

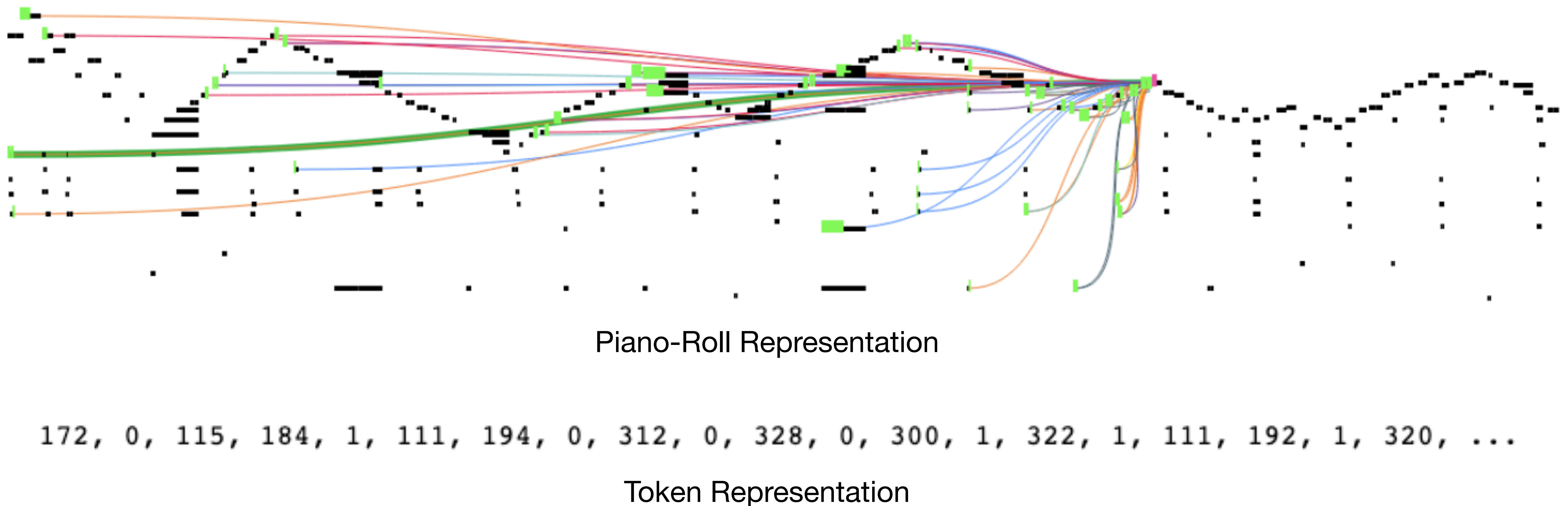
A.I. for Music

Week. 4

Soochul Park

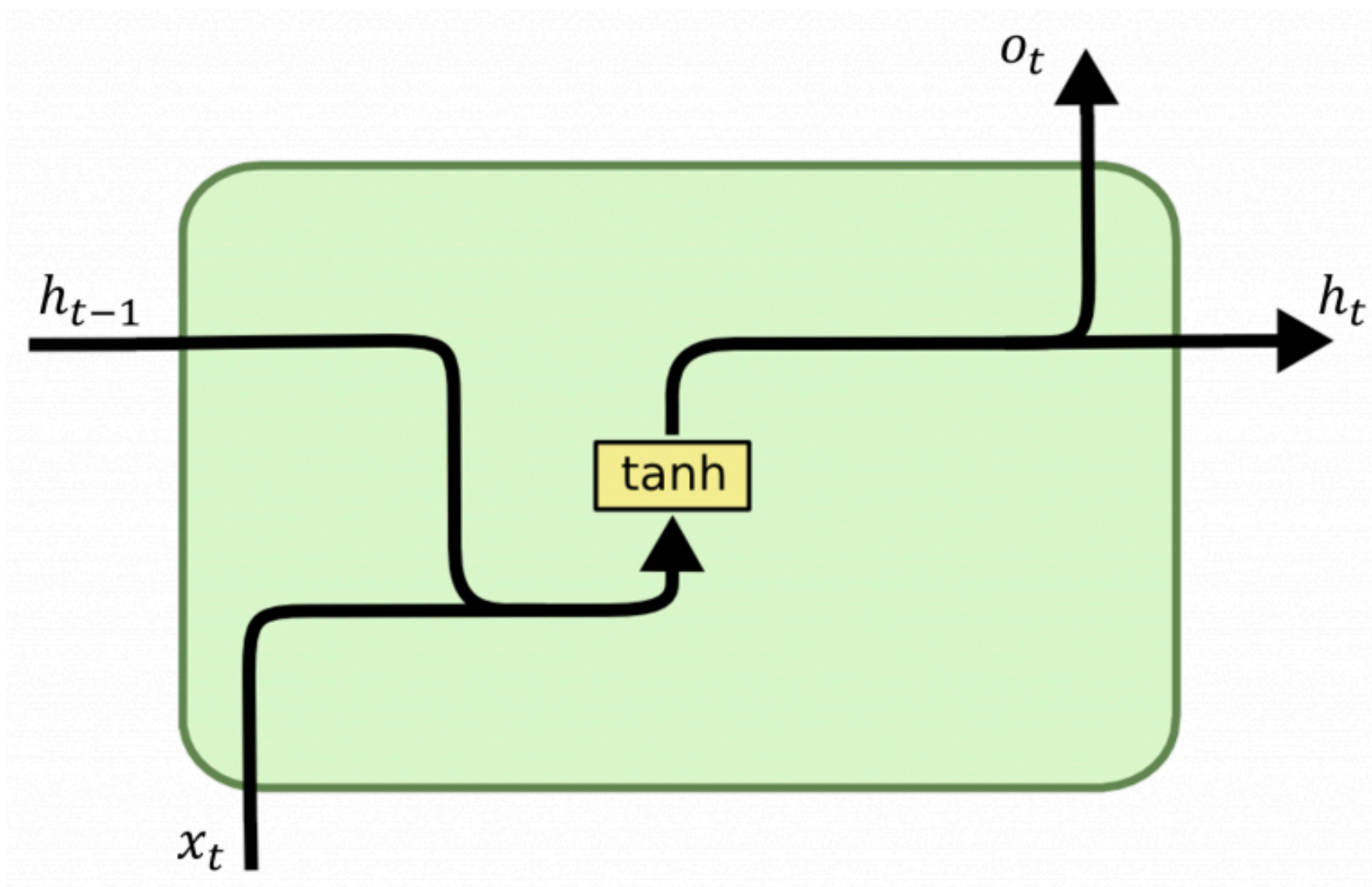
MIDI Generation in Deep Learning

- 딥러닝으로 MIDI를 생성하기 위해서는 MIDI 정보를 token화하여 sequential 데이터로 변환해야 합니다.
- 변환된 token sequence는 autoregressive model을 트레이닝 하는데 사용됩니다.
- Autoregressive model은 과거의 token들의 정보를 이용하여 현재의 token을 예측하는 모델입니다.

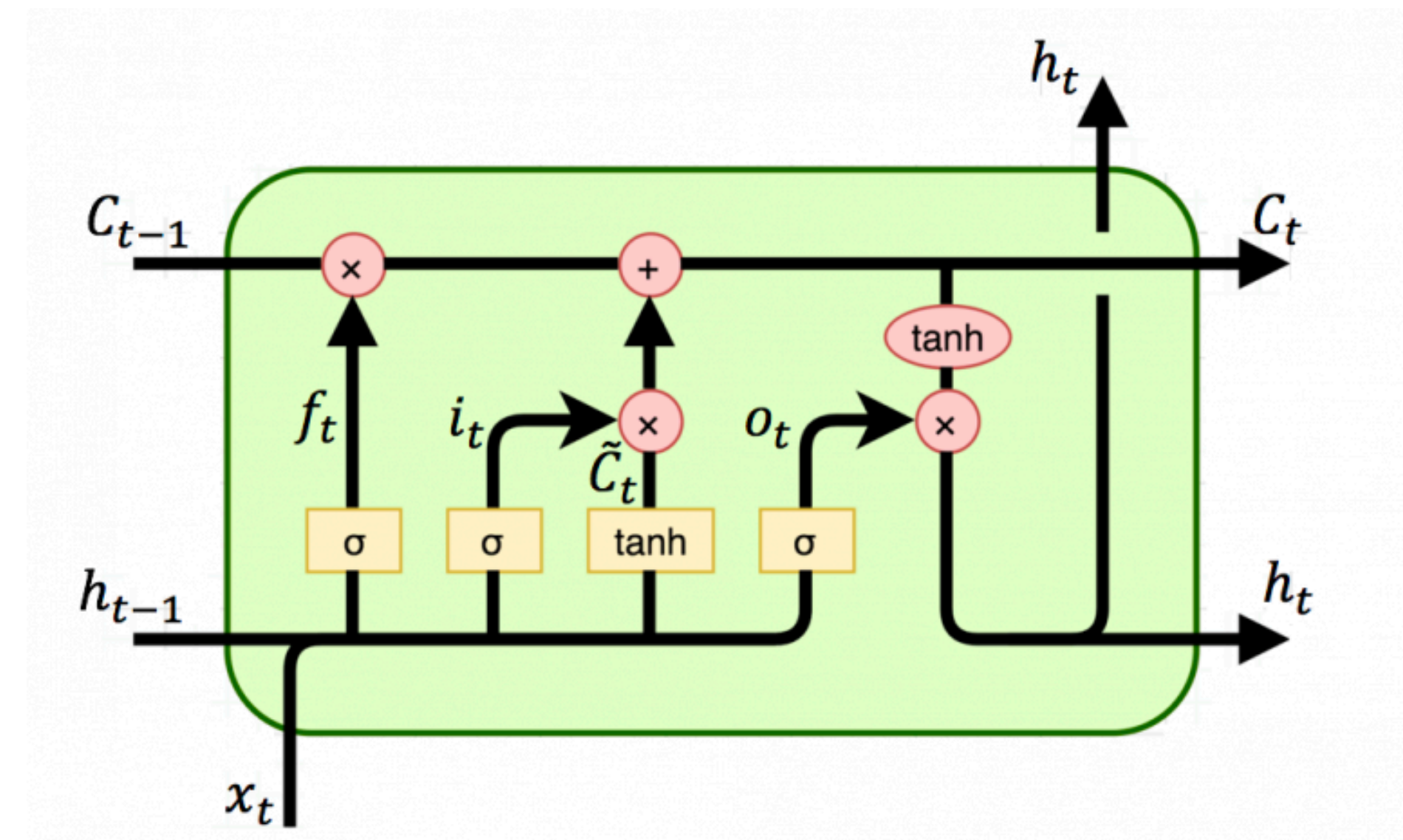


Autoregressive models

- Autoregressive model로는 RNN, GRU, LSTM, Transformer 등이 있습니다.
- RNN, GRU, LSTM은 deep learning 초창기 sequential data를 처리하기 위해 많이 사용되었으며, 최근에는 sequential data뿐만 아니라 이미지와 같은 2D 데이터들도 Transformer를 이용하여 분석되고 있습니다.



RNN Cell

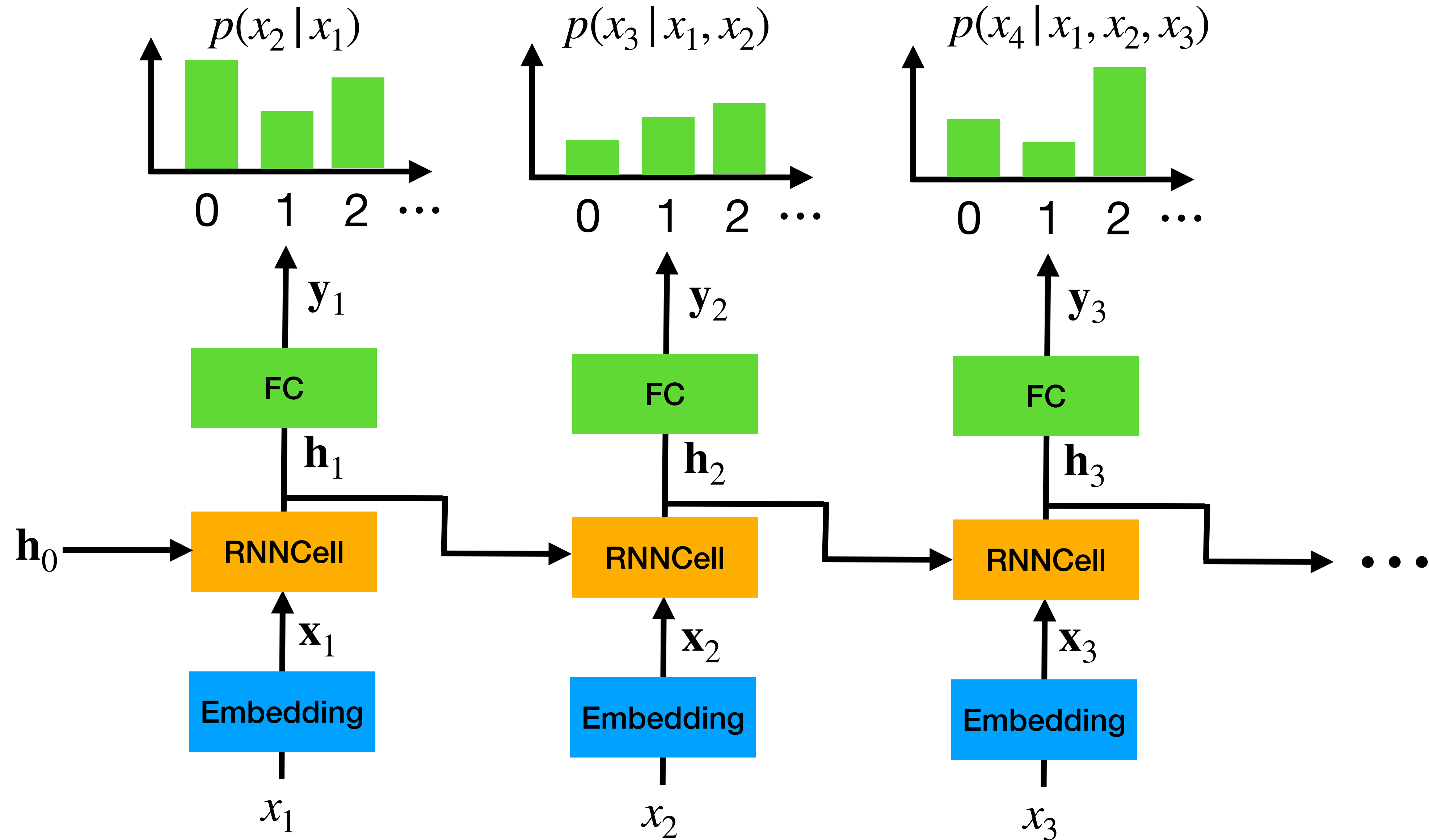


LSTM Cell

Autoregressive model using RNN

- RNN을 이용한 autoregressive model을 구성할 때 RNN만 독립적으로 사용되지는 않고, 대개 embedding layer, output을 위한 fully-connected layer 등과 함께 사용됩니다.
- 1. Embedding layer는 integer인 input token x_t 을 input vector \mathbf{x}_t 로 변환해주는 역할을 합니다.
- 2. RNN은 input vector \mathbf{x}_t 와 이전 시점의 state vector \mathbf{s}_{t-1} 를 입력받아 현재 시점의 state vector \mathbf{s}_t 를 출력합니다.
- 3. 현재 시점의 state vector \mathbf{s}_t 는 fully-connected layer를 통해 다음 시점에 일어날 token에 대한 확률분포를 나타내는 parameter로 변환됩니다.

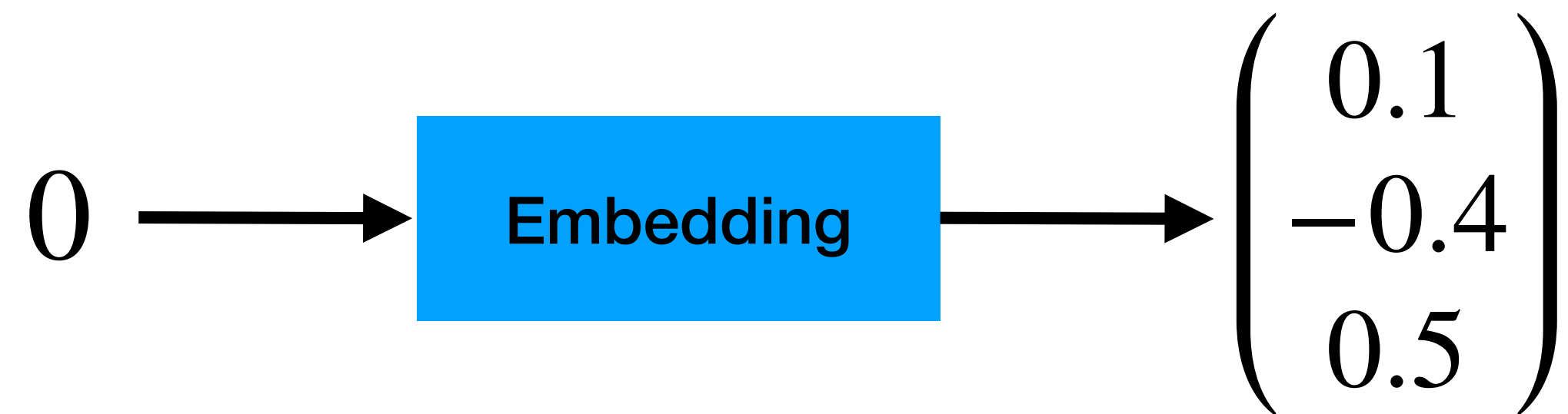
Overall Architecture



Embedding Layer

Integer

Vector

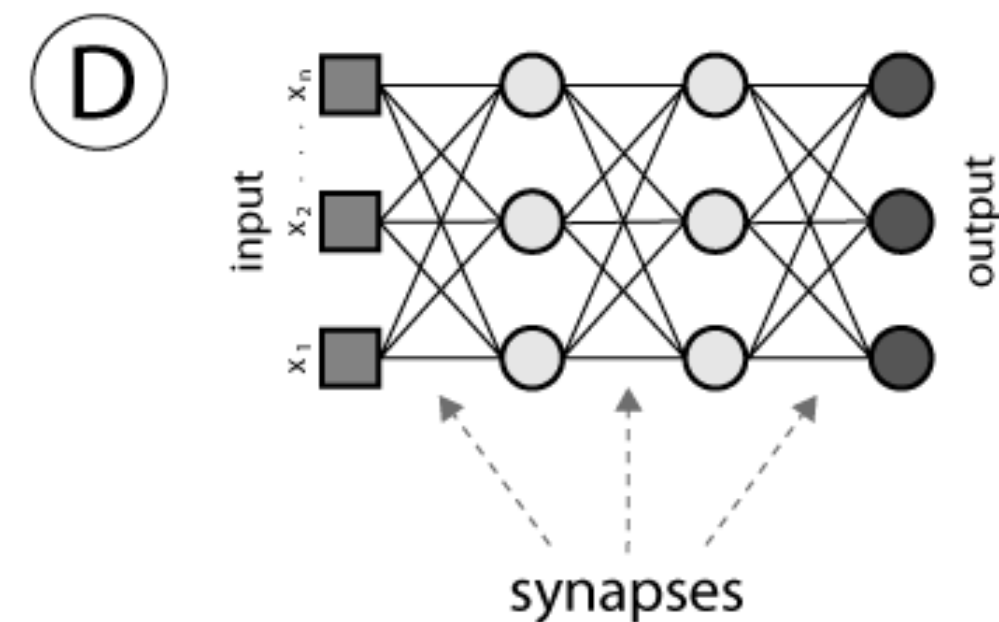
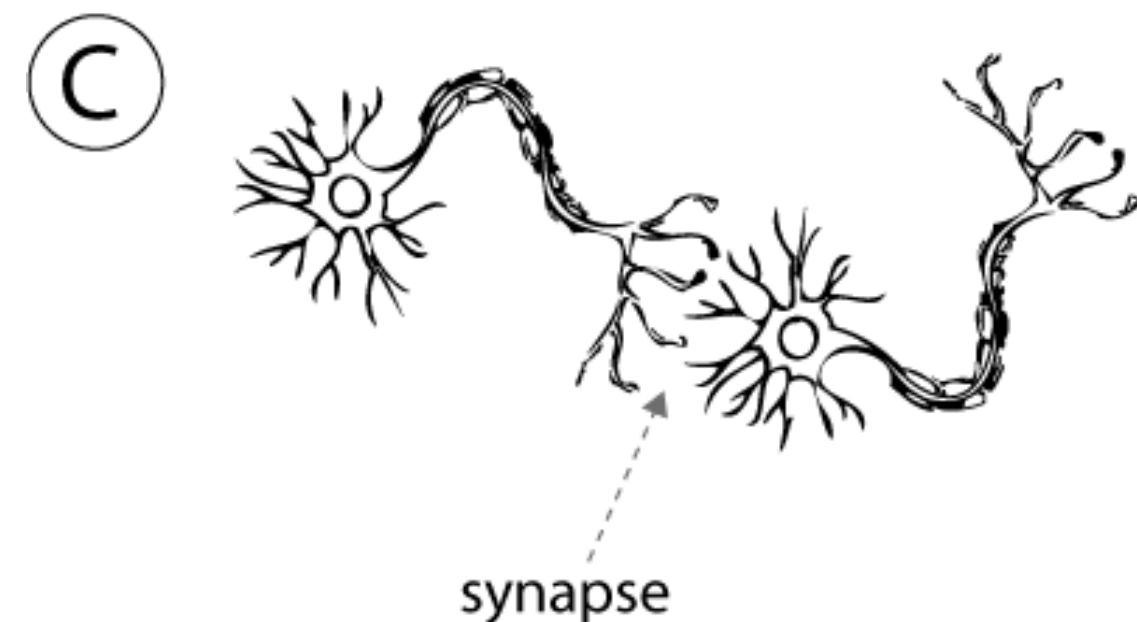
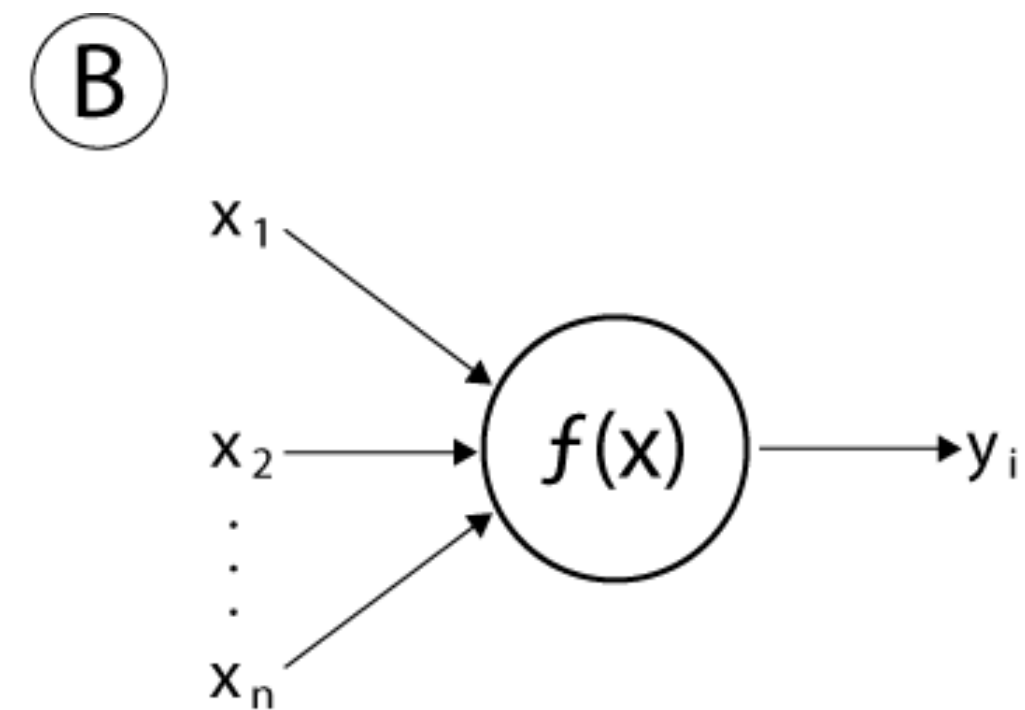
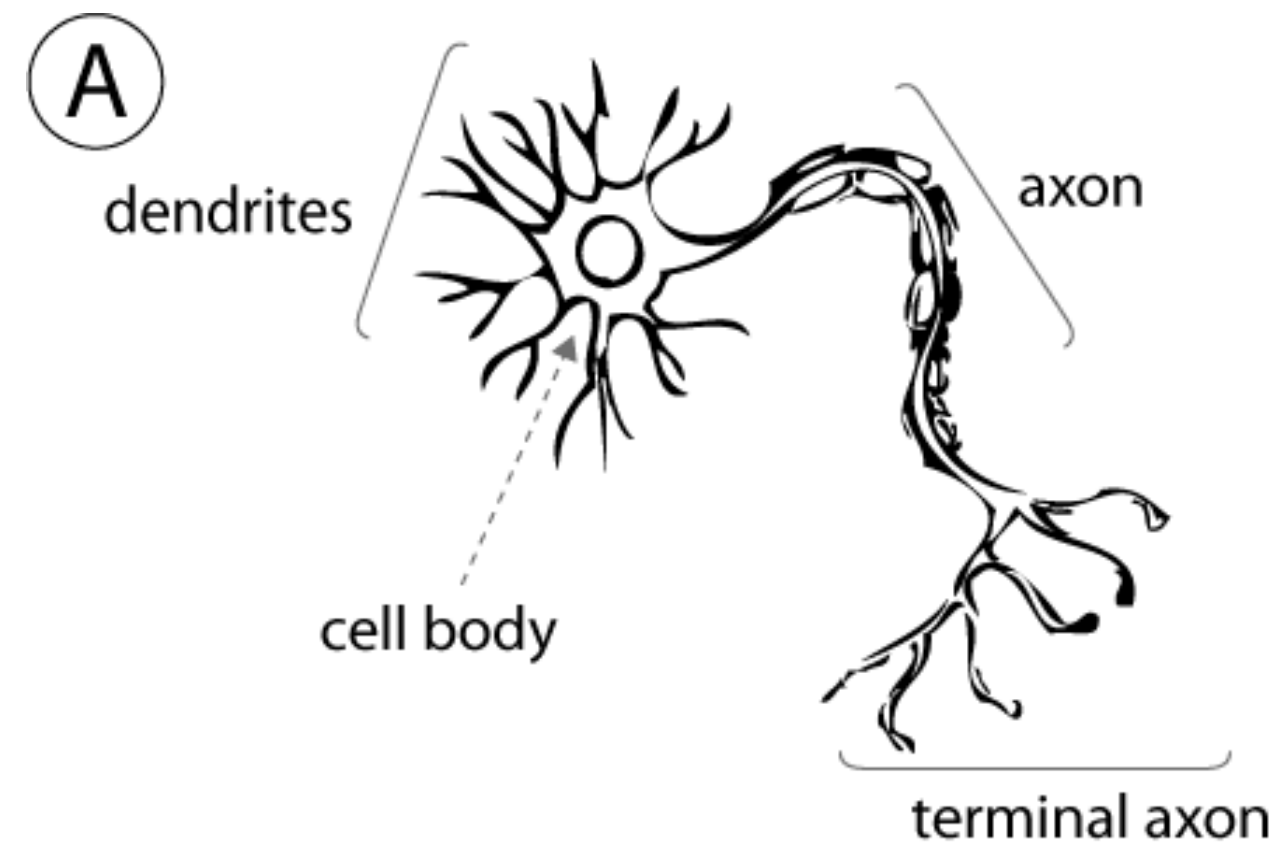


0	1	2	...
0.1	-0.4	-0.4	...
-0.4	0.8	-0.2	...
0.5	1.2	0.7	...

Embedding Table

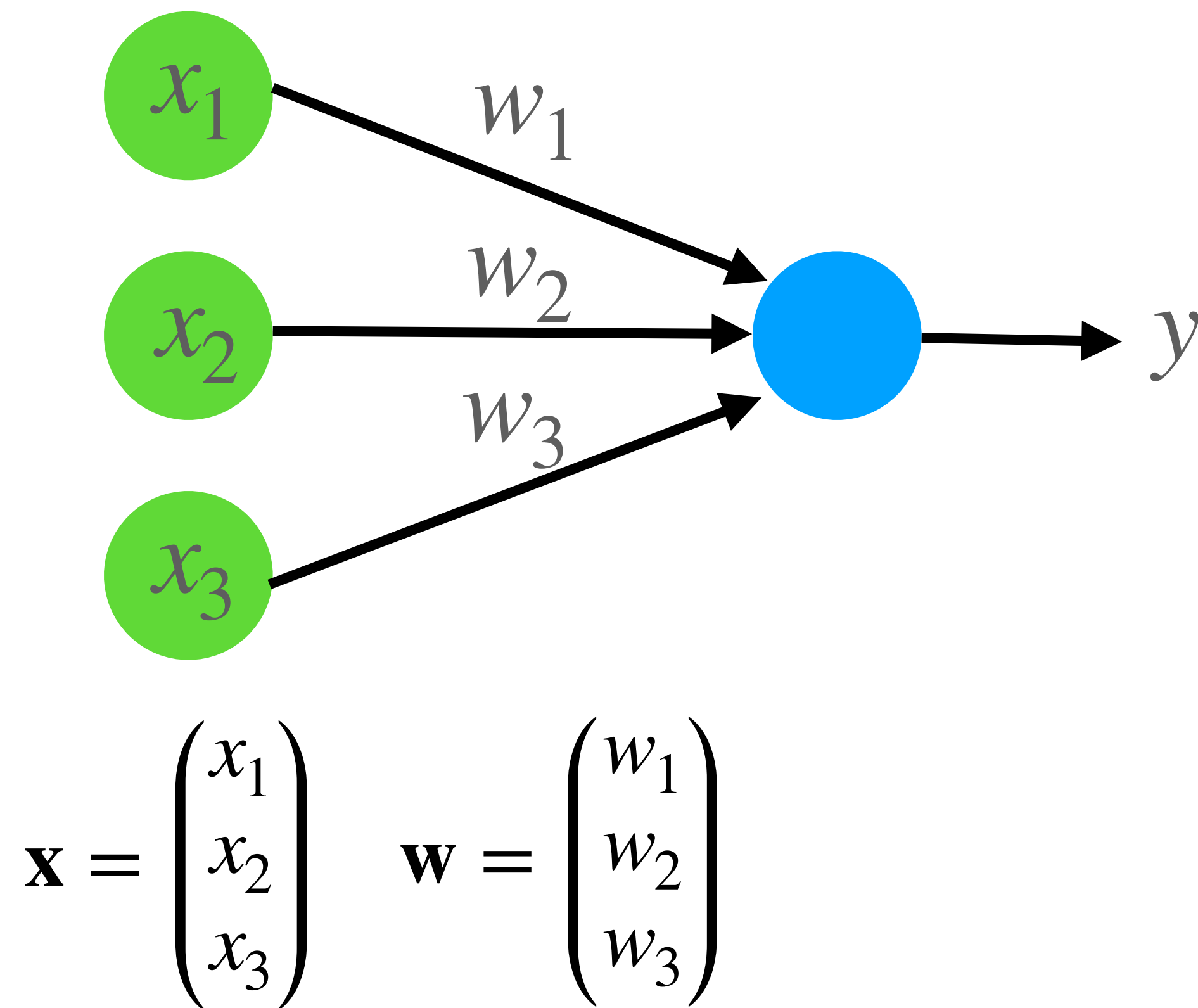
- Embedding layer는 integer 값을 입력으로 받아 vector를 출력합니다.
- Embedding layer 내부에는 embedding table을 갖고 있어 integer값이 들어오면 해당하는 순번의 column vector를 반환하도록 합니다.
- Embedding table의 값들은 트레이닝 가능한 상태로 둡니다.

Fully Connected Layer



- Full connected layer는 딥러닝의 가장 기본적인 유닛입니다.
- 뇌를 구성하는 neuron을 모방하여 만들었습니다.
- Neuron으로 전기 신호가 들어오고 다른 neuron들로 전기 신호가 전달되는 과정을 computer에 모방하여 나타내었습니다.

Fully Connected Layer - One Output



- Fully connected layer는 여러 scalar값을 입력으로 받습니다.

예) x_1, x_2, x_3

- 이 값들을 vector로 묶어서 표현할 수 있습니다.

예) $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

- Fully connected layer의 parameter로 weight

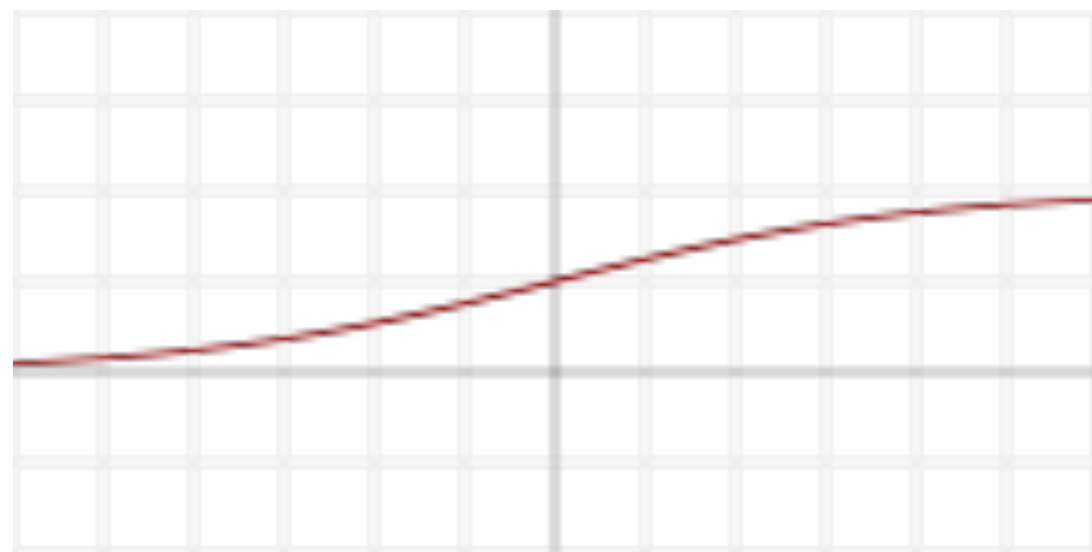
$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$ 가 있으며, 입력 \mathbf{x} 와 내적 (dot-product) 연산을 합니다.

- 내적 연산 후 sigmoid나 tanh와 같은 activation 연산을 수행합니다.

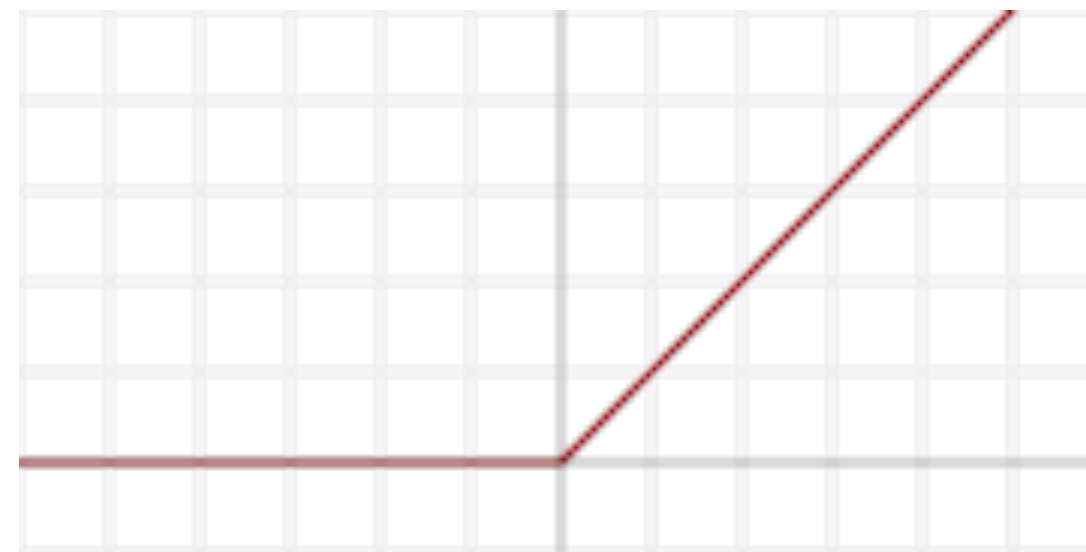
- $$y = \sigma(w_1x_1 + w_2x_2 + w_3x_3)$$
$$= \sigma(\mathbf{w}^T \mathbf{x})$$

Fully Connected Layer - Activations

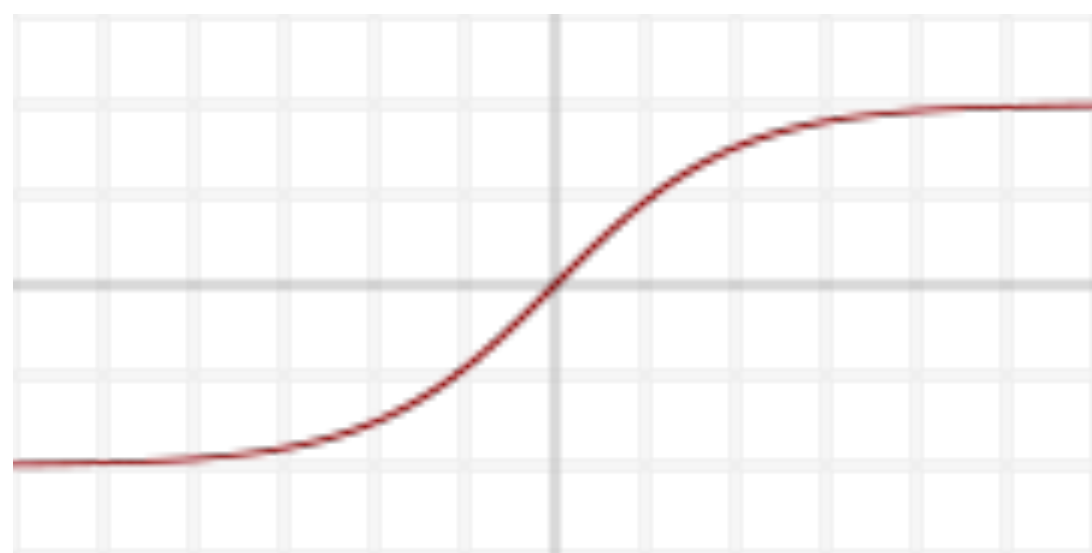
- Activation function은 non-linearity를 연산에 부여하기 위해 사용됩니다.



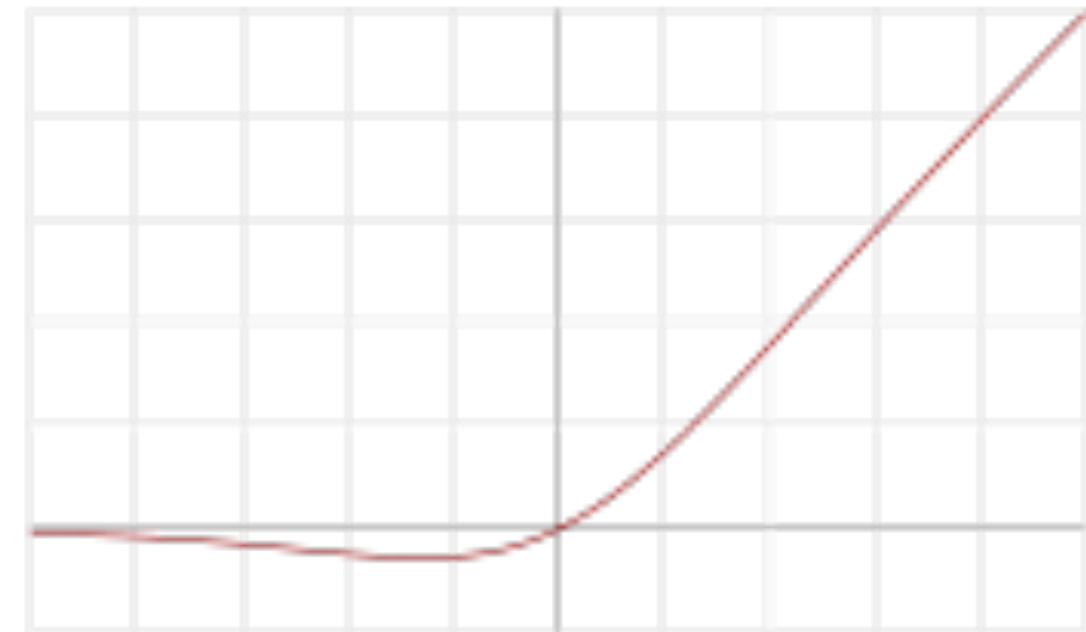
Sigmoid



ReLU



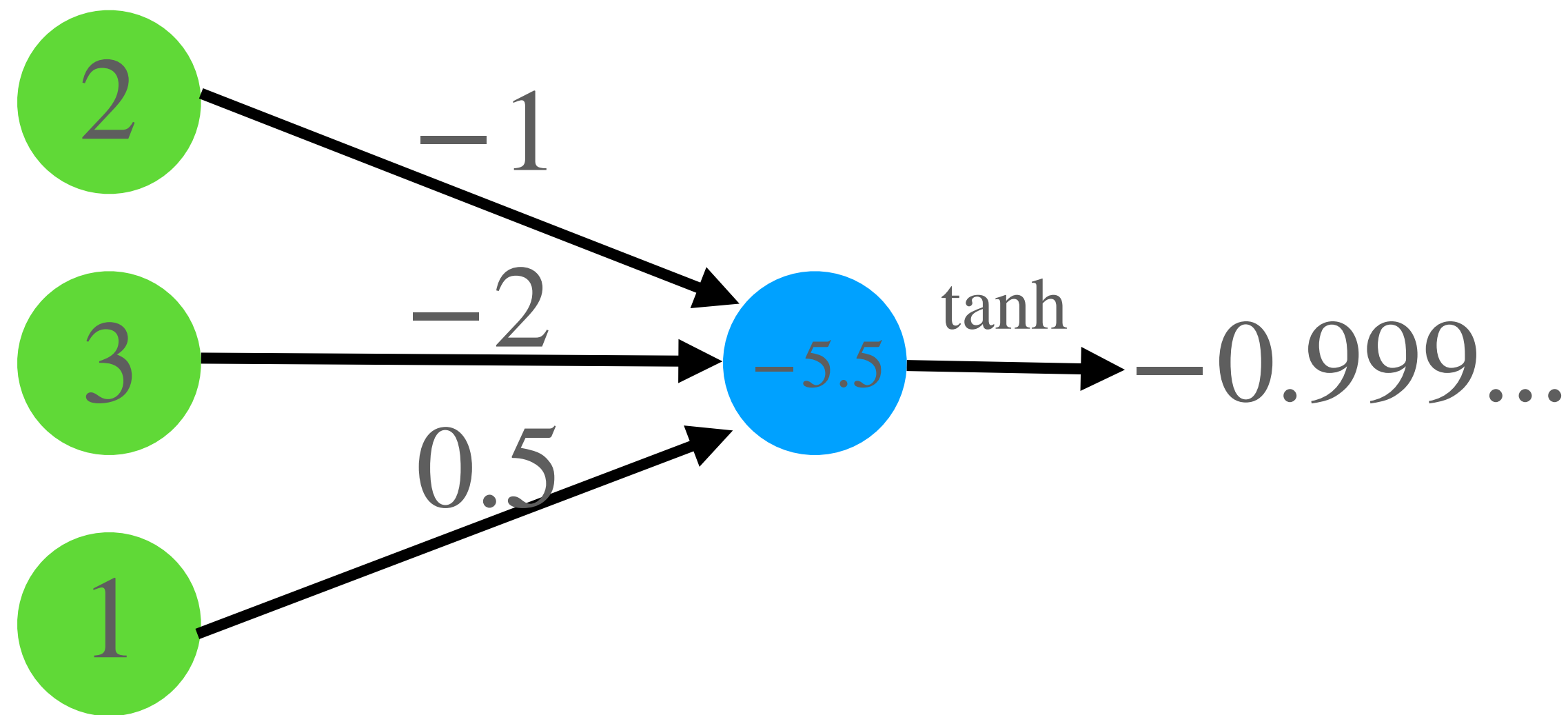
Tanh



SiLU

Fully Connected Layer - One Output Example

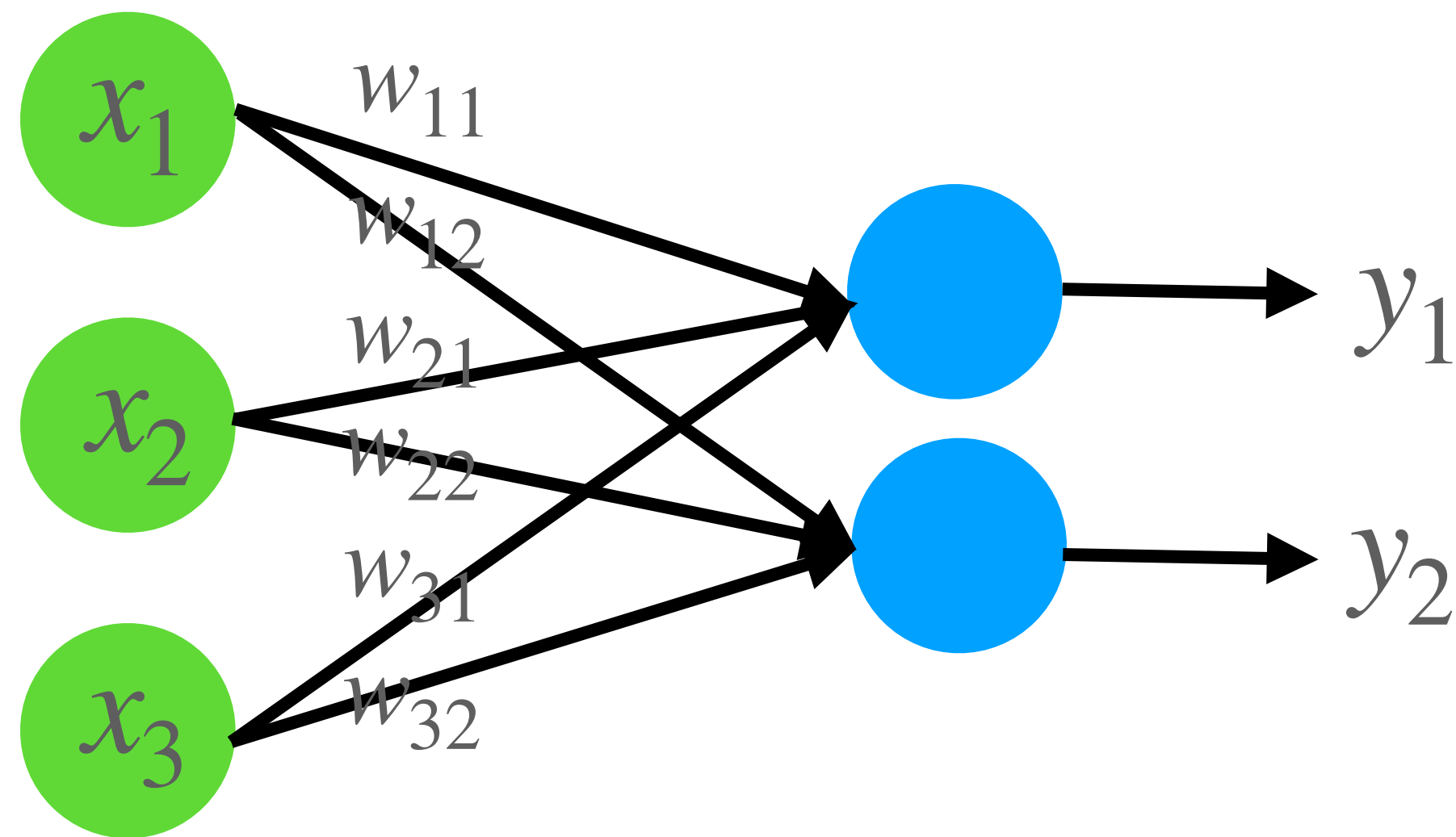
- $$y = \sigma(w_1x_1 + w_2x_2 + w_3x_3)$$
$$= \sigma(2 \cdot (-1) + 3 \cdot (-2) + 1 \cdot 0.5)$$
$$= \sigma(5.5)$$
$$= \tanh(5.5) = -0.999\dots$$



$$\mathbf{x} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} -1 \\ -2 \\ 0.5 \end{pmatrix} \quad y = -0.999\dots$$

Fully Connected Layer - Multi Outputs

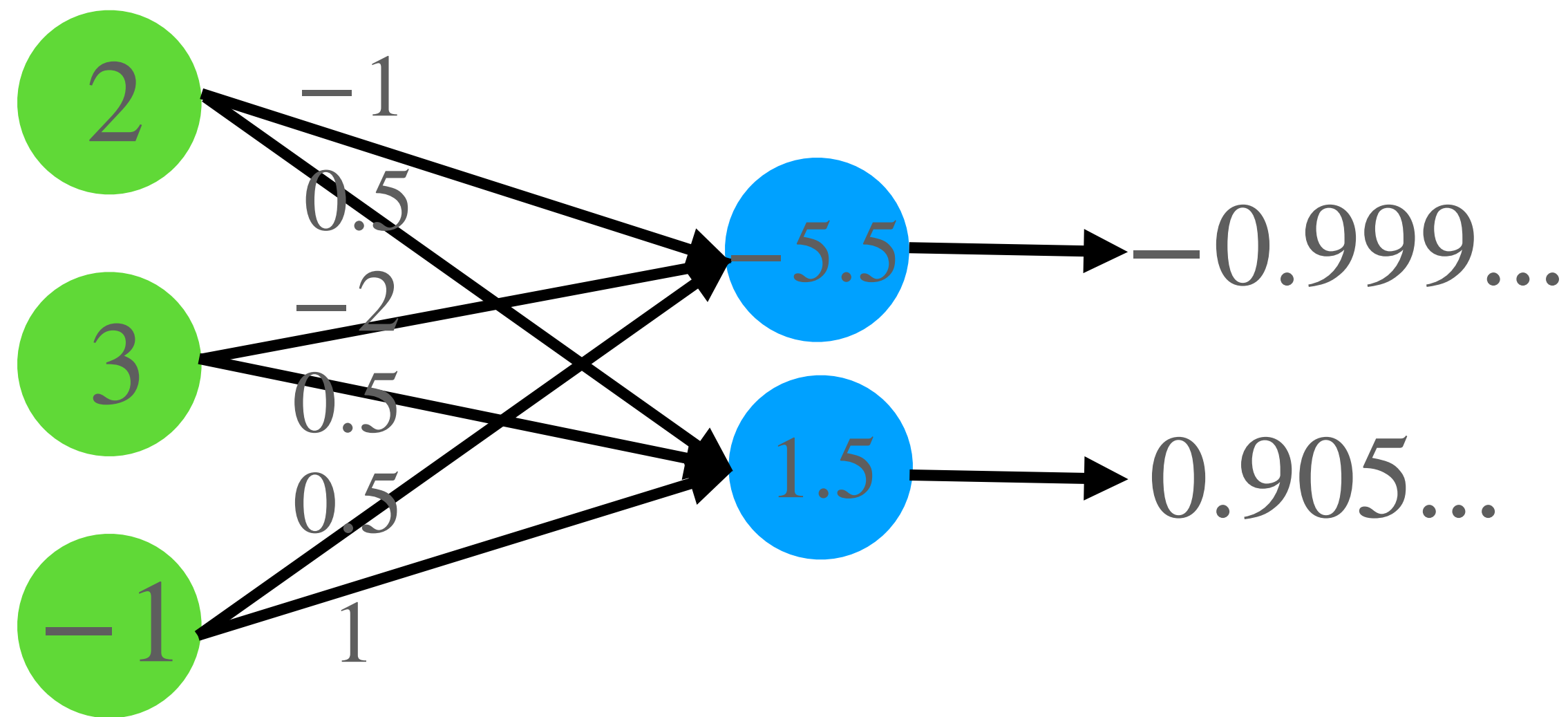
- 출력을 scalar값 여러개인 vector로 내보낼 수도 있습니다.
- 이 때, weight \mathbf{W} 는 (output dim, input dim) shape을 갖는 matrix로 표현할 수 있습니다.



- $$\mathbf{y} = \begin{pmatrix} \sigma(w_{11}x_1 + w_{21}x_2 + w_{31}x_3) \\ \sigma(w_{12}x_1 + w_{22}x_2 + w_{32}x_3) \end{pmatrix}$$
$$= \sigma(\mathbf{W}\mathbf{x})$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \mathbf{W} = \begin{pmatrix} w_{11} & w_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

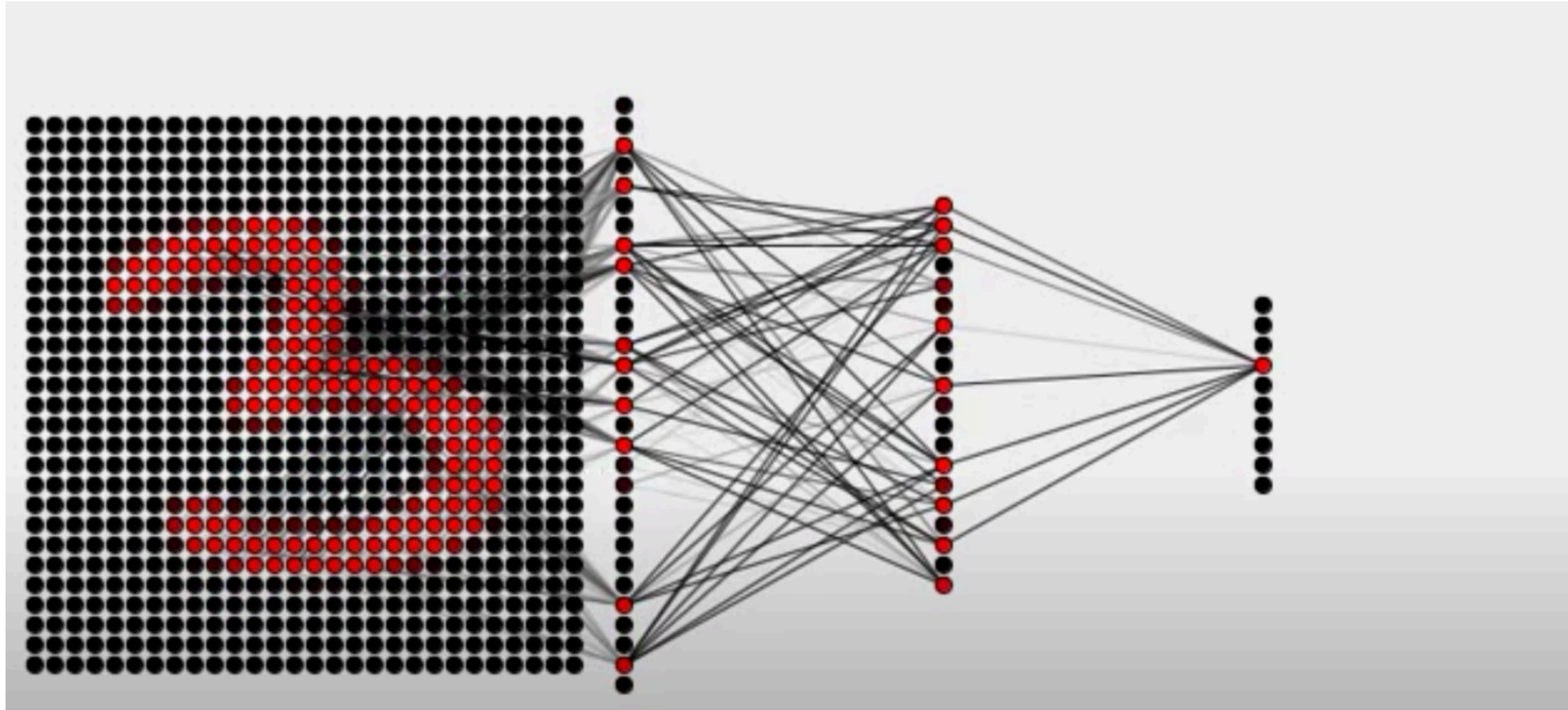
Fully Connected Layer - Multi Outputs Example



$$\mathbf{x} = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} \quad \mathbf{W} = \begin{pmatrix} -1 & -2 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} -0.999... \\ 0.905... \end{pmatrix}$$

$$\begin{aligned} \bullet \quad \mathbf{y} &= \begin{pmatrix} \sigma(w_{11}x_1 + w_{21}x_2 + w_{31}x_3) \\ \sigma(w_{12}x_1 + w_{22}x_2 + w_{32}x_3) \end{pmatrix} \\ &= \begin{pmatrix} \sigma((-1) \cdot 2 + (-2) \cdot 3 + 0.5 \cdot (-1)) \\ \sigma(0.5 \cdot 2 + 0.5 \cdot 3 + 1 \cdot (-1)) \end{pmatrix} \\ &= \begin{pmatrix} \sigma(-5.5) \\ \sigma(1.5) \end{pmatrix} \\ &= \begin{pmatrix} \tanh(-5.5) \\ \tanh(1.5) \end{pmatrix} \\ &= \begin{pmatrix} -0.999... \\ 0.905... \end{pmatrix} \end{aligned}$$

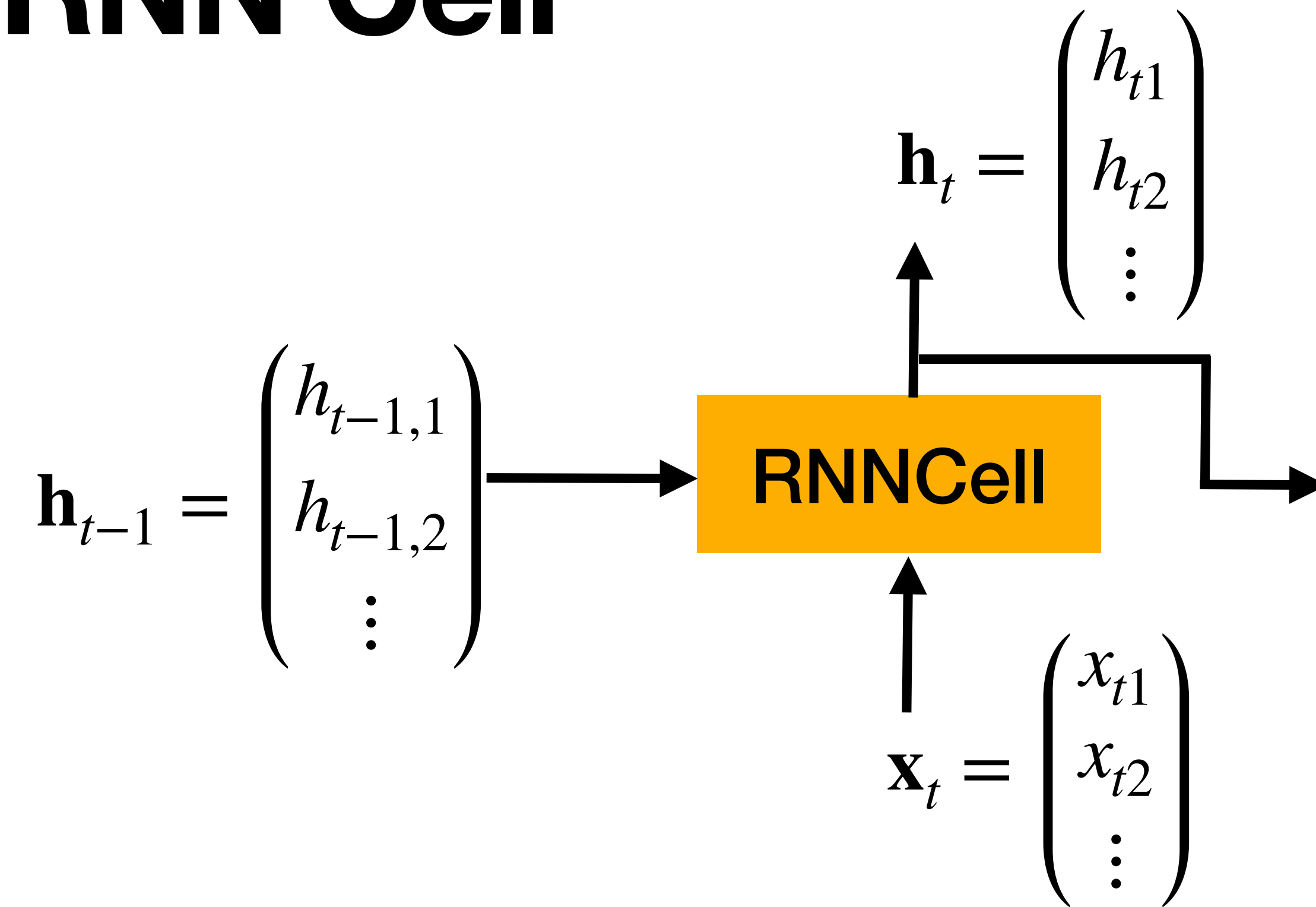
Fully Connected Layer - Image Classification



Deep learning - Neuronix training on MNIST dataset (3 fully connected layers)

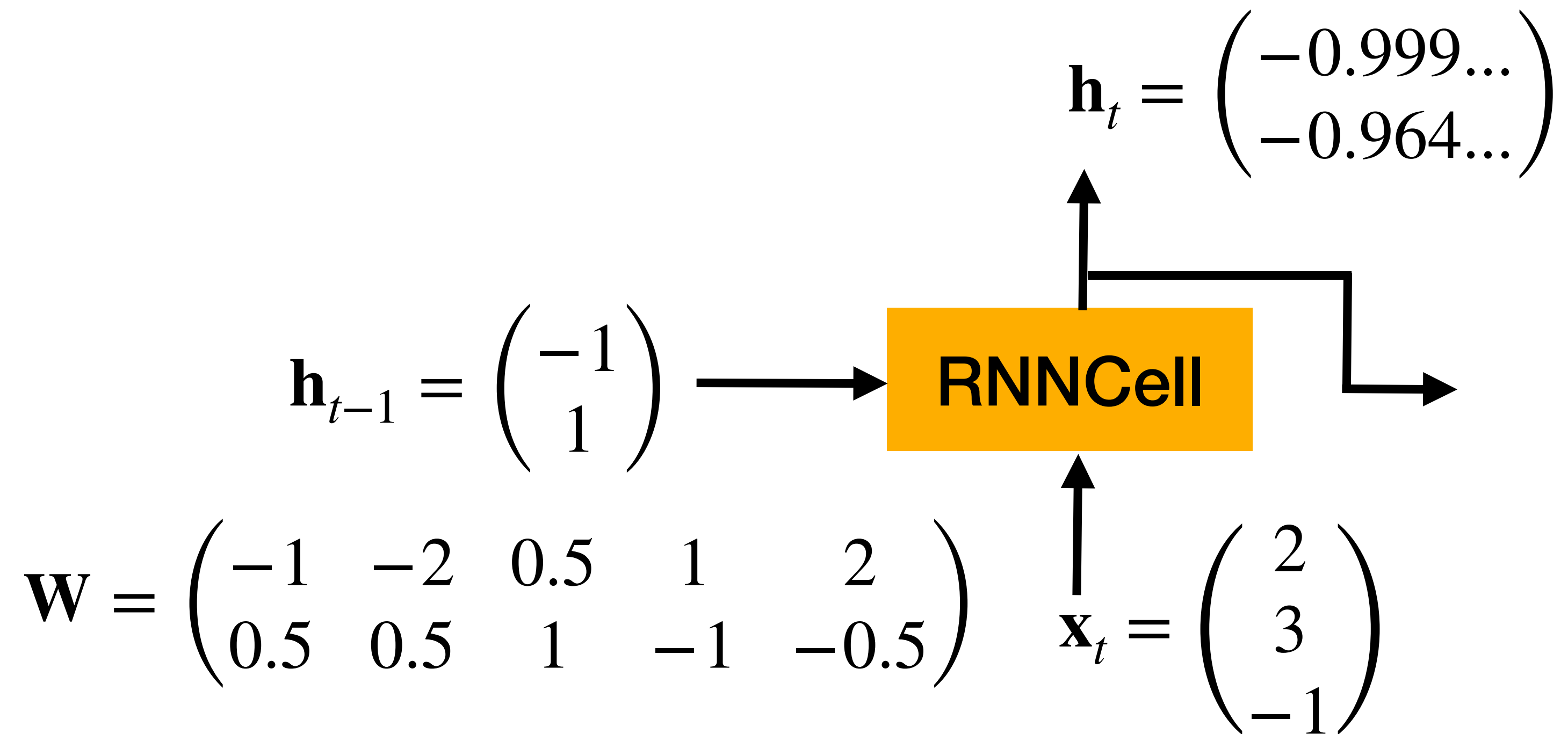
<https://www.youtube.com/watch?v=Ongw6bvlsUA>

RNN - RNN Cell



- $\mathbf{h}_t = \text{RNNCell}(\mathbf{x}_t, \mathbf{h}_{t-1}) = \tanh(\mathbf{W}(\mathbf{x}_t; \mathbf{h}_{t-1}))$
- RNN Cell은 현재 시점의 입력 \mathbf{x}_t 와 이전 시점의 hidden state \mathbf{h}_{t-1} 를 입력받아 현재 시점의 hidden state \mathbf{h}_t 를 출력합니다.
- 입력 \mathbf{x}_t 와 이전 시점의 hidden state \mathbf{h}_{t-1} 를 concatenate한 후 RNN의 parameter \mathbf{W} 를 곱하고 (matrix multiplication), activation (주로 tanh)를 적용하여 현재 시점의 hidden state \mathbf{h}_t 을 출력합니다.

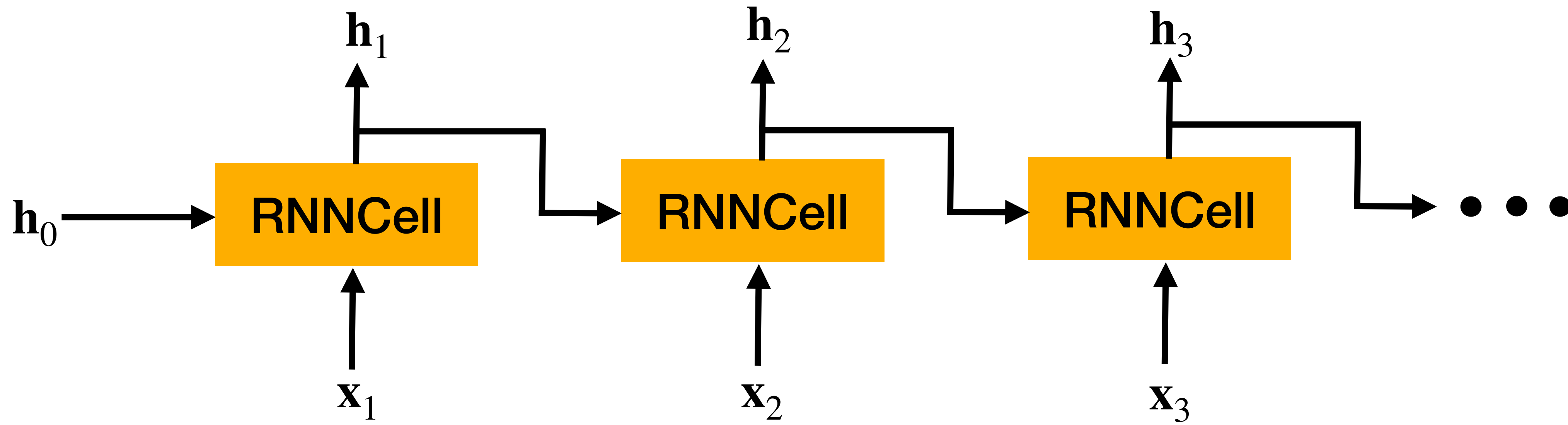
RNN - RNN Cell Example



- $$\mathbf{h}_t = \text{RNNCell}(\mathbf{x}_t, \mathbf{h}_{t-1}) = \tanh(\mathbf{W}(\mathbf{x}_t; \mathbf{h}_{t-1}))$$

$$= \tanh \left(\begin{pmatrix} -1 & -2 & 0.5 & 1 & 2 \\ 0.5 & 0.5 & 1 & -1 & -0.5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ -1 \\ -1 \\ 1 \end{pmatrix} \right) = \tanh \begin{pmatrix} -7.5 \\ 2 \end{pmatrix} = \begin{pmatrix} -0.999... \\ 0.964... \end{pmatrix}$$

RNN - Unfolded



- $h_1 = \text{RNNCell}(x_1, h_0)$
 $h_2 = \text{RNNCell}(x_2, h_1)$
 $h_3 = \text{RNNCell}(x_3, h_2)$
⋮

- RNN Cell은 매 시점마다 다른 입력 x_1, x_2, x_3, \dots 에 대해 같은 연산을 수행하며,
- 매 시점의 연산의 결과 h_1, h_2, h_3, \dots 를 출력합니다.

RNN - Example

QUEENE:

I had thought thou hadst a Roman; for the oracle,
Thus by All bids the man against the word,
Which are so weak of care, by old care done;
Your children were in your holy love,
And the precipitation through the bleeding throne.

BISHOP OF ELY:

Marry, and will, my lord, to weep in such a one were prettiest;
Yet now I was adopted heir
Of the world's lamentable day,
To watch the next way with his father with his face?

ESCALUS:

The cause why then we are all resolved more sons.

VOLUMNIA:

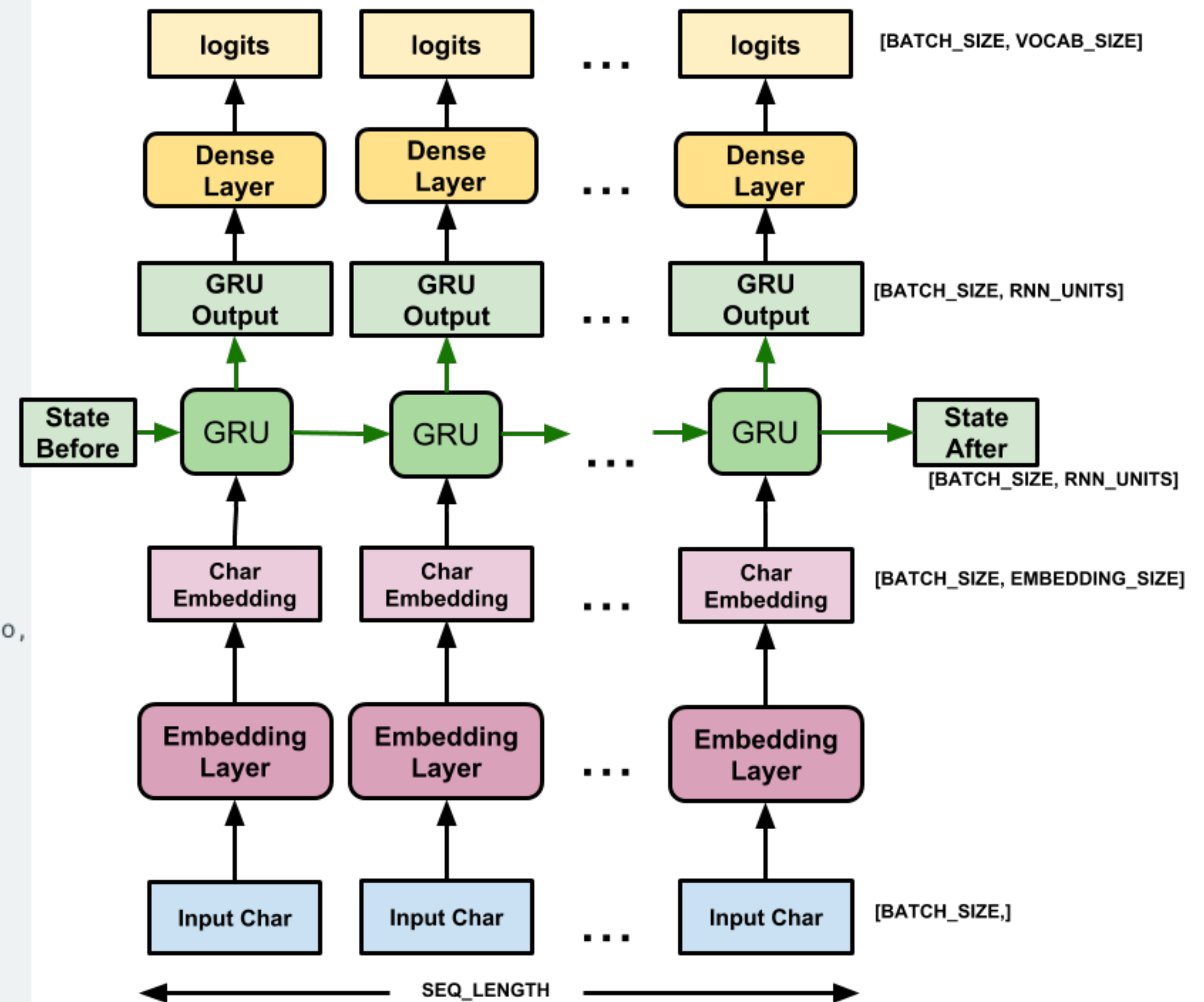
O, no, no, no, no, no, no, no, no, no, no, no, no, no, no, no, no,
And love and pale as any will to that word.

QUEEN ELIZABETH:

But how long have I heard the soul for this world,
And show his hands of life be proved to stand.

PETRUCHIO:

I say he look'd on, if I must be content
To stay him from the fatal of our country's bliss.
His lordship pluck'd from this sentence then for prey,
And then let us twain, being the moon,
were she such a case as fills m



Tensorflow RNN Tutorial

https://www.tensorflow.org/text/tutorials/text_generation

Shakespear의 소설 데이터를 이용하여 RNN으로 생성한 결과