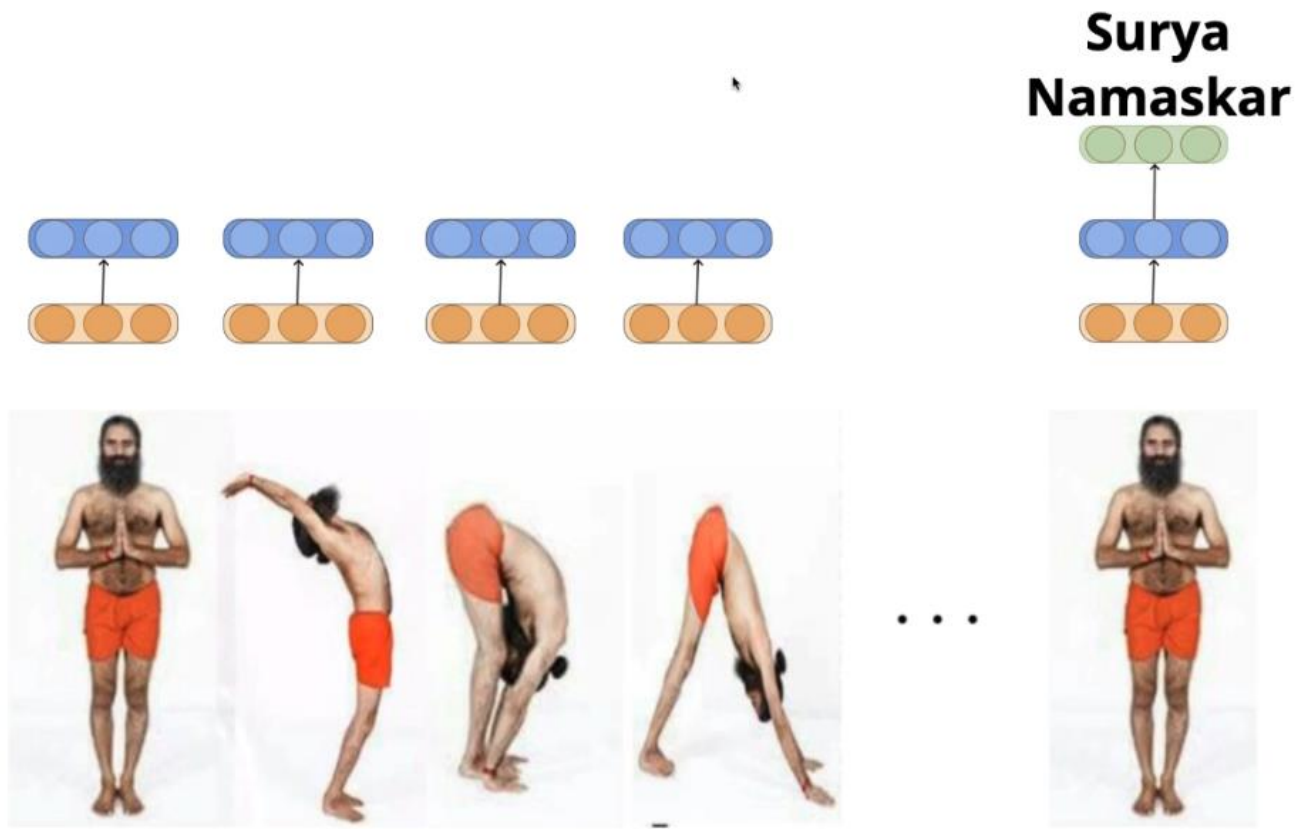


Speech

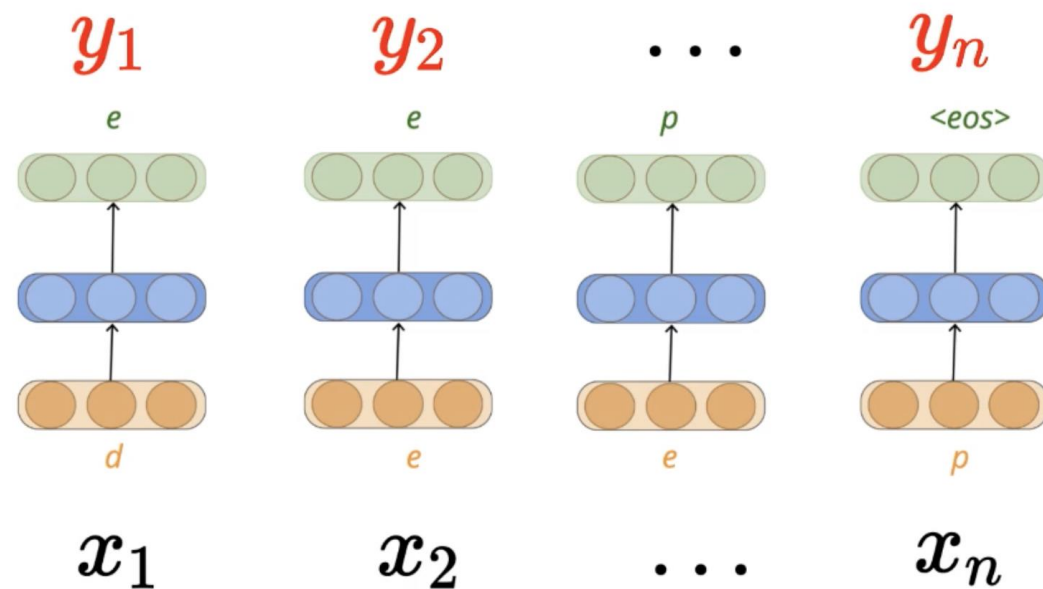


Video

What are some other types of sequences that we encounter?



What are some other types of sequences that we encounter?



$$y_t = \hat{f}(x_1, x_2, \dots, x_t)$$

How do we model such sequence learning problems?

Wishlist while modelling such sequence problems

1

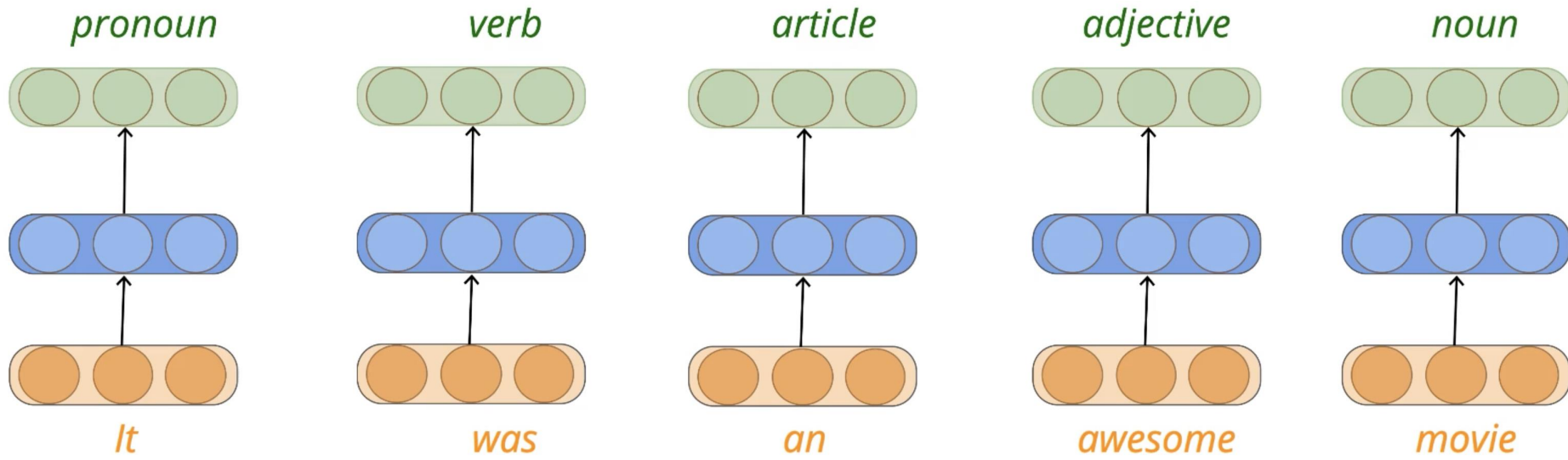
Ensure that Y_t is dependent on previous inputs also

2

Ensure that the function can deal with variable number of inputs

3

Ensure that the function executed at each time step is the same.



$$h_i = \sigma(W_1 x_i + b_1)$$

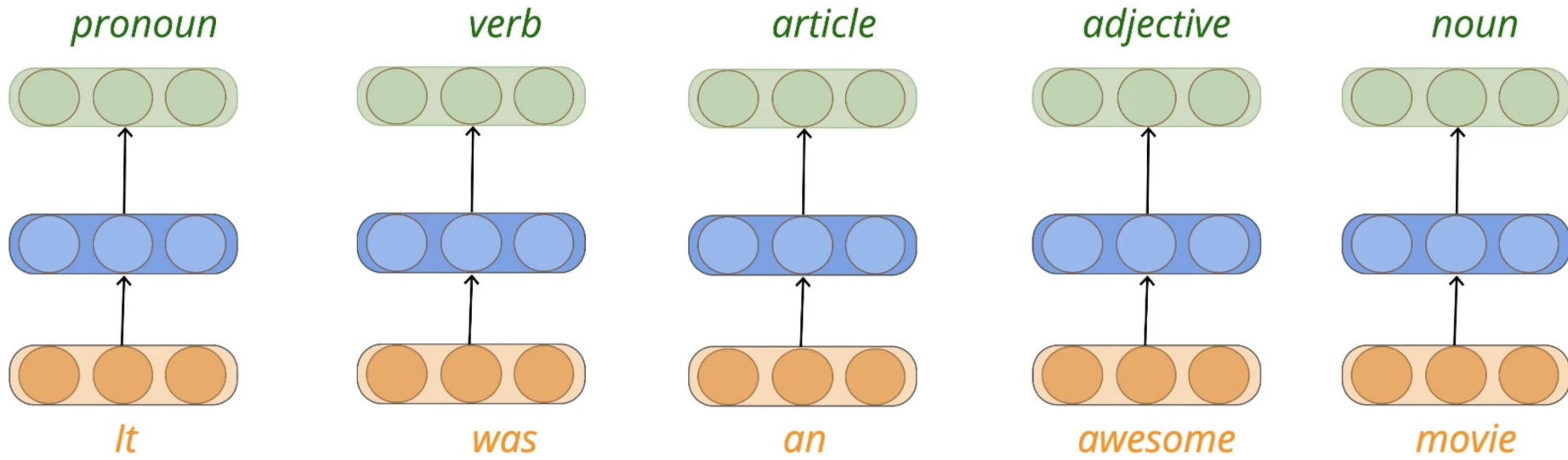
$$y_i = O(W_2 h_i + b_2)$$

$$i = \text{timestep}$$

$$s_i = \sigma(U x_i + b)$$

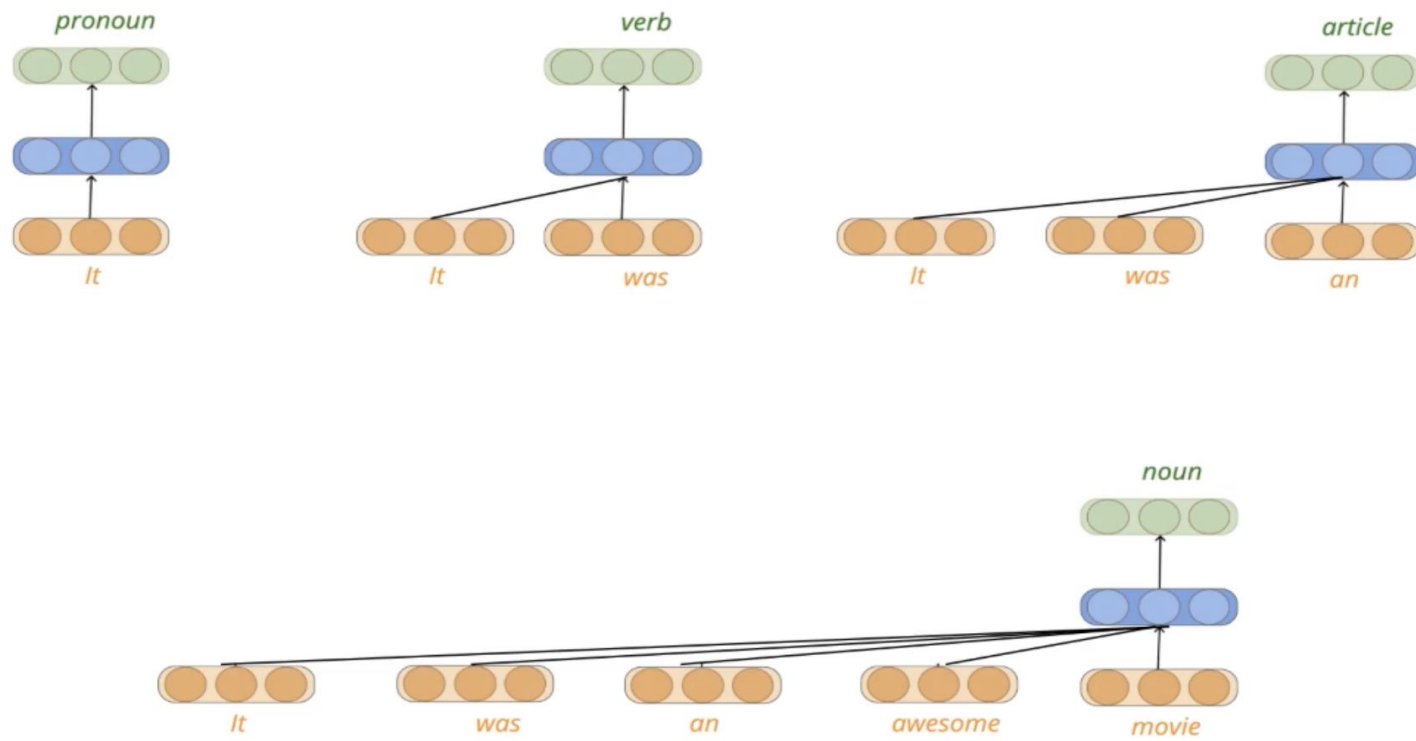
$$y_i = O(V s_i + c)$$

Ensure that the function executed at each time step is the same.



$$s_i = \sigma(Ux_i + b)$$
$$y_i = O(Vs_i + c)$$

Ensure that the function can deal with variable number of inputs



$$s_i = \sigma(Ux_i + b)$$

$$y_i = O(Vs_i + c)$$

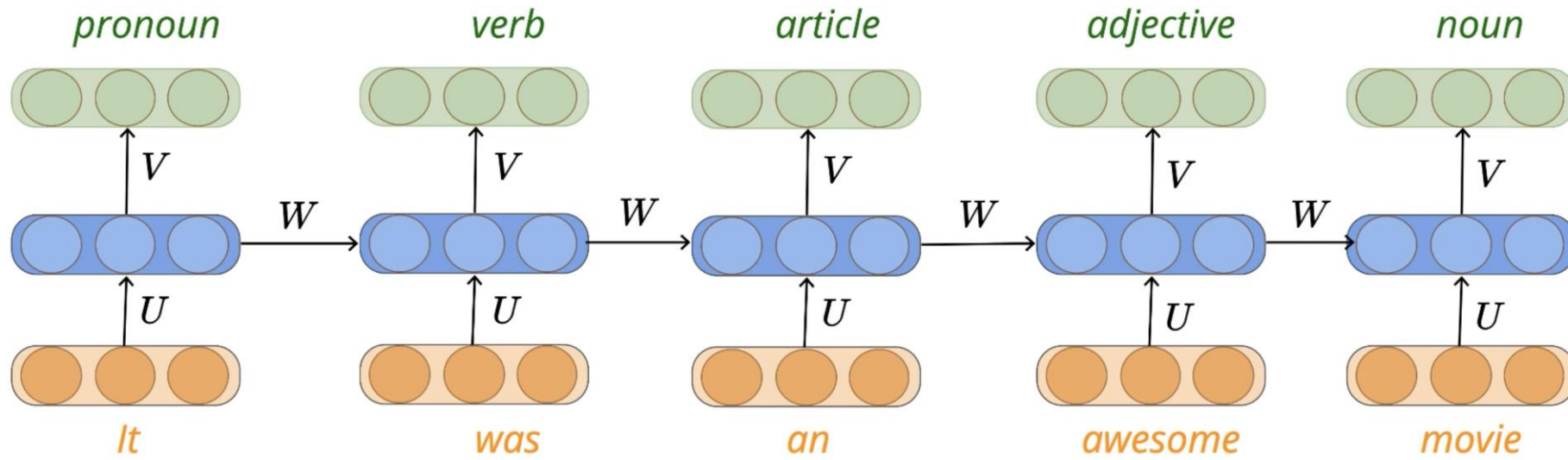
$$y_1 = f(x_1)$$

$$y_2 = f(x_1, x_2)$$

$$y_3 = f(x_1, x_2, x_3)$$

Ensure that Y_t is dependent on previous inputs also

What is the problem in this mathematical model?



$$s_i = \sigma(Ux_i + Ws_{i-1} + b)$$

$$y_i = O(Vs_i + c)$$

Recurrent Neural Network (RNN)

Sequence Problem	I/P	O/P
Sequence Classification	N	1
Sequence Labelling	N	N
Sequence Generation	N	M

- ✓ How do you represent words and characters as numbers ? **(data and tasks)**
- ✓ What is an appropriate loss function ? **(loss)**
- ✓ How do you train the model ? **(learning algorithm)**

Recurrent Neural Network (RNN)

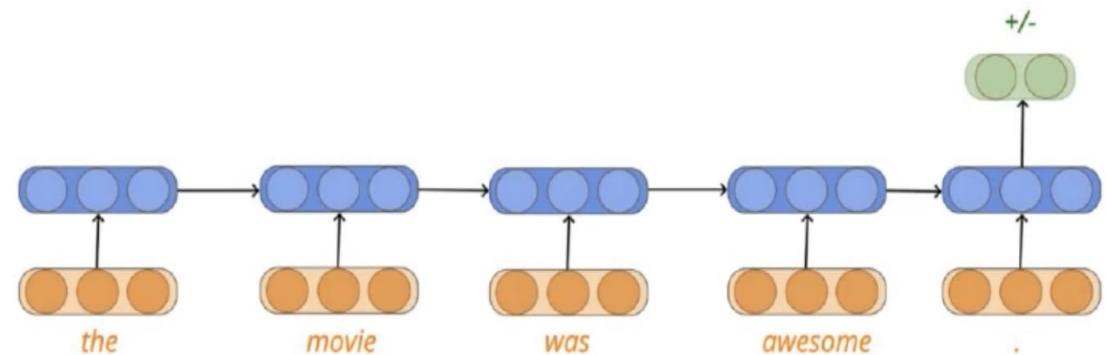
- ✓ **Sequence Classification** (sentiment classification, video classification)
- ✓ **Sequence Labelling** (part of speech tagging, named entity recognition)
- ✓ **Sequence Generation** (machine translation, transliteration)

Recurrent Neural Network (RNN)

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
<sos>	The	first	half	was	very	boring	.	<eos>	<pad>
<sos>	Great	performance	by	all	the	lead	actors	.	<eos>
<sos>	The	visual	effects	were	stunning	.	<eos>	<pad>	<pad>
<sos>	The	movie	was	a	waste	of	time	.	<eos>

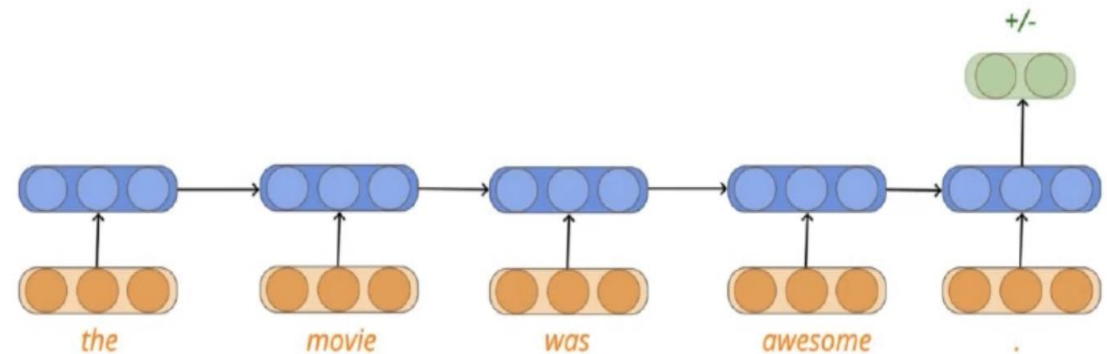
y
0
1
1
0

Pre-process the data
(Sequence Classification Problem)



- ✓ Define special symbols: <sos>, <eos>, <pad>
- ✓ Prepend/Append <sos>, <eos> to each sequence
- ✓ Find maximum input length across all sequences (say, 10)
- ✓ Add special word <pad> to all shorter sequences so that they become of the same length (10, in this case)

Pre-process the data (Sequence Classification Problem)

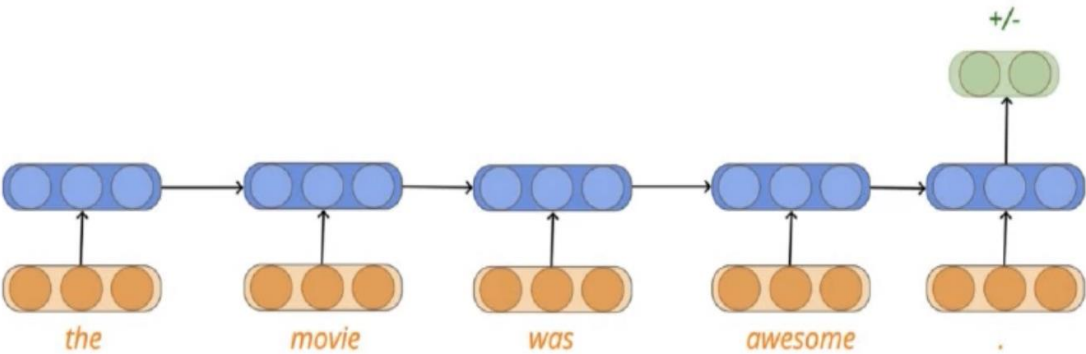


x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
<sos>	The	first	half	was	very	boring	.	<eos>	<pad>	0
<sos>	Great	performance	by	all	the	lead	actors	.	<eos>	1
<sos>	The	background	music	was	awesome	.	<eos>	<pad>	<pad>	1
<sos>	The	movie	was	a	waste	of	time	.	<eos>	0

word	id
<sos>	1
<eos>	2
<pad>	3
the	4
first	5
half	6
...	

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

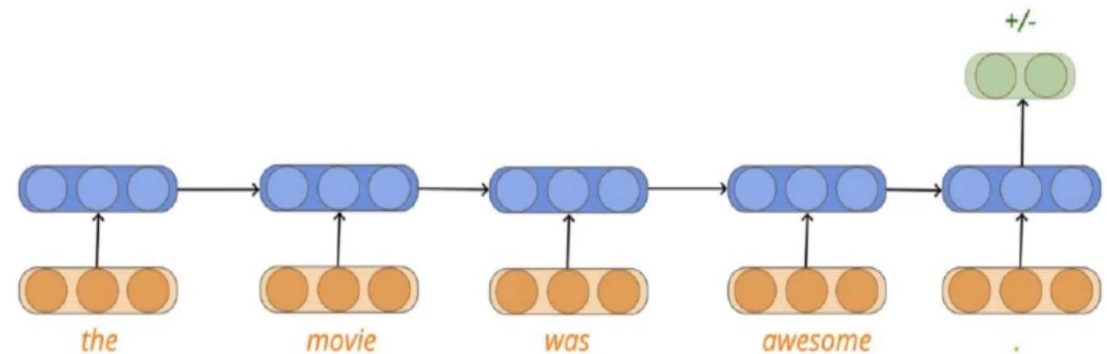
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]



Pre-process the data
(Sequence Classification Problem)

- ✓ lower case all words
- ✓ compute the total number of unique words across all sentences (say, $L \rightarrow 24$ in the above case)
- ✓ Assign a unique id to each word (between 1 to L)
- ✓ Represent each word using a L dimensional binary vector with only the bit corresponding to the word id set to 1

Pre-process the data
(Sequence Classification Problem)



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
The	first	half	was	very	boring	.	
Great	performance	by	all	the	lead	actors	.
The	background	music	was	awesome	.		
The	movie	was	a	waste	of	time	.

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
DT	AJ	NN	VB	JJ	JJ	PC	
AJ	NN	PP	PN	DT	JJ	NN	PC
DT	JJ	BB	VB	JJ	PC		
DT	NN	VB	DT	NN	PP	NN	PC

Pre-Process the data (Sequence Labelling Problem)

word	id
<sos>	1
<eos>	2
<pad>	3
the	4
first	5
half	6
...	
...	
time	24

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

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

[0, 1]

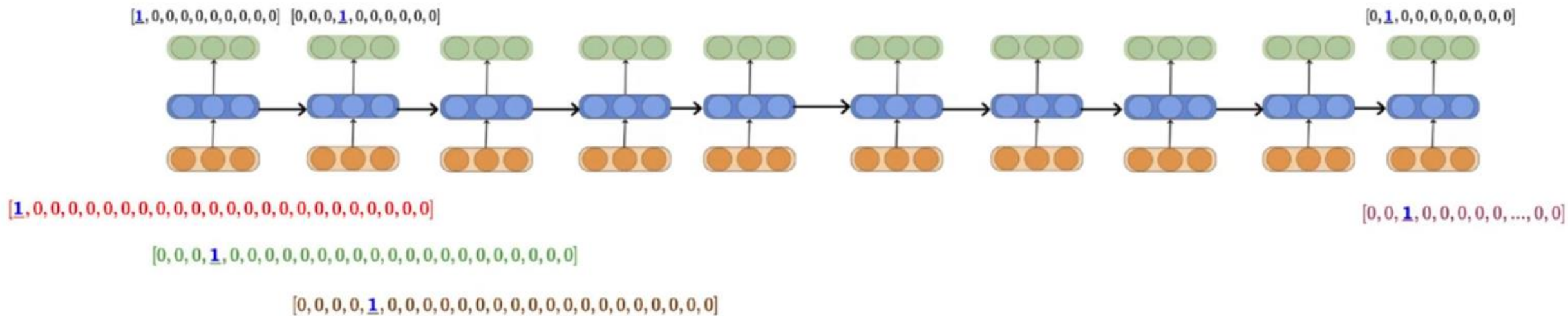
label	id
<sos>	1
<eos>	2
<pad>	3
DT	4
JJ	5
NN	6
...	
...	
PN	10

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

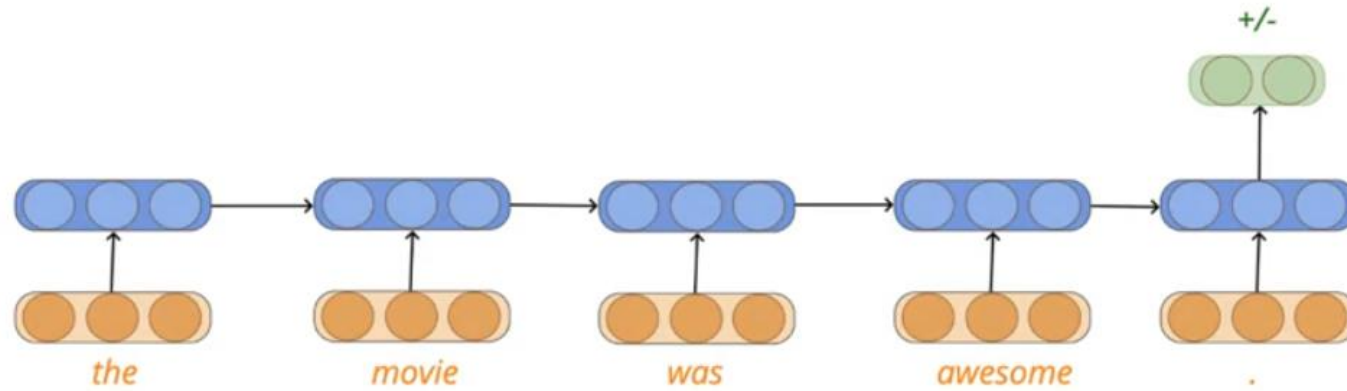
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0]

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

Pre-Process the data (Sequence Labelling Problem)



What does final input and output look like?



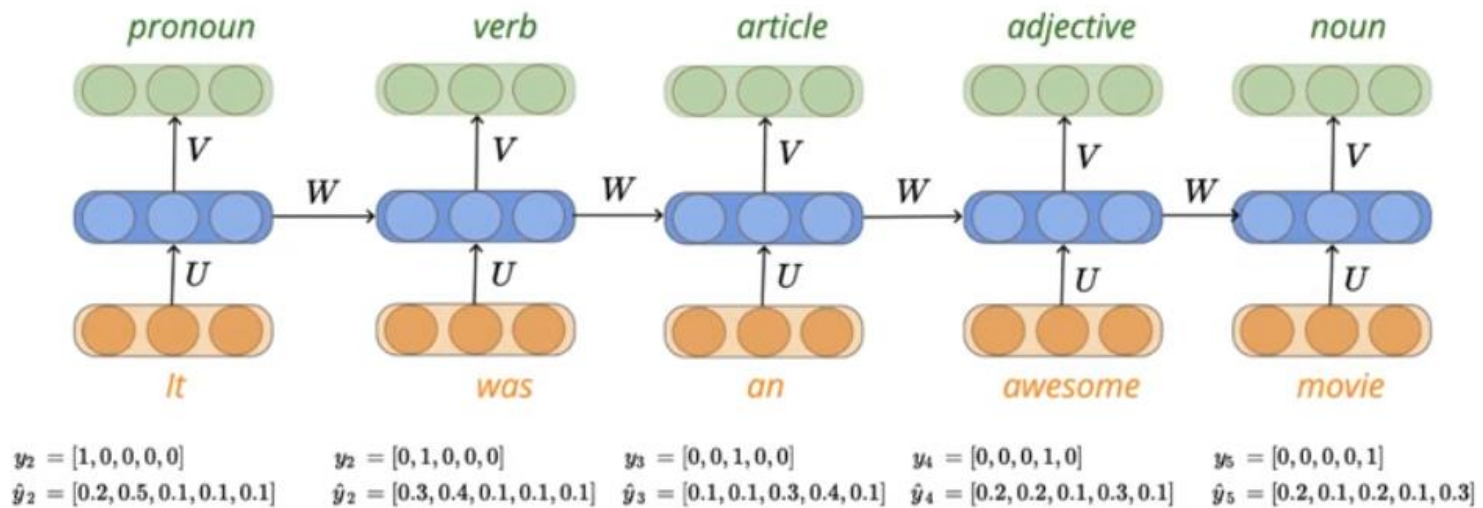
$$y = [1, 0]$$

$$\hat{y} = [0.7, 0.3]$$

$$\begin{aligned}\mathcal{L}(\theta) &= - \sum_{i=0}^1 y[i] \log \hat{y}[i] \\ &= - \log y_c \\ &= - \log 0.7\end{aligned}$$

$$\mathcal{L}(\theta) = - \frac{1}{m} \sum_{i=1}^m \log y_{ic}$$

Loss function for Sequence Classification Task



$$\mathcal{L}(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^T \log y_{ijc}$$

Loss function for Sequence Labelling Task