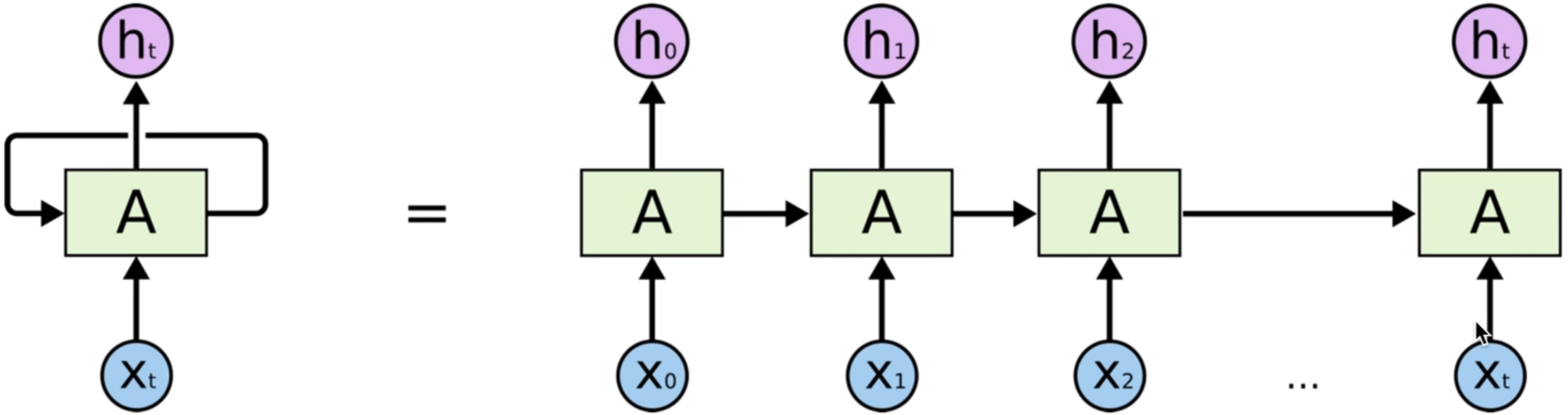
Lecture 12

Recurrent Neural Network (RNN)

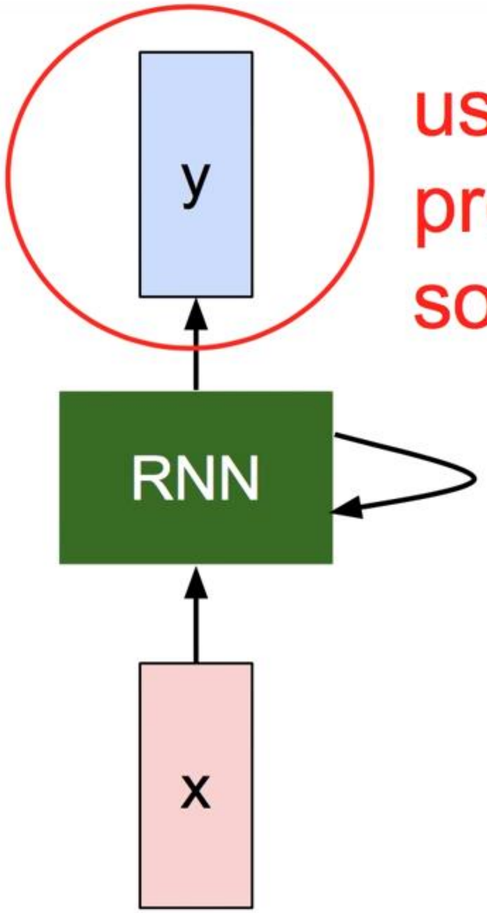
Sequence data

- We don’t understand one word only
- We understand based on the previous words
- NN/CNN cannot do this

Sequence data



Recurrent Neural Network



usually want to predict a vector at some time steps

Recurrent Neural Network

- We can process a sequence of vectors x by applying a recurrence formula at every time step:

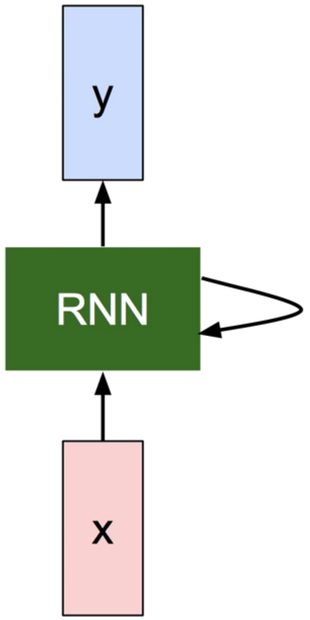
$$h_t = f_W(h_{t-1}, x_t)$$

new state

some function with parameters W

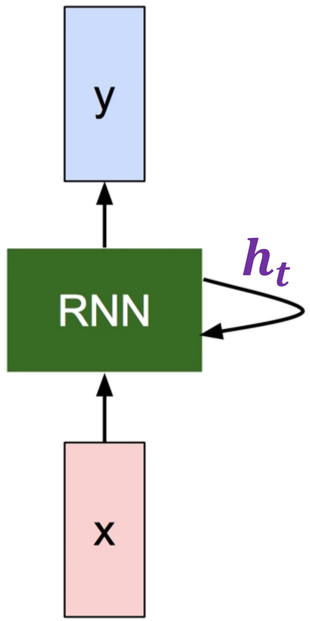
old state

input vector at some time step



(Vanilla) Recurrent Neural Network

- The state consists of a single **"hidden"** vector **h**:



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

WX

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

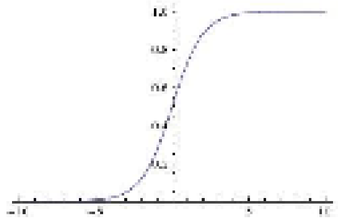
$$y_t = W_{hy} h_t$$

Activation Functions

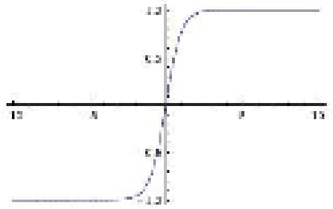
Activation Functions

Sigmoid

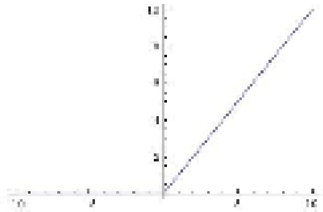
$\sigma(x) = 1/(1+e^{-x})$



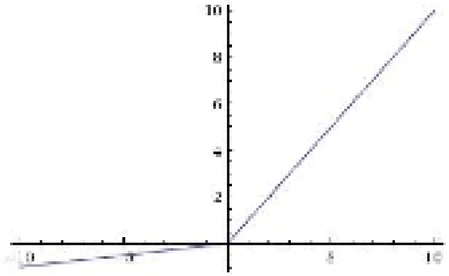
tanh tanh(x)



ReLU max(0,x)



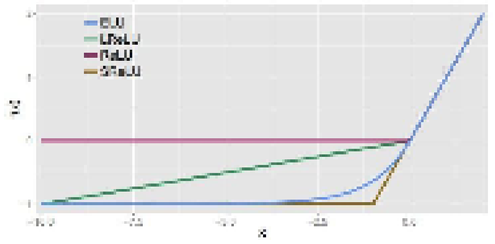
Leaky ReLU
max(0.1x, x)



Maxout $\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

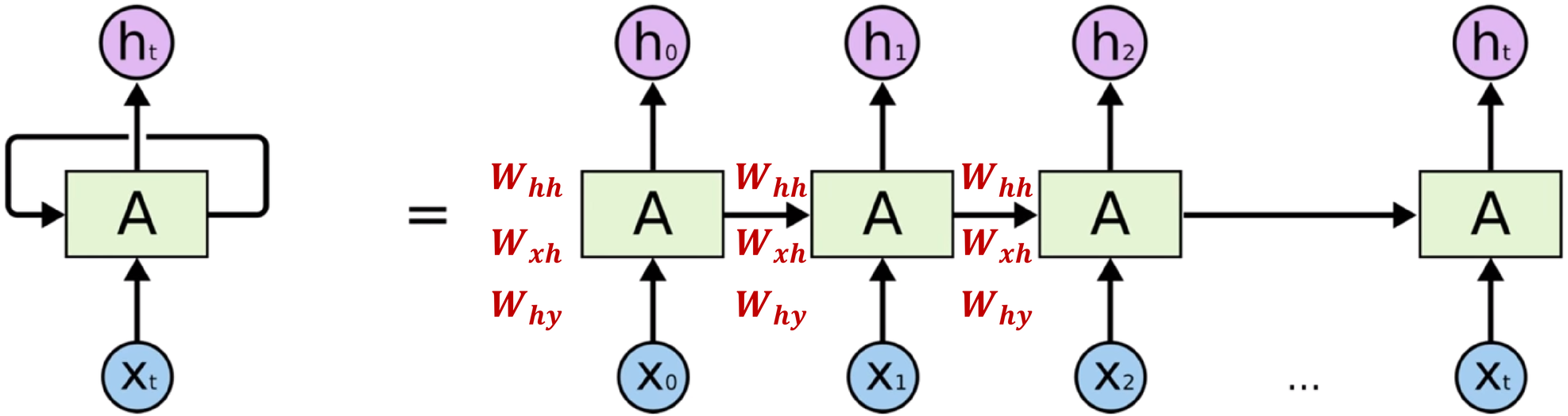
$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha\,(\exp(x) - 1) & \text{if } x \leq 0 \end{cases}$$



(Vanilla) Recurrent Neural Network

- Notice : **the same function** and **the same set of parameters** are used at every time step.

W_{hh}
W_{xh}
W_{hy}

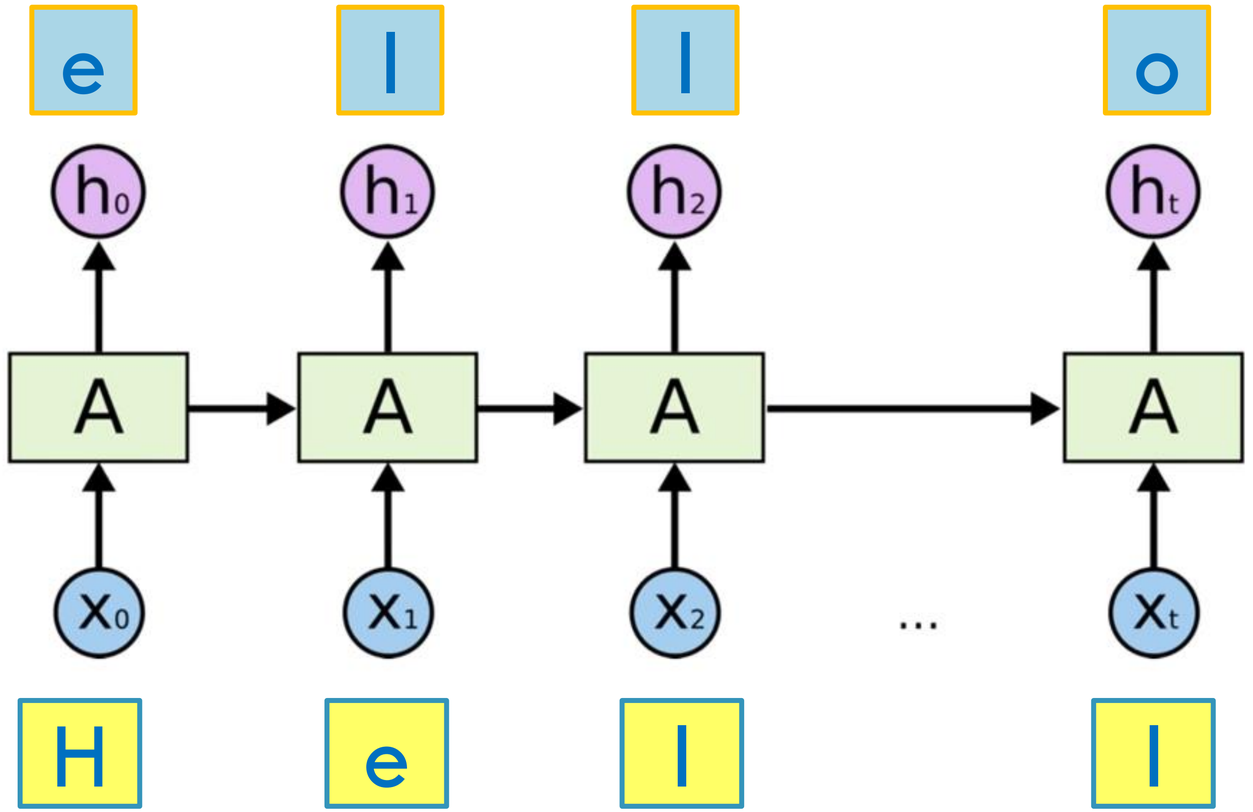


(Vanilla) Recurrent Neural Network

- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training
Sequence :
“hello”

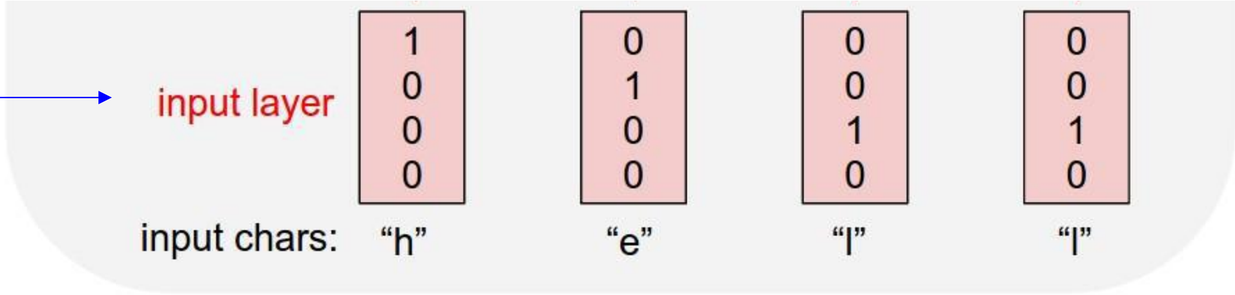


(Vanilla) Recurrent Neural Network

- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training
Sequence :
“hello”

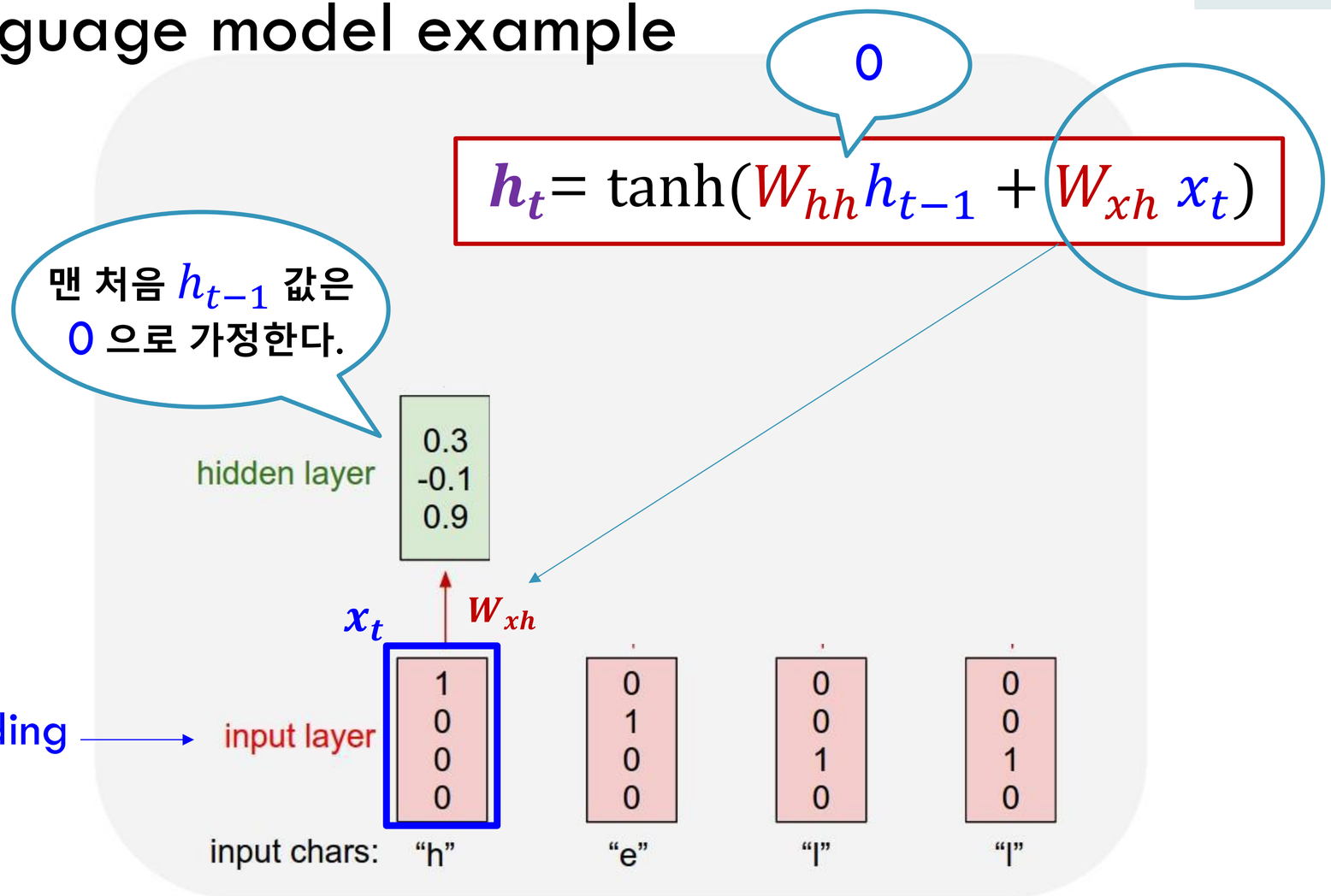


(Vanilla) Recurrent Neural Network

- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training Sequence :
“hello”

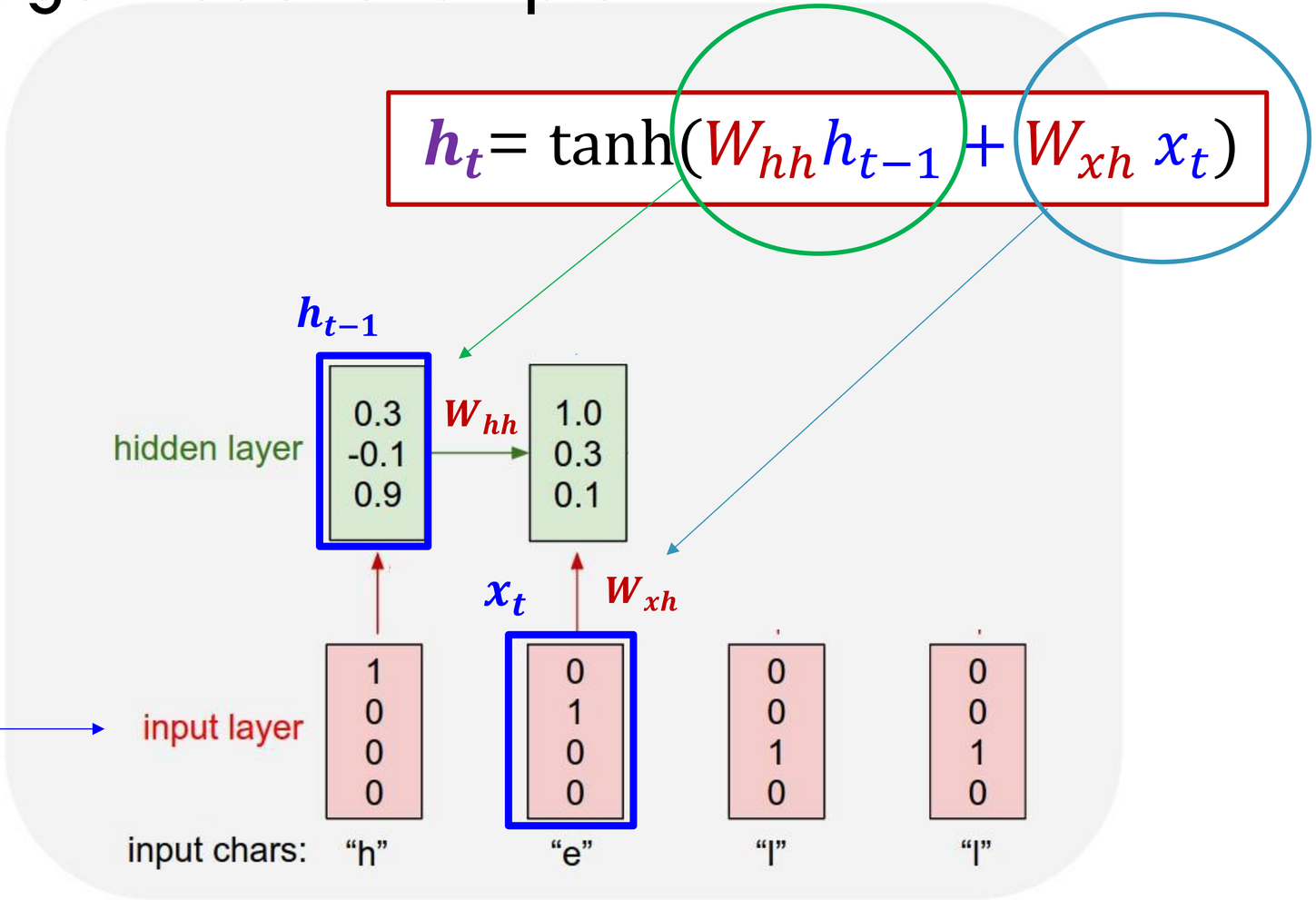


(Vanilla) Recurrent Neural Network

- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training Sequence :
“hello”



(Vanilla) Recurrent Neural Network

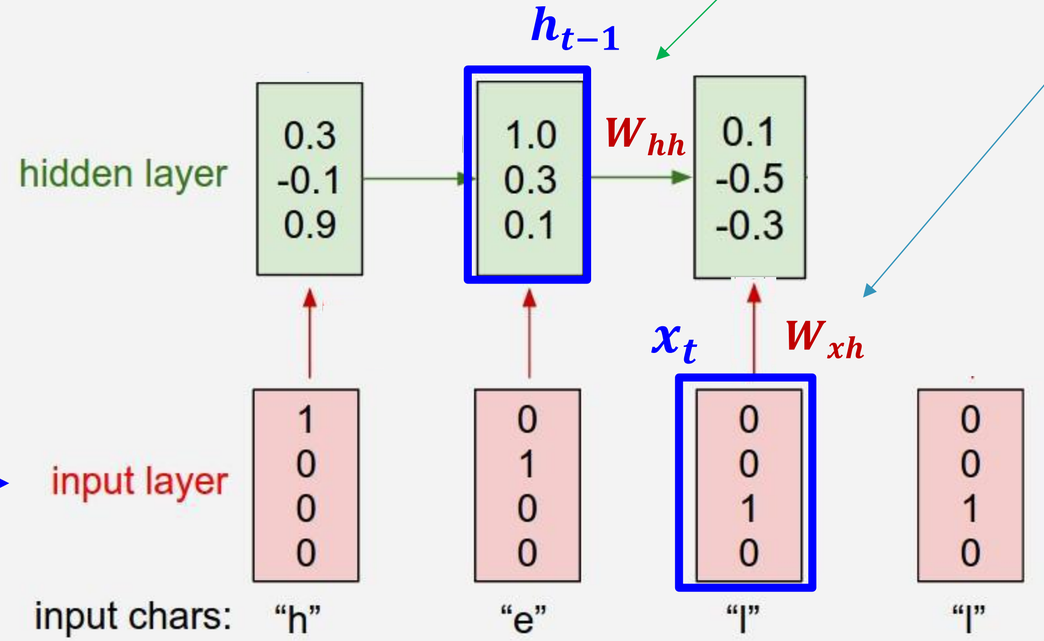
- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training Sequence :
“hello”

one-hot encoding

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

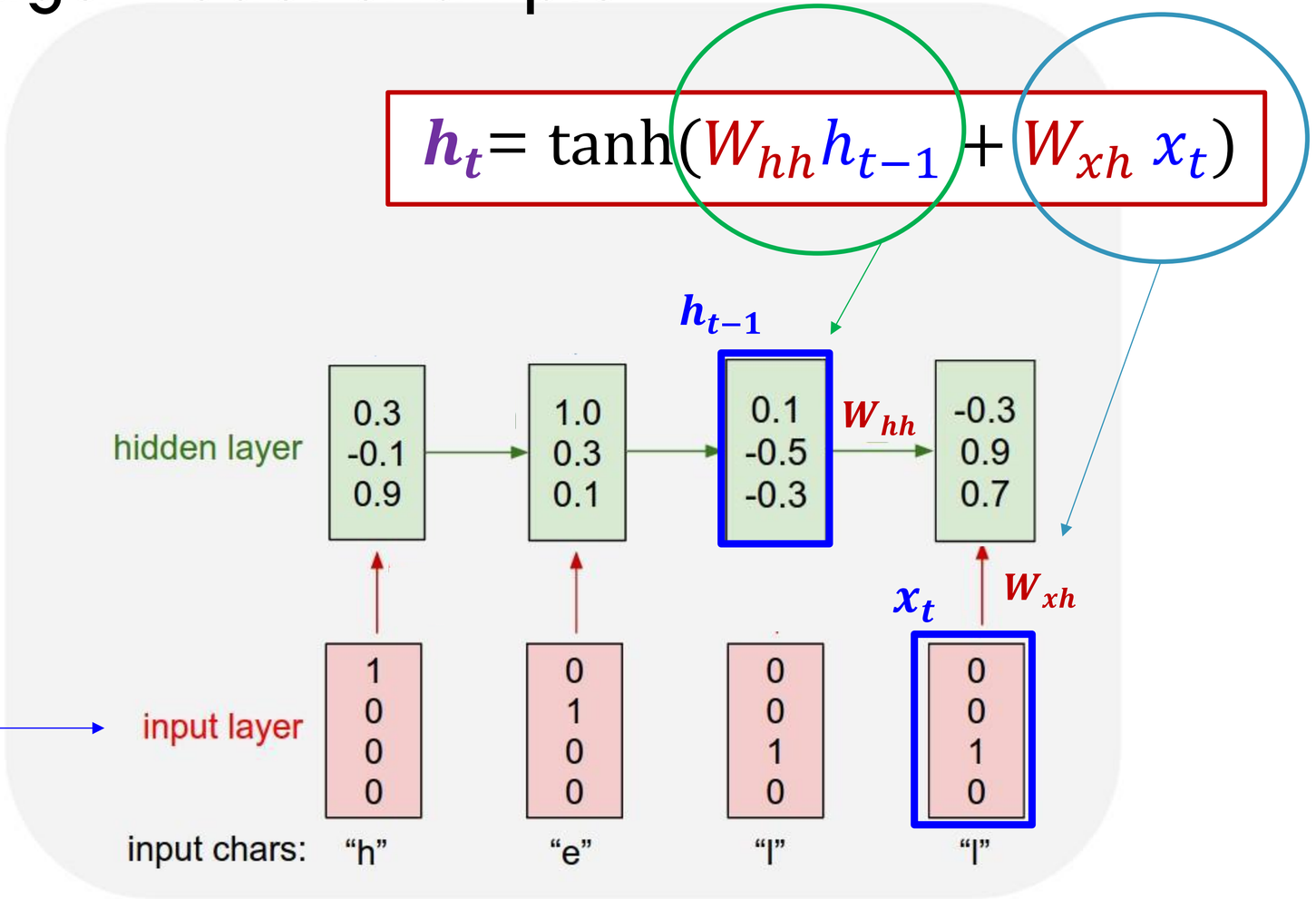


(Vanilla) Recurrent Neural Network

- Character-level language model example

Vocabulary :
[h, e, l, o]

Example training Sequence :
“hello”



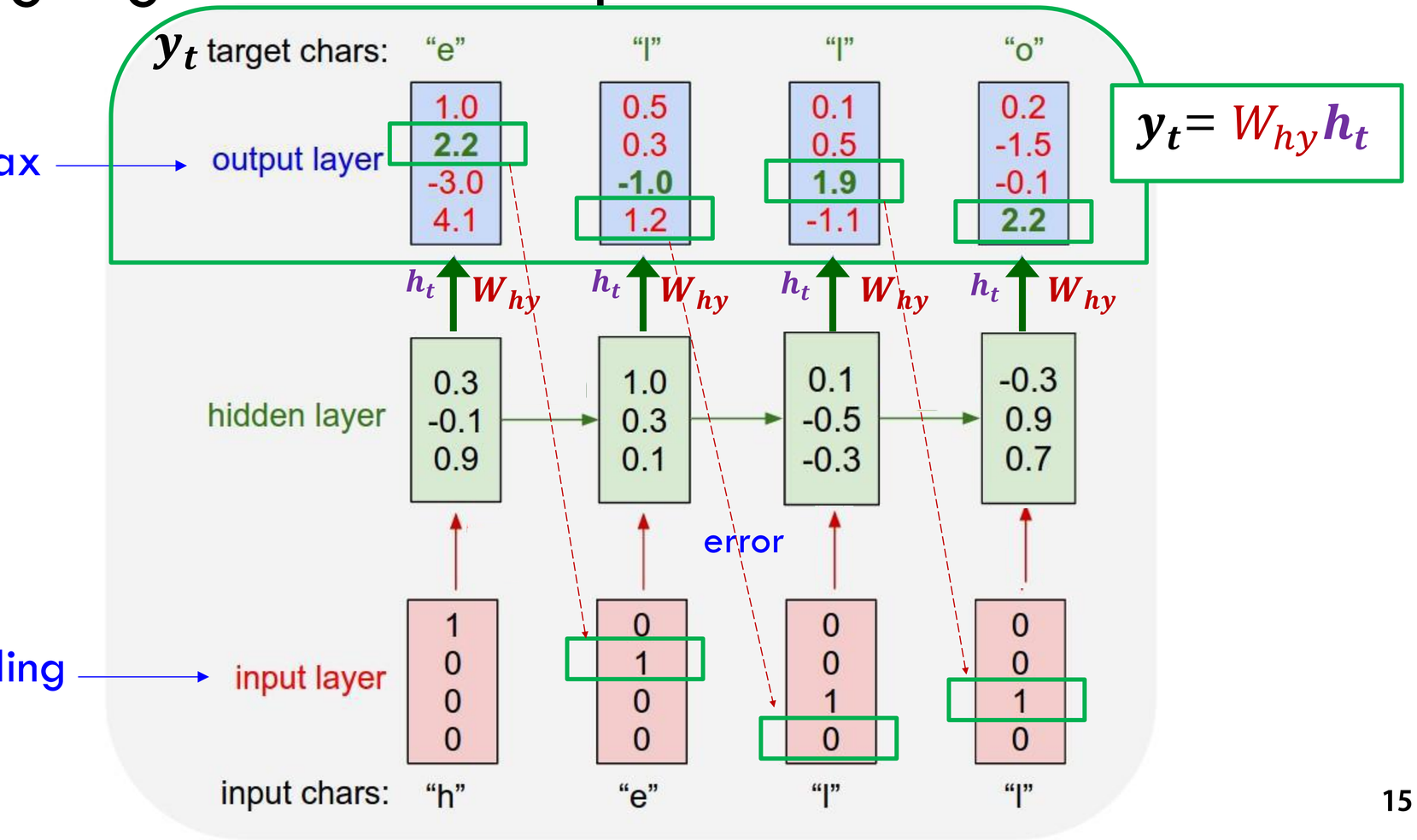
(Vanilla) Recurrent Neural Network

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

- Character-level language model example

Vocabulary :
[h, e, l, o]

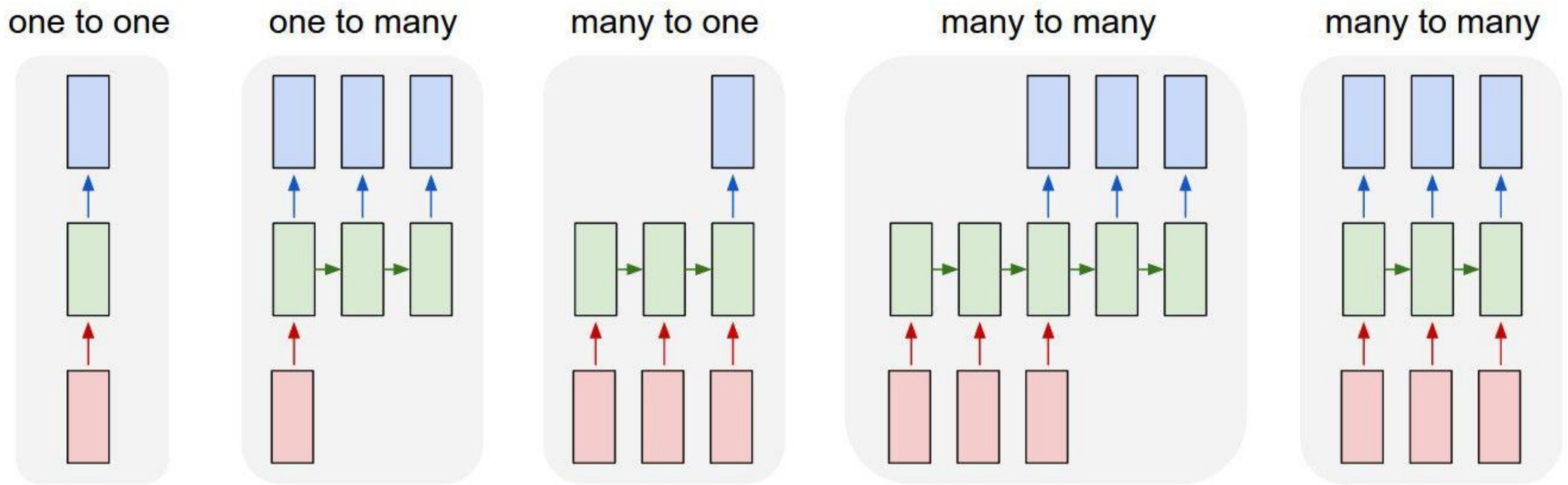
Example training Sequence :
“hello”



RNN applications

- Language Modeling
- Speech Recognition
- Machine Translation
- Conversation Modeling / Question Answering
- Image / Video Captioning
- Image / Music / Dance Generation

Recurrent Networks offer a lot of flexibility:



Vanilla Neural Networks

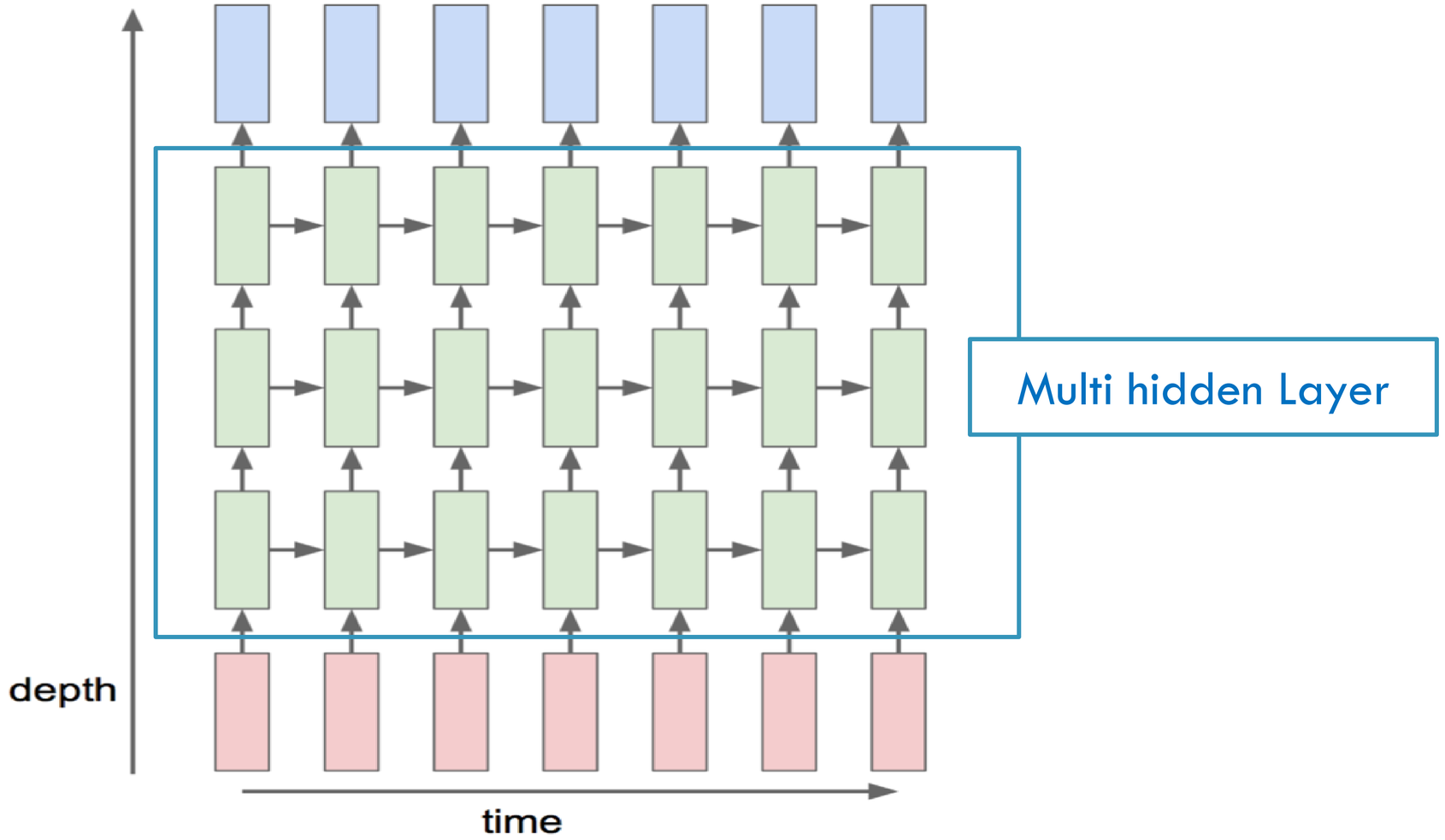
e.g. Image Captioning
image
→ sequence of words

e. g. Sentiment Classification
sequence of words
→ sentiment

e.g. Machine Translation
sequence of words
→ sequence of words

e.g. Video classification on frame level

Multi-Layer RNN



Training RNNs is challenging

- Several advanced models
  - Long Short Term Memory (LSTM)
  - GRU by Cho et al. 2014