

# CHƯƠNG 5

## MẠNG NEURAL HỒI QUY (P1)

Khoa Khoa học và Kỹ thuật thông tin  
Bộ môn Khoa học dữ liệu

# NỘI DUNG

1. Sequence models.
2. Recurrent Neural model.
3. Các loại RNN.
4. Language models.

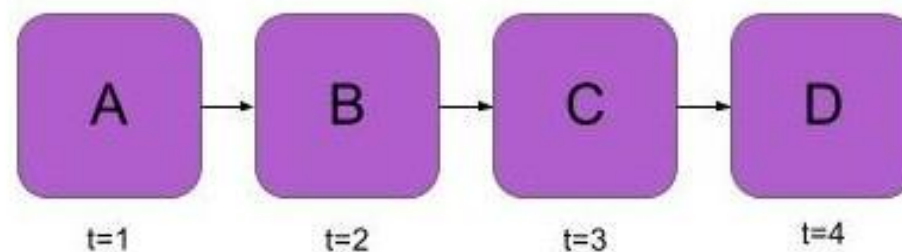
# SEQUENCE MODEL

# Sequence model

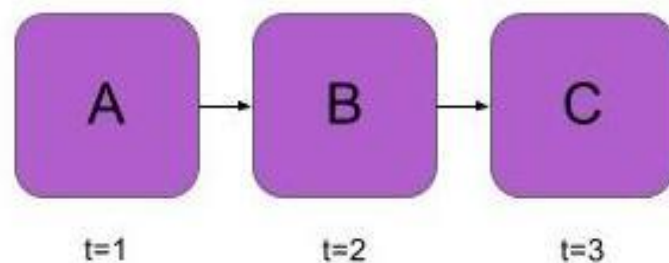
- Sequence model là một tác vụ dự đoán một đối tượng nào đó sẽ xuất hiện tiếp theo trong một chuỗi các đối tượng.
- + VD:
  - “con” → từ xuất hiện tiếp theo có thể là: “bò”, “chó”, ...
  - “đi” → từ xuất hiện tiếp theo có thể là: “đến trường”, “học”, “màn”.
- Trong sequence model, input của lớp hiện tại sẽ phụ thuộc vào output của lớp trước đó. Độ dài của dữ liệu input vào sẽ không cố định.

# Minh họa

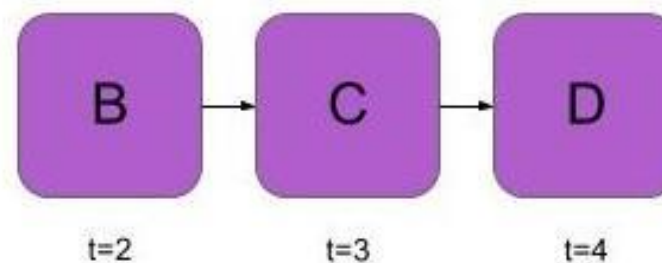
Sequence



Input Sequence



Target Sequence



# Các dữ liệu dạng sequence - Text

DNA sequence analysis    AGCCCCTGTGAGGAACTAG →    AG**CCCCTGTGAGGAACTAG**

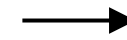
Machine translation    Voulez-vous chanter avec moi? →    Do you want to sing with me?

Name entity recognition    Yesterday, Harry Potter met Hermione Granger. →    Yesterday, **Harry Potter** met **Hermione Granger**.

Sentiment classification    “There is nothing to like in this movie.” →    ★☆☆☆☆

# Các dữ liệu dạng sequence – multi media

Speech  
recognition



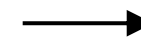
“The quick brown fox jumped  
over the lazy dog.”

Music generation

∅



Video activity recognition



Running

# SEQUENCE MODEL – BÀI TOÁN

## NAME ENTITIES RECOGNITION

X Harry Potter and Hermione Granger invented a new spell.

$X^{<1>}$   $X^{<2>}$   $X^{<3>}$  .....  $X^{<9>}$

y 1 1 0 1 1 0 0 0 0  
 $y^{<1>}$   $y^{<2>}$   $y^{<3>}$  .....  $y^{<9>}$

$T_x = 9$ : Độ dài X sẽ là 9



# Một số ký hiệu mới trong Sequence model

| Ký hiệu      | Ý nghĩa   |
|--------------|---|
| $X^{<i>}$    | Phần tử vị trí thứ $i$ trong sequence   |
| $X^{(i)<t>}$ | Phần tử thứ $t$ trong sequence thứ $i$ ( <i>sequence thứ <math>i</math> ở đây là điểm dữ liệu thứ <math>i</math> trong tập huấn luyện</i> ) |
| $T_X$        | Độ dài của input sequence $X$   |
| $T_X^{(i)}$  | Độ dài của input sequence $X$ ở điểm dữ liệu thứ $i$ trong tập huấn luyện   |

# Biểu diễn từ trong Sequence model

X Harry Potter and Hermione Granger invented a new spell.

$X^{<1>}$

$X^{<2>}$

$X^{<3>}$

.....

$X^{<9>}$

- **Bước 1:** Xây dựng bộ từ vựng (Vocabulary) cho toàn bộ dữ liệu.
- **Bước 2:** Biểu diễn mỗi  $X^{<t>}$  dưới dạng một vector có số chiều bằng độ dài của bộ từ vựng.
  - Các dạng biểu diễn thường dùng: one-hot vector, term frequency, tf-idf, ...

# Biểu diễn từ trong Sequence model

X Harry Potter and Hermione Granger invented a new spell.

$X^{<1>}$     $X^{<2>}$     $X^{<3>}$    .....    $X^{<9>}$

Vocabulary: [Harry, Potter, and, Hermione, Granger, invented, a, new, spell]

Vocab\_size = 9

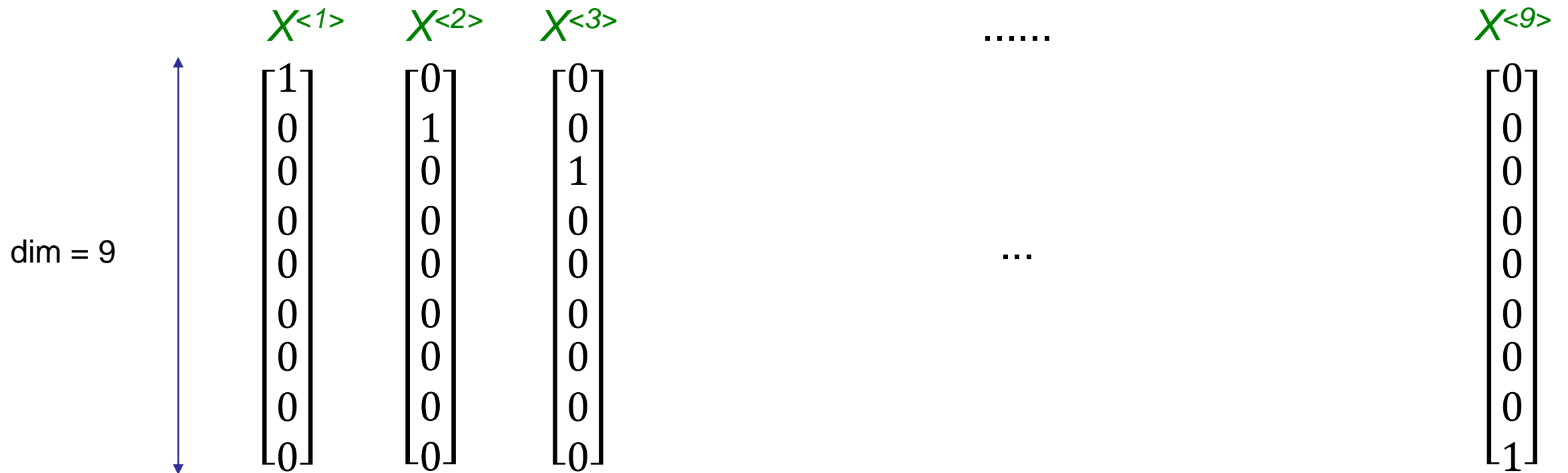
$$X^{<1>} = [1, 0, 0, 0, 0, 0, 0, 0, 0]$$

$$X^{<2>} = [0, 1, 0, 0, 0, 0, 0, 0, 0]$$

...

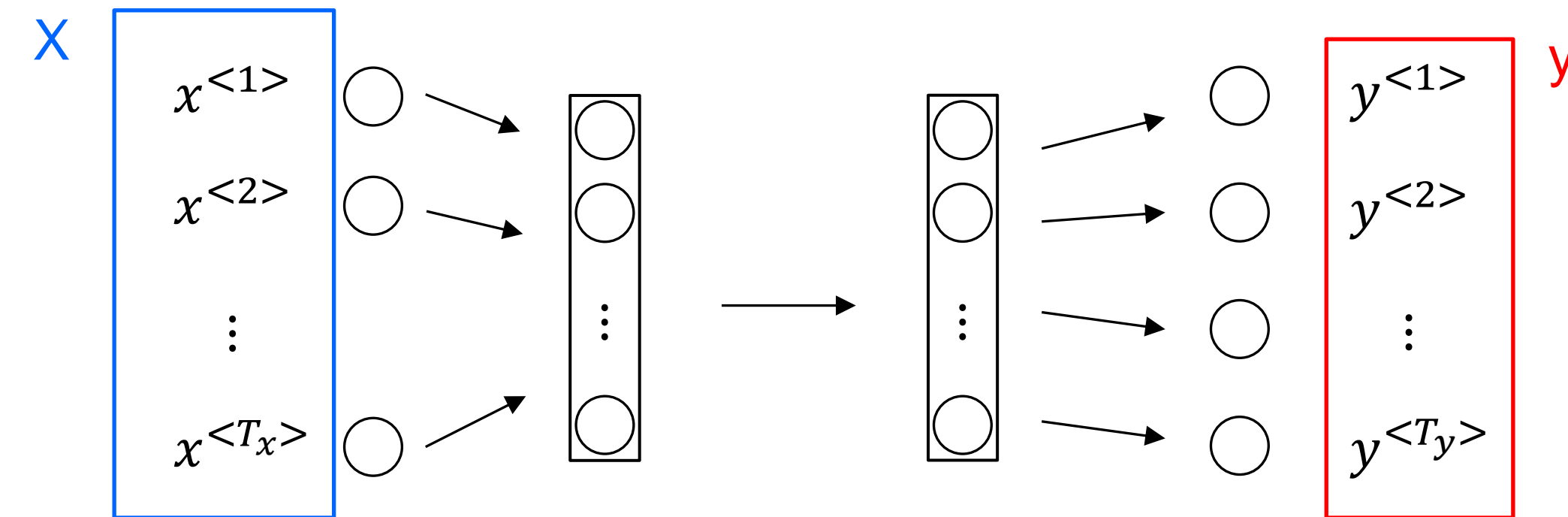
# Biểu diễn từ trong Sequence model

X Harry Potter and Hermione Granger invented a new spell.



# Recurrent Neural Network

# Mạng neural truyền thống



X Harry Potter and Hermione Granger invented a new spell.

y 1 1 0 1 1 0 0 0 0

# Vấn đề xảy ra khi dùng mạng neural truyền thống cho sequence data

- Input và output có thể khác nhau về độ dài ở các điểm dữ liệu khác nhau trong tập dữ liệu.
  - + Khác so với CNN: các ảnh trong tập dữ liệu có cùng kích thước (VD: 64x64x3).
- Không thể share feature với các vị trí khác trong sequence model.

VD:

*Harry Potter* and *Hermione Granger* invented a new spell.



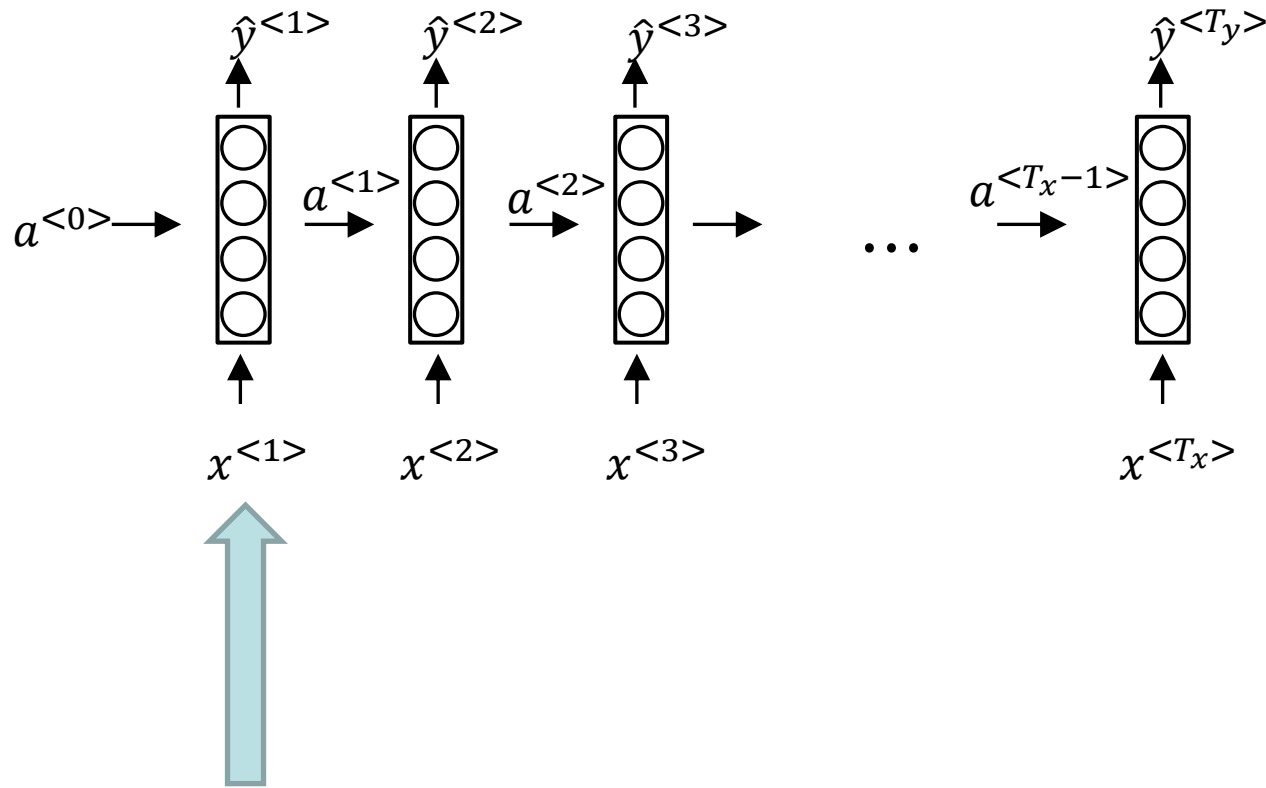
Hai vị trí trên trong các câu khác có khả năng là các tên riêng

# Recurrent Neural network

## Truyền qua từng layer

— Lớp thứ 1:

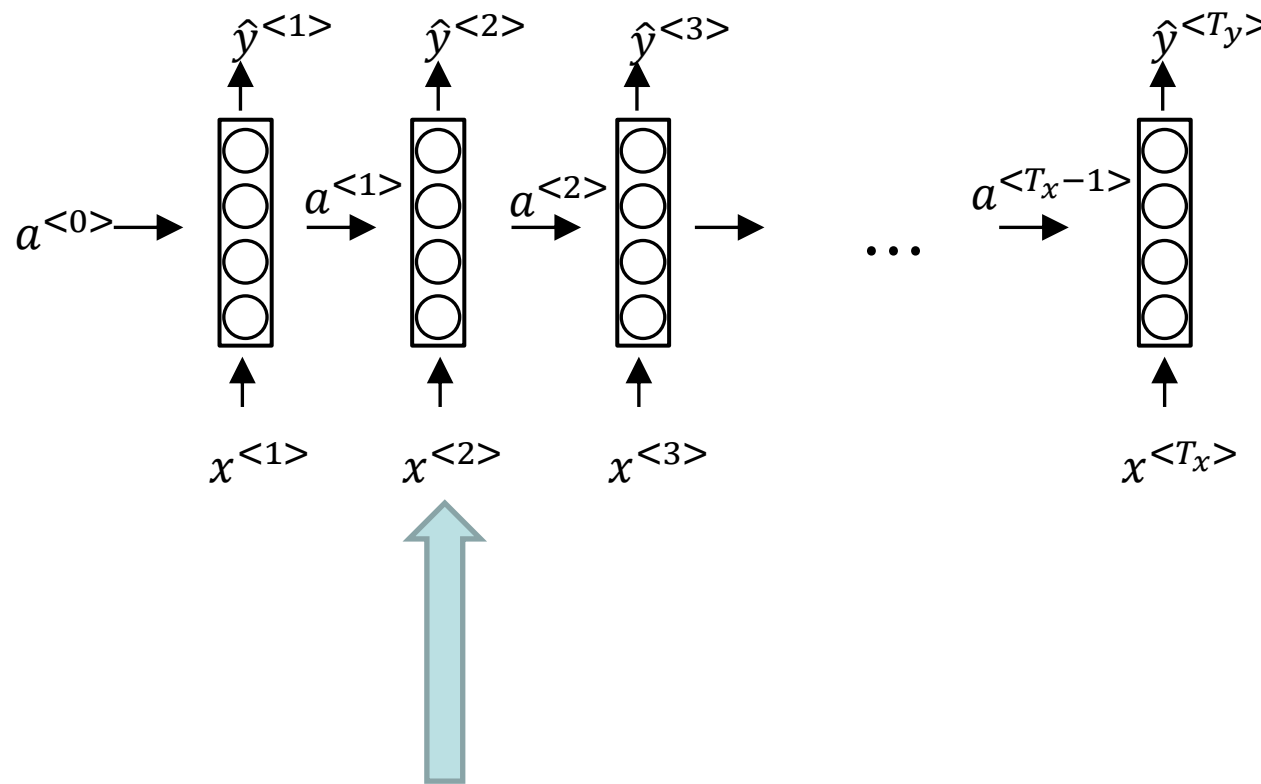
- + *Input*: sequence  $x^{<1>}$  + đầu vào lớp 0 –  $a^{<0>}$ : *vector of zero*.
- + *Output*: giá trị dự đoán  $y^{<1>}$ .





# Recurrent Neural network

## Truyền qua từng layer



### — Lớp thứ 1:

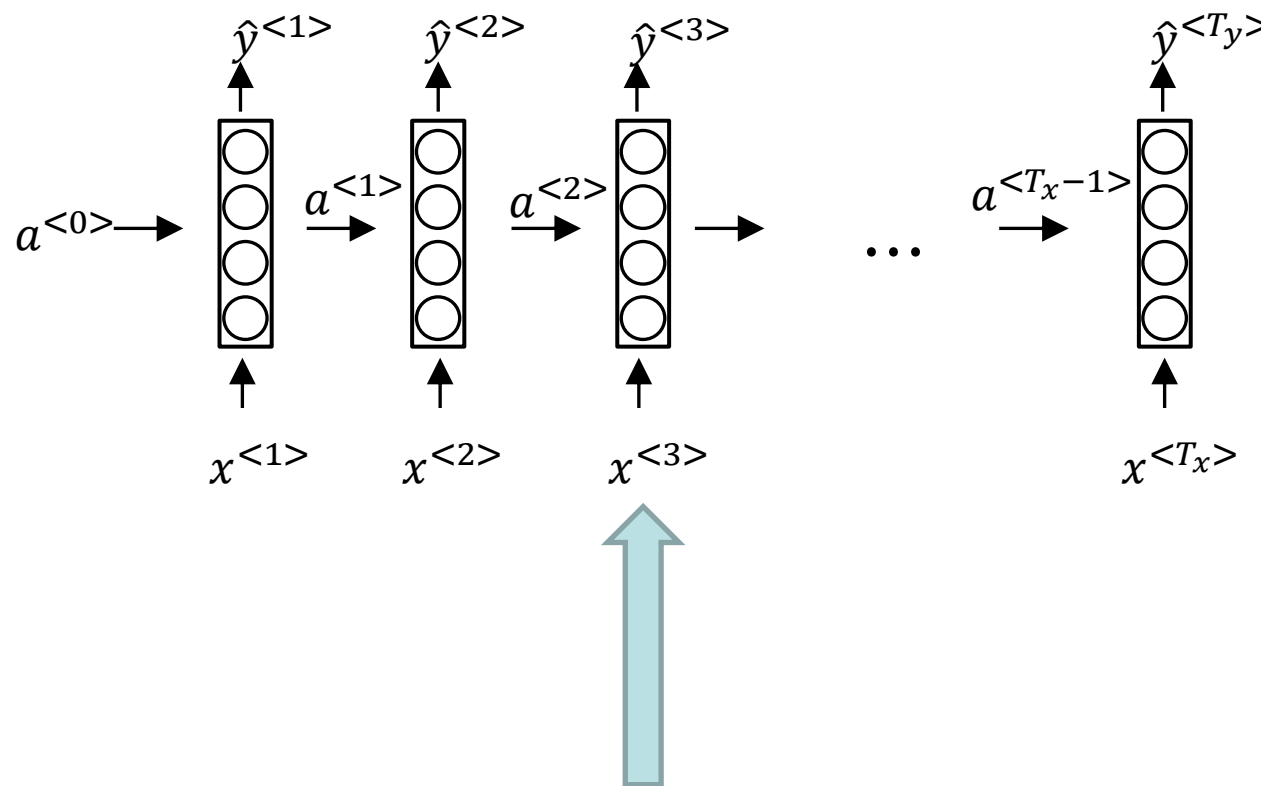
- + *Input*: sequence  $x^{<1>}$  + đầu vào lớp 0 –  $a^{<0>}$ : *vector of zero*.
- + *Output*: giá trị dự đoán  $y^{<1>}$ .

### — Lớp thứ 2:

- + *Input*: sequence  $x^{<2>}$  + giá trị deactivation của lớp 1 –  $a^{<1>}$ .
- + *Output*: giá trị dự đoán  $y^{<2>}$ .

# Recurrent Neural network

## Truyền qua từng layer



### – Lớp thứ 1:

- + *Input*: sequence  $x^{<1>}$  + đầu vào lớp 0 –  $a^{<0>}$ : *vector of zero*.
- + *Output*: giá trị dự đoán  $y^{<1>}$ .

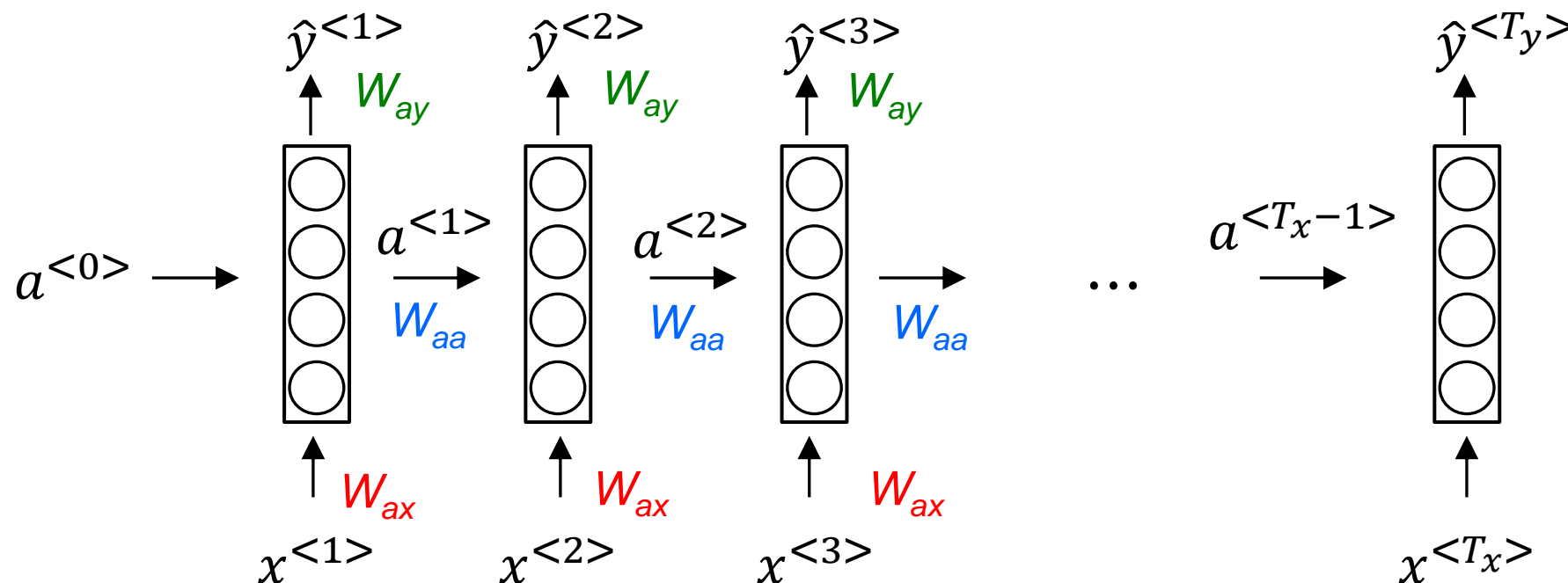
### – Lớp thứ 2:

- + *Input*: sequence  $x^{<2>}$  + giá trị deactivation của lớp 1 –  $a^{<1>}$ .
- + *Output*: giá trị dự đoán  $y^{<2>}$ .

### – Lớp thứ 3:

- + *Input*: sequence  $x^{<3>}$  + giá trị deactivation của lớp 2 –  $a^{<2>}$ .
- + *Output*: giá trị dự đoán  $y^{<3>}$ .

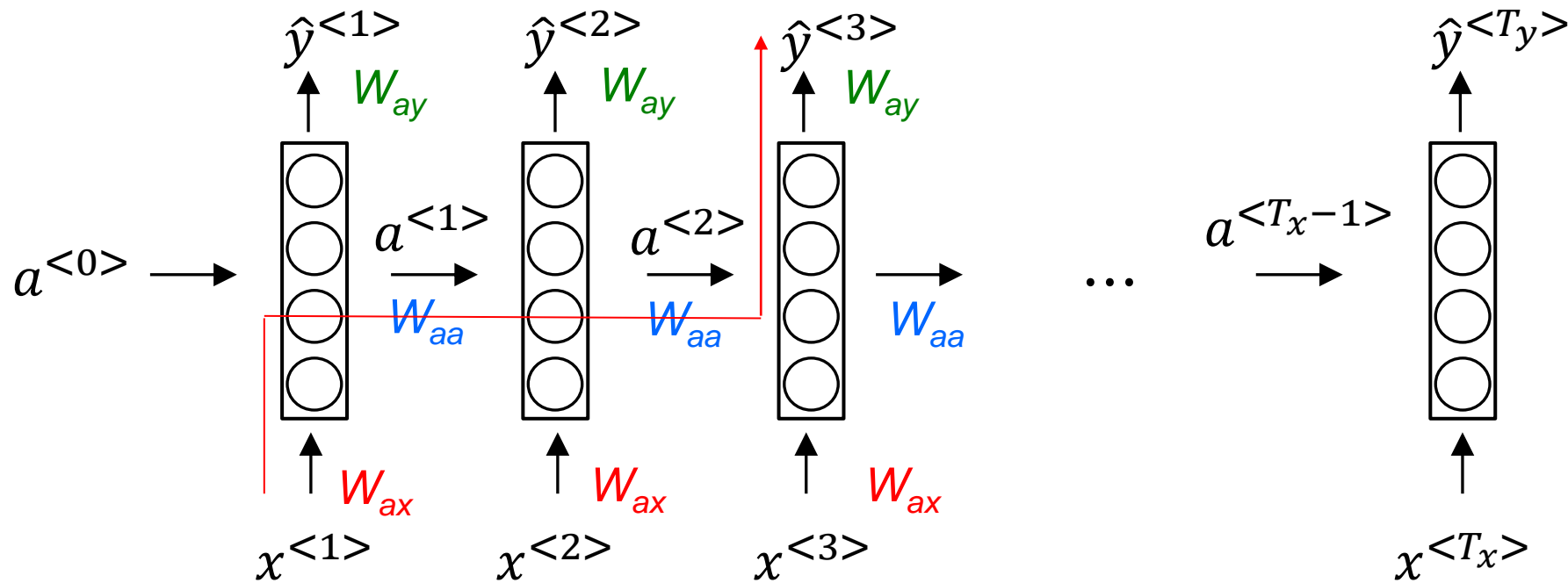
# Tham số trong mô hình RNN



Mạng RNN quét dữ liệu sequence từ trái sang phải. Có 3 loại tham số chính:

- *Governing parameter*:  $W_{ax}$ .
- *Deactivation parameter (horizontal connection)*:  $W_{aa}$ .
- *Output prediction*:  $W_{ay}$ .

# Truyền tham số trong mô hình RNN



Khi dự đoán giá trị  $y^{<3>}$ , mô hình RNN nhận các input gồm:  $x^{<3>}, x^{<1>}, x^{<2>}$ .

# Ví dụ: bài toán nhận dạng tên người

— Câu 1:

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

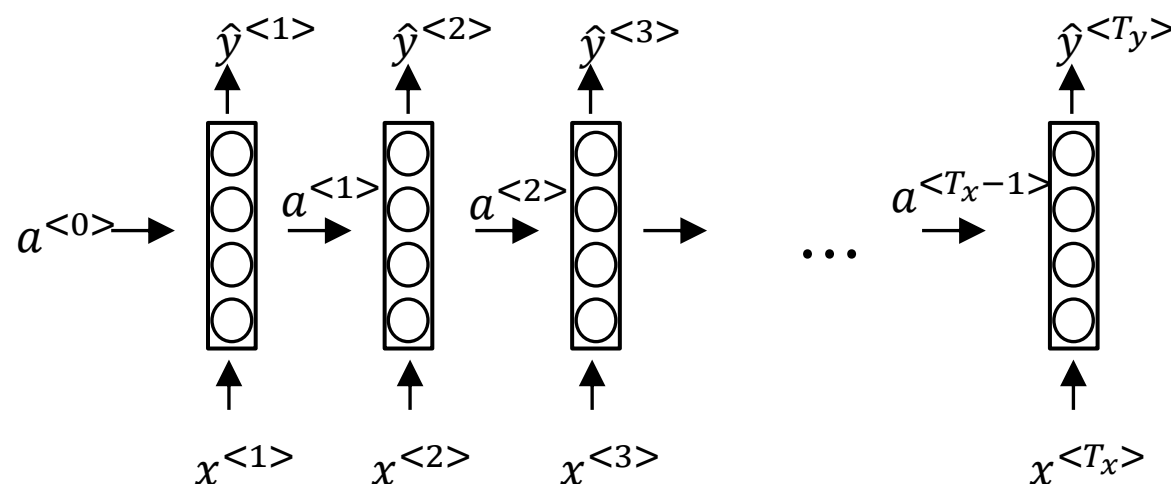
→ Teddy là tên người (Teddy Roosevelt – Tổng thống thứ 26 của Hoa Kỳ)

— Câu 2:

He said, “Teddy bears are on sale!”

→ Teddy lúc này không phải là tên người.

# Forward Propagation



Tổng quát:

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

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

– Lớp 0:

$$a^{<0>} = \vec{0}$$

– Lớp 1:

$$a^{<1>} = g_1(W_{aa} * a^{<0>} + W_{ax} * X^{<1>} + b_a)$$

$$y^{<1>} = g_2(W_{ya} * a^{<1>} + b_y)$$

– Lớp 2:

$$a^{<2>} = g_1(W_{aa} * a^{<1>} + W_{ax} * X^{<2>} + b_a)$$

$$y^{<2>} = g_2(W_{ya} * a^{<2>} + b_y)$$

# Forward Propagation

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

(100, 100)

(100, 10000)

$$[W_{aa}; W_{ax}] = W_a$$

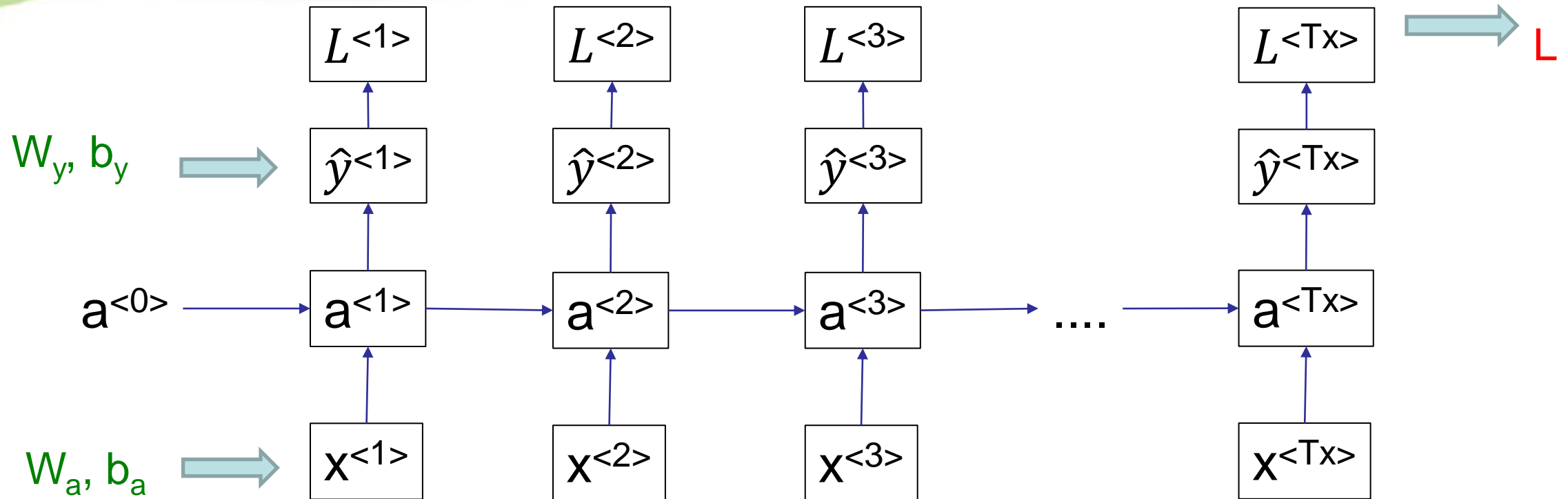
$$a^{<t>} = g(W_a * [a^{<t>}, x^{<t>}] + b_a)$$

Tham số:  $W_a, b_a$

$$y^{<t>} = g(W_y * a^{<t>} + b_y)$$

Tham số:  $W_y, b_y$

# Backward propagation cho RNN

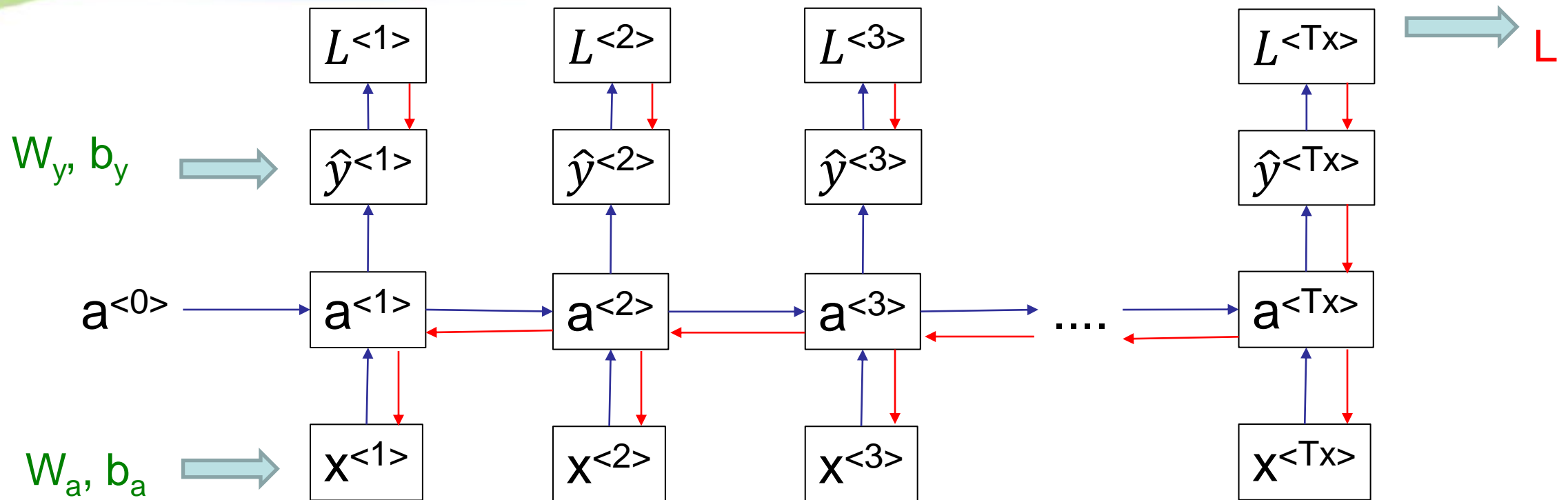


Hàm chi phí:  $L(\hat{y}^{<t>}, y^{<t>}) = -y^{<t>} * \log(\hat{y}^{<t>}) + (1 - y^{<t>}) * \log(1 - \hat{y}^{<t>})$

$$= \sum_{t=1}^{Tx} L^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

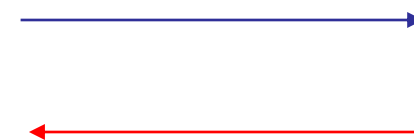


# Backward propagation cho RNN

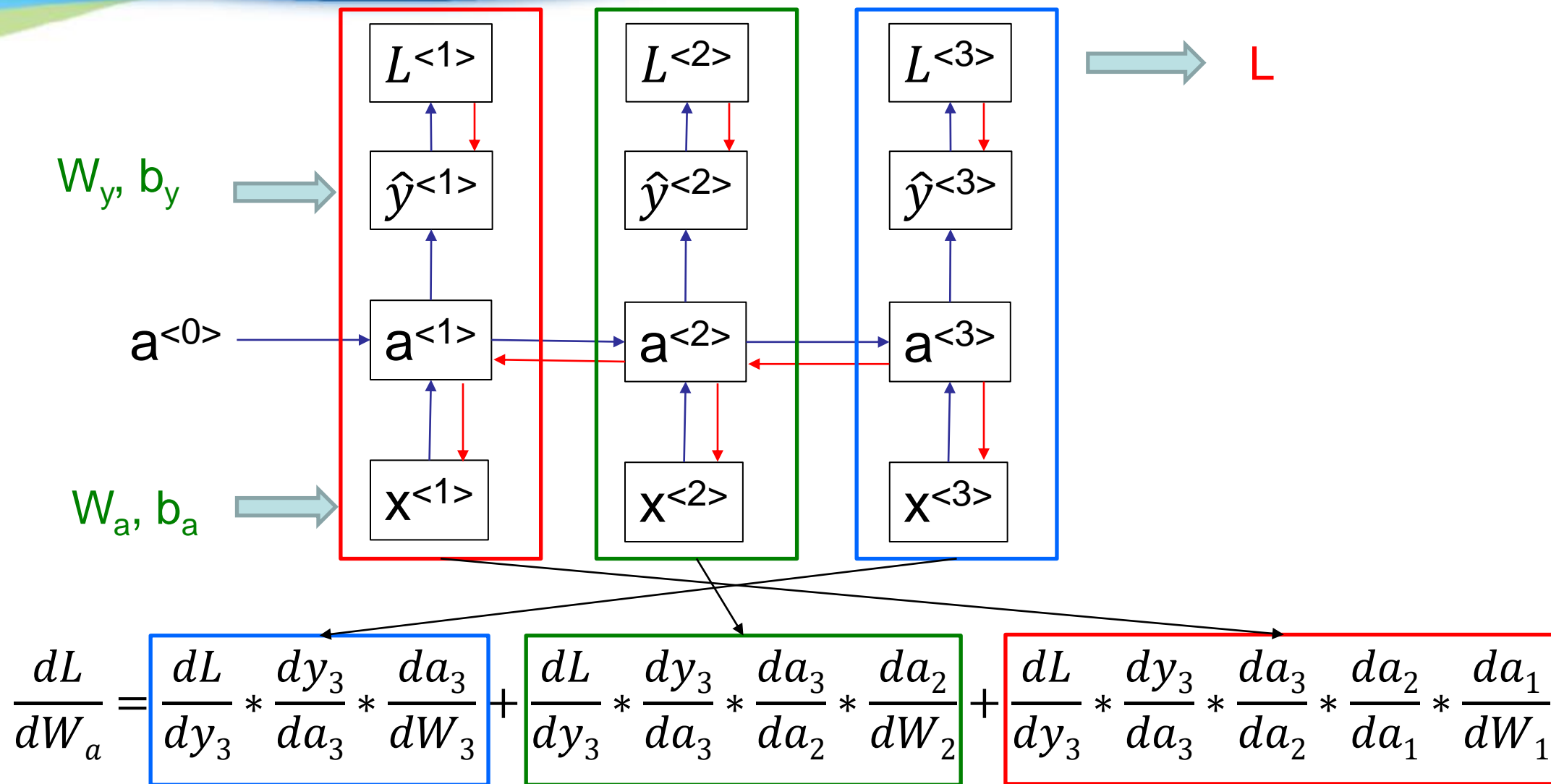


**Forward:** đi từ trái sang phải (t tăng dần).

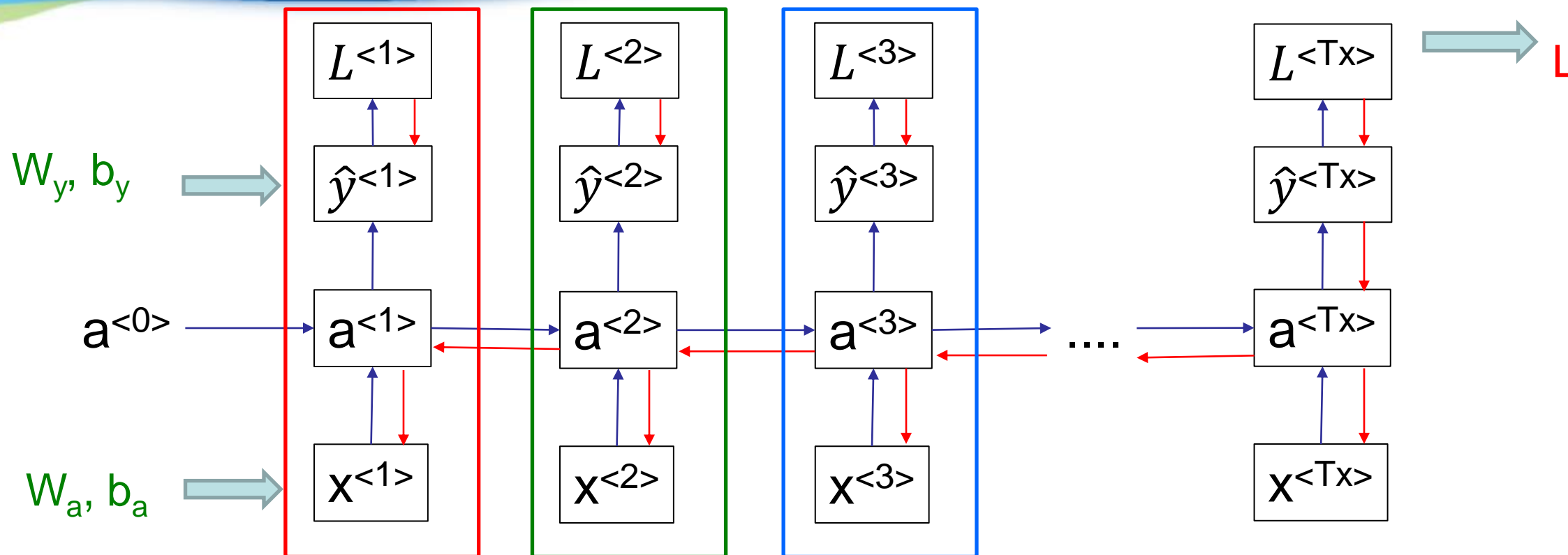
**Backward:** đi từ phải sang trái (t giảm dần).



# Backward propagation through time



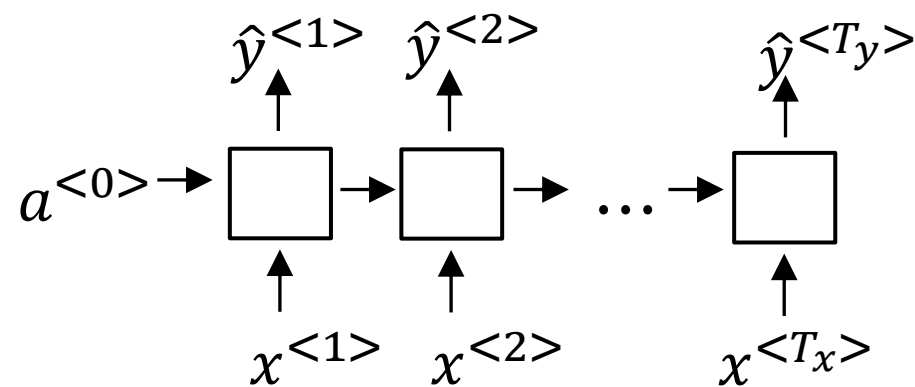
# Backward propagation through time



Tổng quát: 
$$\frac{dL}{dW} = \sum_{t=1}^T \left. \frac{dL^{(t)}}{dW} \right|_{((t))}$$

# Các loại RNN

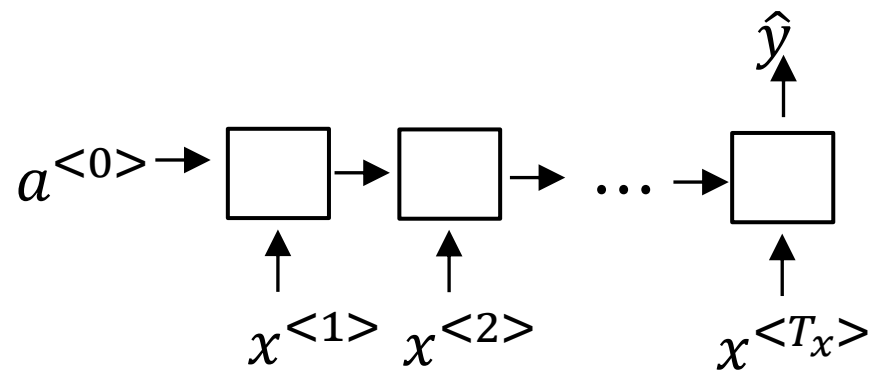
# Many-to-many



Many to many

- Đầu vào: 1 sequence (input sequence).
- Đầu ra: 1 sequence với độ dài bằng độ dài đầu vào.

# Many-to-one



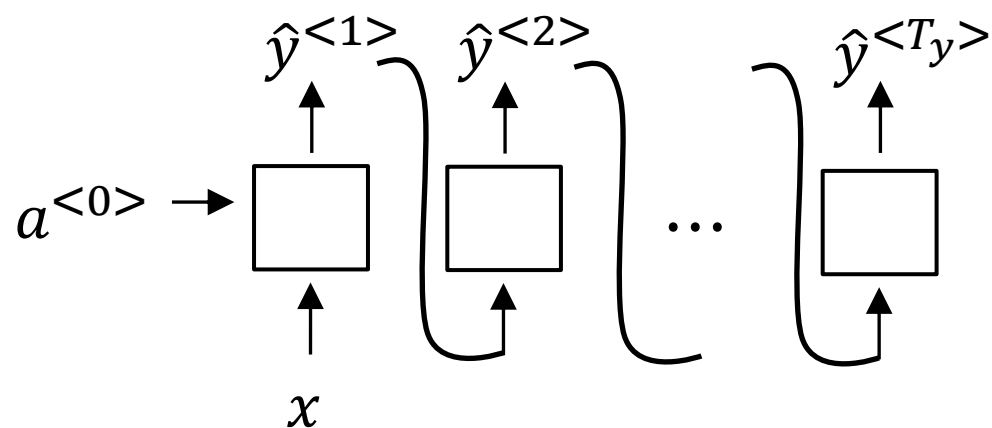
Many to one

- Đầu vào: 1 sequence.
- Đầu ra: một giá trị dự đoán  $\hat{y}$

Ứng dụng: bài toán **classification**:

- + Đầu vào (X): 1 chuỗi text.
- + Đầu ra (y): nhãn (giá trị) dự đoán.

# One-to-many



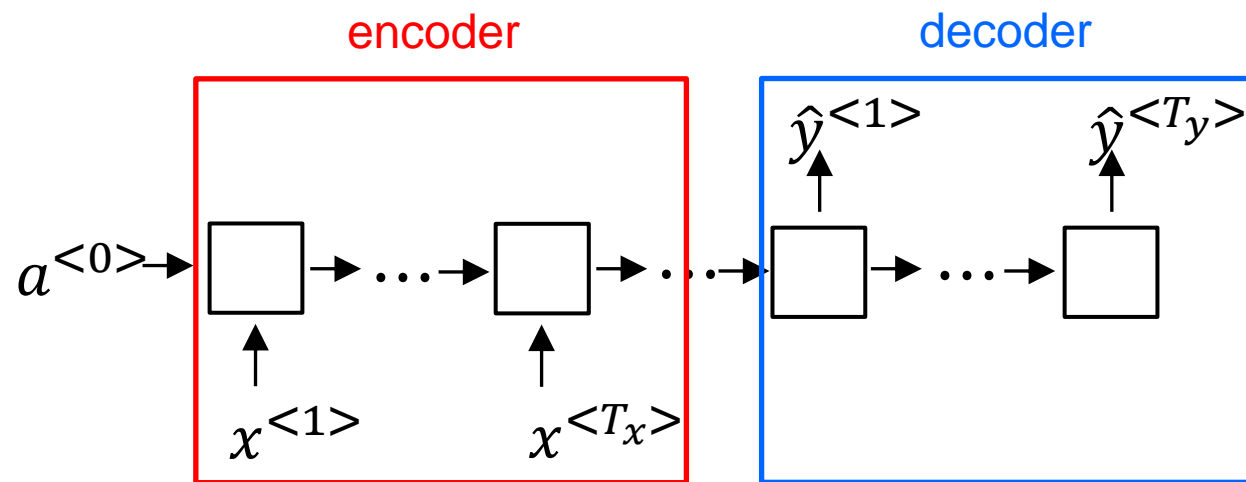
One to many

- Đầu vào: 1 giá trị (có thể là 1 null input).
- Đầu ra: 1 sequence.

VD: Bài toán sinh ra 1 đoạn nhạc:

- + Đầu vào (X): thể loại nhạc.
- + Đầu ra (y): Một sequence gồm các nốt nhạc.

# Many-to-many



Many to many

- Đầu vào: 1 sequence.
- Đầu ra: 1 sequence khác với đầu vào.

VD: bài toán dịch máy

- + Đầu vào (X): một câu tiếng Anh.
- + Đầu ra (y): một câu tiếng Pháp.



# Language models

# Định nghĩa Language model

- Language model (tạm dịch: mô hình ngôn ngữ) là một phân phối xác suất (probability distribution) giữa các chuỗi ký tự (bao gồm từ - words) trong một đoạn văn bản.
- Định nghĩa của D. Jurafsky: *Models that assign probabilities to sequences of words are called language models or LMs.*

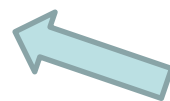
# Ví dụ:

- Sentence 1: The apple and pair salad.
- Sentence 2: The apple and pear salad.

$P(\text{The apple and pair salad}) = 3.2 \times 10^{-3}$ .

$P(\text{The apple and pear salad}) = 5.7 \times 10^{-3}$ .

$P(\text{next\_sentence}) = ?$  – xác suất xuất hiện câu tiếp theo là bao nhiêu ?



Language model

# Token in language model

## Sentences

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

$y^{<1>}$   $y^{<2>}$   $y^{<3>}$   $y^{<4>}$   $y^{<5>}$   $y^{<6>}$   $y^{<7>}$   $y^{<8>}$   $y^{<9>}$

token

<EOS>: là  
token định  
nghĩa kết thúc  
một câu  
(End of  
sentence)

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

$y^{<1>}$   $y^{<2>}$   $y^{<3>}$   $y^{<4>}$   $y^{<5>}$   $y^{<6>}$   $y^{<7>}$   $y^{<8>}$   $y^{<9>}$

token

<UNK>: token đặc biệt, được dùng khi token đó **không có trong bộ từ vựng** (Vocabulary)

## Vocabulary:

Cats  
average  
15  
hours  
of  
sleep  
a  
day  
The  
Egyptian  
is  
bread

# Huấn luyện RNN model cho language model

— Mục tiêu huấn luyện: dự đoán xác suất xuất hiện từ tiếp theo trong một câu dựa vào từ trước đó.

VD:  $(y^{<1>}, y^{<2>}, y^{<3>})$  là một câu dự đoán (gồm 3 token).

Tính xác suất: cho từng token trên theo language model.

$P(y^{<1>})$ : xác suất xuất hiện token  $y^{<1>}$ .

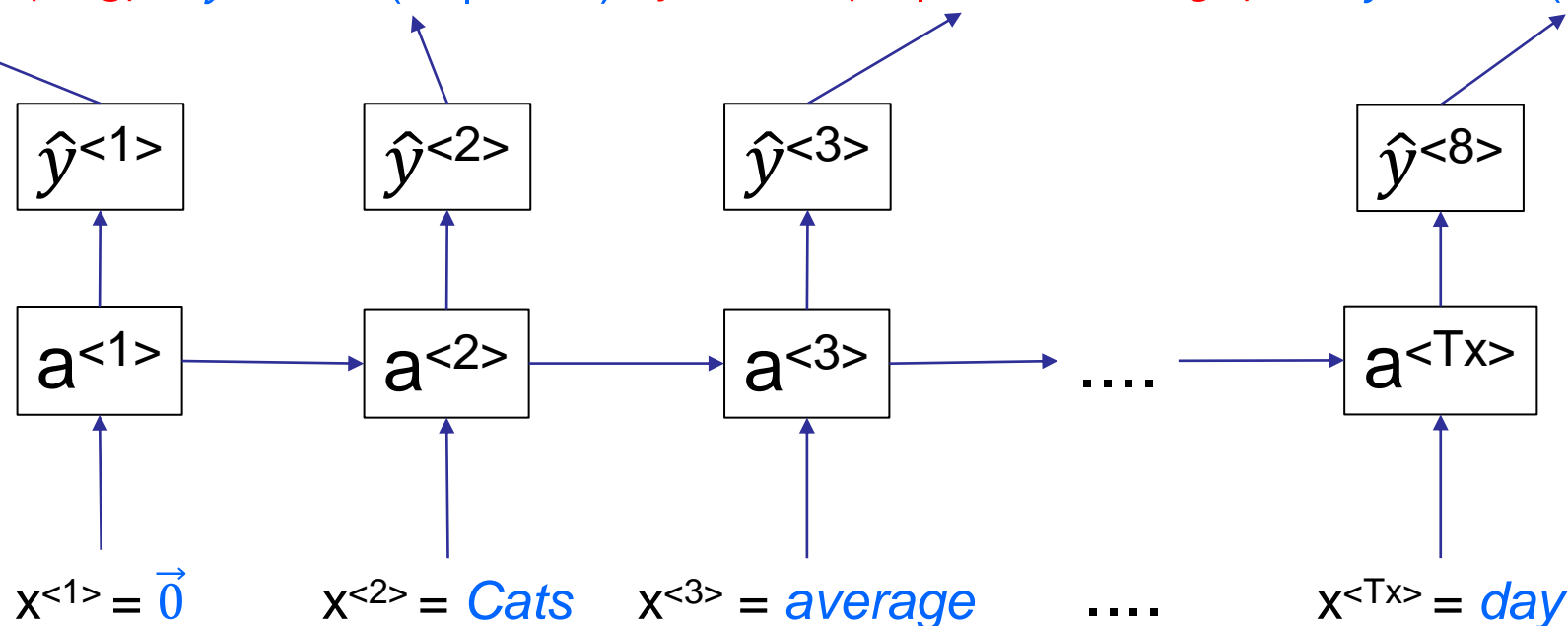
$P(y^{<2>} | y^{<1>})$ : xác suất xuất hiện token  $y^{<2>}$  khi biết  $y^{<1>}$ .

$P(y^{<3>} | y^{<2>}, y^{<1>})$ : xác suất xuất hiện  $y^{<3>}$  khi biết  $y^{<1>}$  và  $y^{<2>}$ .

Đánh giá độ chính xác: dựa vào hàm Cost function của mạng RNN.

# Ví dụ

$$\hat{y}^{<1>} = P(\text{cats}) \dots P(\text{dog}) \quad \hat{y}^{<2>} = P(\_ | \text{cats}) \quad \hat{y}^{<3>} = P(\_ | \text{cats average}) \quad \hat{y}^{<8>} = P(\text{day} | \dots)$$



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

# Tổng kết

1. Sequence models: định nghĩa, loại dữ liệu, và ứng dụng.
2. Kiến trúc RNN: định nghĩa, ký hiệu.
3. Truyền xuôi và truyền ngược trên RNN.
4. Language model: định nghĩa, bài toán language model sử dụng RNN để giải quyết.

# TÀI LIỆU THAM KHẢO

1. Khoá học *Neural Network and Deep learning*, deeplearning.ai.
2. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep learning*, MIT Press, 2016.
3. Andrew Ng., *Machine Learning Yearning*. Link:  
<https://www.deeplearning.ai/machine-learning-yearning/>
4. Vũ Hữu Tiệp, *Machine Learning cơ bản*, NXB Khoa học và Kỹ thuật, 2018.