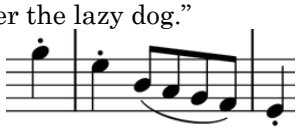
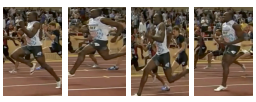
deeplearning.ai

Recurrent Neural Networks

Why sequence models?

1

Examples of sequence data

Speech recognition		→	"The quick brown fox jumped over the lazy dog."
Music generation	∅	→	
Sentiment classification	"There is nothing to like in this movie."	→	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	→	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	→	Do you want to sing with me?
Video activity recognition		→	Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	→	Yesterday, Harry Potter met Hermione Granger.

Andrew Ng

2

deeplearning.ai

Recurrent Neural Networks

Notation

3

Motivating example

x: Harry Potter and Hermione Granger invented a new spell.

Andrew Ng

4

Representing words

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ … $x^{<9>}$

Andrew Ng

5

Representing words

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ … $x^{<9>}$

And = 367
Invented = 4700
A = 1
New = 5976
Spell = 8376
Harry = 4075
Potter = 6830
Hermione = 4200
Gran… = 4000

Andrew Ng

6

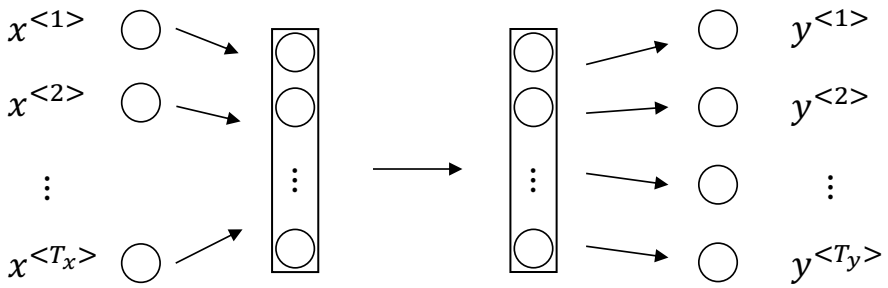
deeplearning.ai

Recurrent Neural Networks

Recurrent Neural Network Model

7

Why not a standard network?



Problems:

- Inputs, outputs can be different lengths in different examples.
- Doesn't share features learned across different positions of text.

Andrew Ng

8

Recurrent Neural Networks

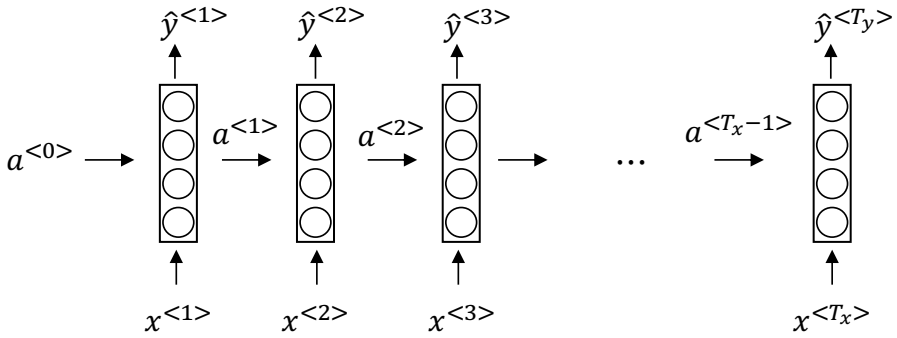
He said, “Teddy Roosevelt was a great President.”

He said, “Teddy bears are on sale!”

Andrew Ng

9

Forward Propagation



Andrew Ng

10

Simplified RNN notation

$$a^{<t>} = g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

$$\hat{y}^{<t>} = g(W_{ya}a^{<t>} + b_y)$$

Andrew Ng

11

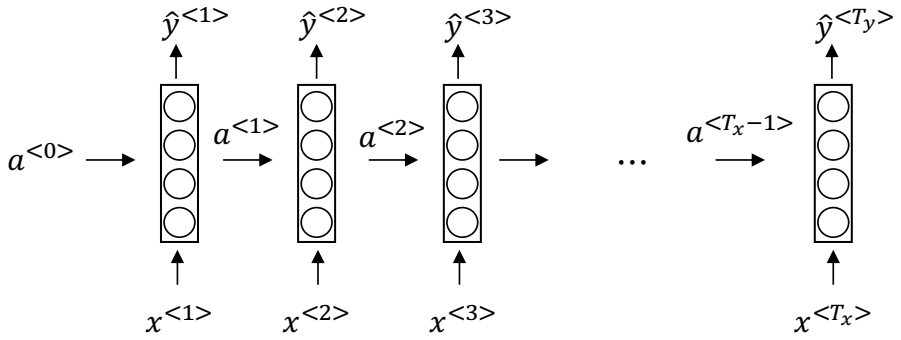
deeplearning.ai

Recurrent Neural Networks

Backpropagation through time

12

Forward propagation and backpropagation



Andrew Ng

13

Forward propagation and backpropagation

$\mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>}) =$

Backpropagation through time

Andrew Ng

14

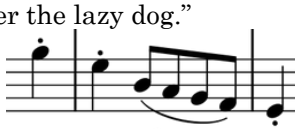
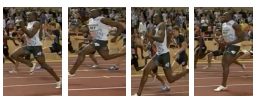
deeplearning.ai

Recurrent Neural Networks

Different types of RNNs

15

Examples of sequence data

Speech recognition		→	"The quick brown fox jumped over the lazy dog."
Music generation	∅	→	
Sentiment classification	"There is nothing to like in this movie."	→	★☆☆☆☆
DNA sequence analysis	AGCCCCTGTGAGGAACTAG	→	AGCCCCTGTGAGGAACTAG
Machine translation	Voulez-vous chanter avec moi?	→	Do you want to sing with me?
Video activity recognition		→	Running
Name entity recognition	Yesterday, Harry Potter met Hermione Granger.	→	Yesterday, Harry Potter met Hermione Granger.

Andrew Ng

16

Examples of RNN architectures

Andrew Ng

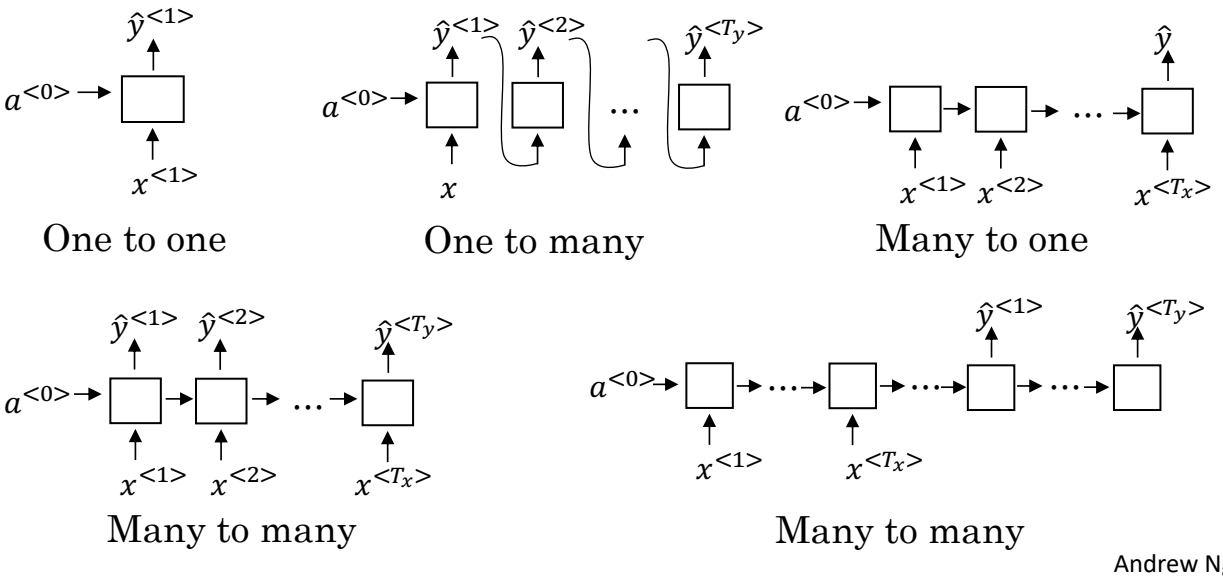
17

Examples of RNN architectures

Andrew Ng

18

Summary of RNN types



Andrew Ng

19

deeplearning.ai

Recurrent Neural Networks

Language model and sequence generation

20

What is language modelling?

Speech recognition

The apple and pair salad.

The apple and pear salad.

$P(\text{The apple and pair salad}) =$

$P(\text{The apple and pear salad}) =$

Andrew Ng

21

Language modelling with an RNN

Training set: large corpus of english text.

Cats average 15 hours of sleep a day.

The Egyptian Mau is a bread of cat. <EOS>

Andrew Ng

22

RNN model

Cats average 15 hours of sleep a day. <EOS>

$$\mathcal{L}(\hat{y}^{<t>}, y^{<t>}) = -\sum_i y_i^{<t>} \log \hat{y}_i^{<t>}$$

$$\mathcal{L} = \sum_t \mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

Andrew Ng

23

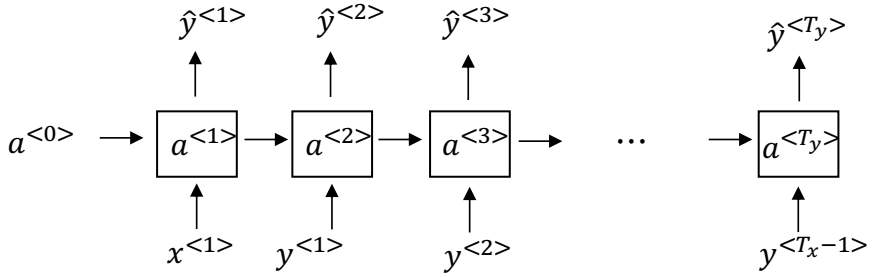
deeplearning.ai

Recurrent Neural Networks

Sampling novel sequences

24

Sampling a sequence from a trained RNN

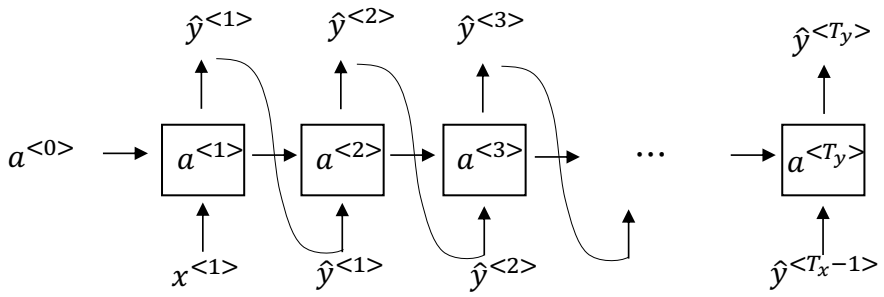


Andrew Ng

25

Character-level language model

Vocabulary = [a, aaron, …, zulu, <UNK>]



Andrew Ng

26

Sequence generation

News

President enrique peña nieto, announced sench's sulk former coming football langston paring.

"I was not at all surprised," said hich langston.

"Concussion epidemic", to be examined.

The gray football the told some and this has on the uefa icon, should money as.

Shakespeare

The mortal moon hath her eclipse in love.

And subject of this thou art another this fold.

When besser be my love to me see sabl's.

For whose are ruse of mine eyes heaves.

Andrew Ng

27

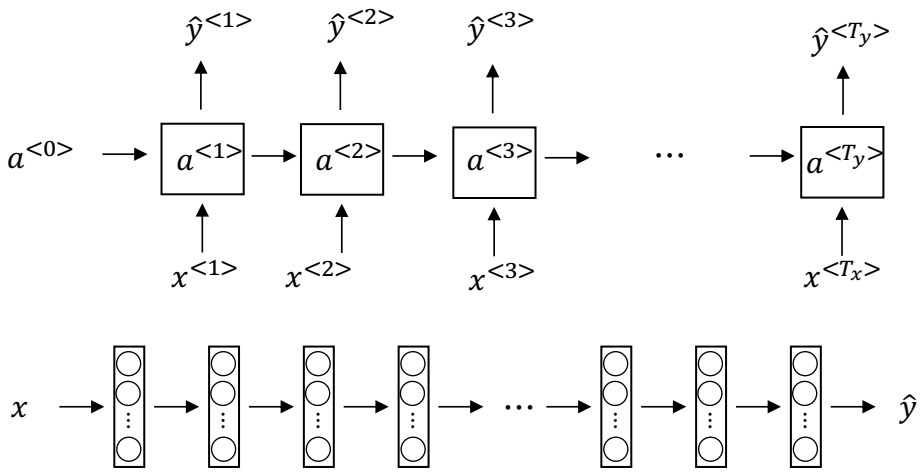
deeplearning.ai

Recurrent Neural Networks

Vanishing gradients with RNNs

28

Vanishing gradients with RNNs



Exploding gradients.

Andrew Ng

29

deeplearning.ai

Recurrent Neural Networks

Gated Recurrent Unit (GRU)

30

RNN unit

$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

Andrew Ng

31

GRU (simplified)

The cat, which already ate …, was full.

[Cho et al., 2014. On the properties of neural machine translation: Encoder-decoder approaches]
[Chung et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling]

Andrew Ng

32

Full GRU

$$\tilde{c}^{<t>} = \tanh(W_c[\, c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\, c^{<t-1>}, x^{<t>}] + b_u)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) + c^{<t-1>}$$

The cat, which ate already, was full.

Andrew Ng

33

deeplearning.ai

Recurrent Neural Networks

LSTM (long short term memory) unit

34

GRU and LSTM

GRU

$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$

$\Gamma_u = \sigma(W_u[\, c^{<t-1>}, x^{<t>}] + b_u)$

$\Gamma_r = \sigma(W_r[\, c^{<t-1>}, x^{<t>}] + b_r)$

$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$

$a^{<t>} = c^{<t>}$

LSTM

[Hochreiter & Schmidhuber 1997. Long short-term memory]

Andrew Ng

35

LSTM units

GRU

$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$

$\Gamma_u = \sigma(W_u[\, c^{<t-1>}, x^{<t>}] + b_u)$

$\Gamma_r = \sigma(W_r[\, c^{<t-1>}, x^{<t>}] + b_r)$

$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$

$a^{<t>} = c^{<t>}$

LSTM

$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$

$\Gamma_u = \sigma(W_u[\, a^{<t-1>}, x^{<t>}] + b_u)$

$\Gamma_f = \sigma(W_f[\, a^{<t-1>}, x^{<t>}] + b_f)$

$\Gamma_o = \sigma(W_o[\, a^{<t-1>}, x^{<t>}] + b_o)$

$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$

$a^{<t>} = \Gamma_o * c^{<t>}$

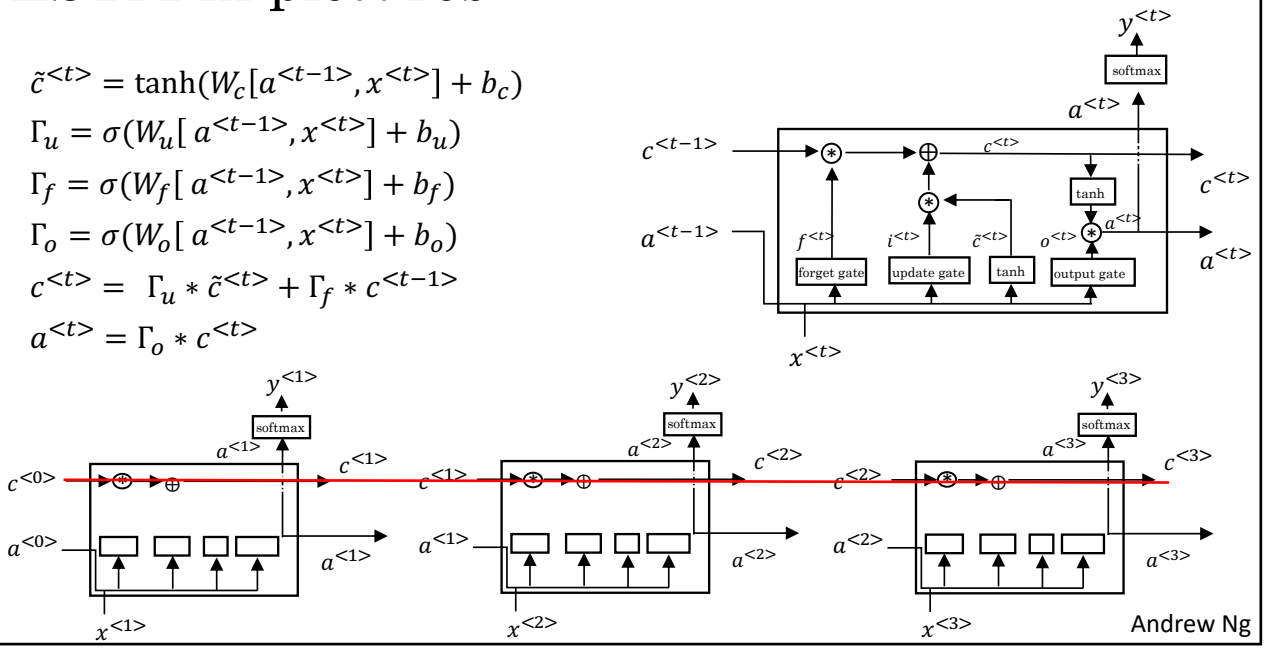
[Hochreiter & Schmidhuber 1997. Long short-term memory]

Andrew Ng

36

LSTM in pictures

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$
$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$
$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$
$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$
$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$
$$a^{<t>} = \Gamma_o * c^{<t>}$$



37

deeplearning.ai

Recurrent Neural Networks

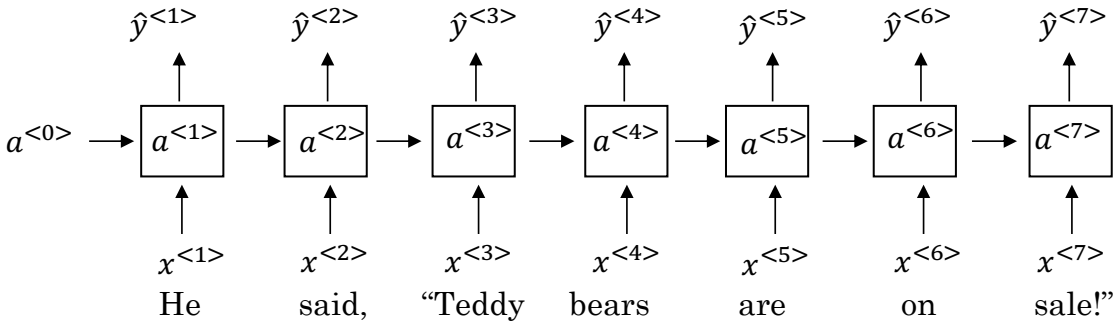
Bidirectional RNN

38

Getting information from the future

He said, “Teddy bears are on sale!”

He said, “Teddy Roosevelt was a great President!”



Andrew Ng

39

Bidirectional RNN (BRNN)

Andrew Ng

40

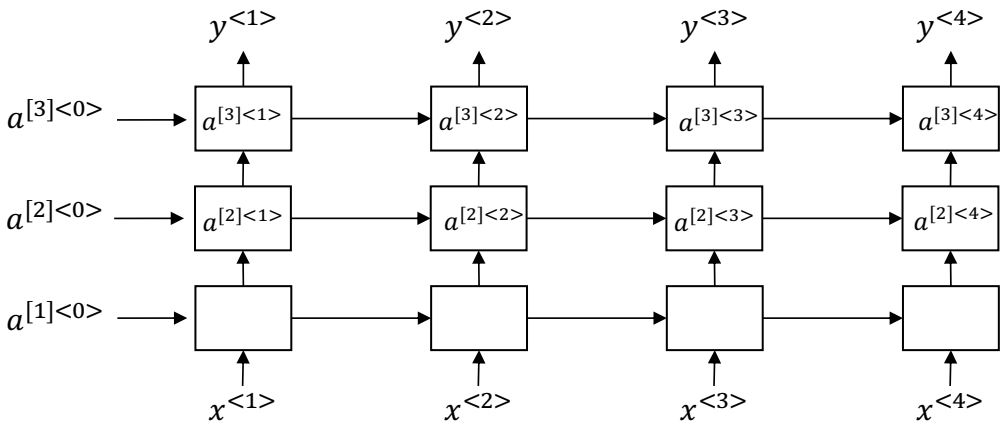
deeplearning.ai

Recurrent Neural Networks

Deep RNNs

41

Deep RNN example



Andrew Ng

42