

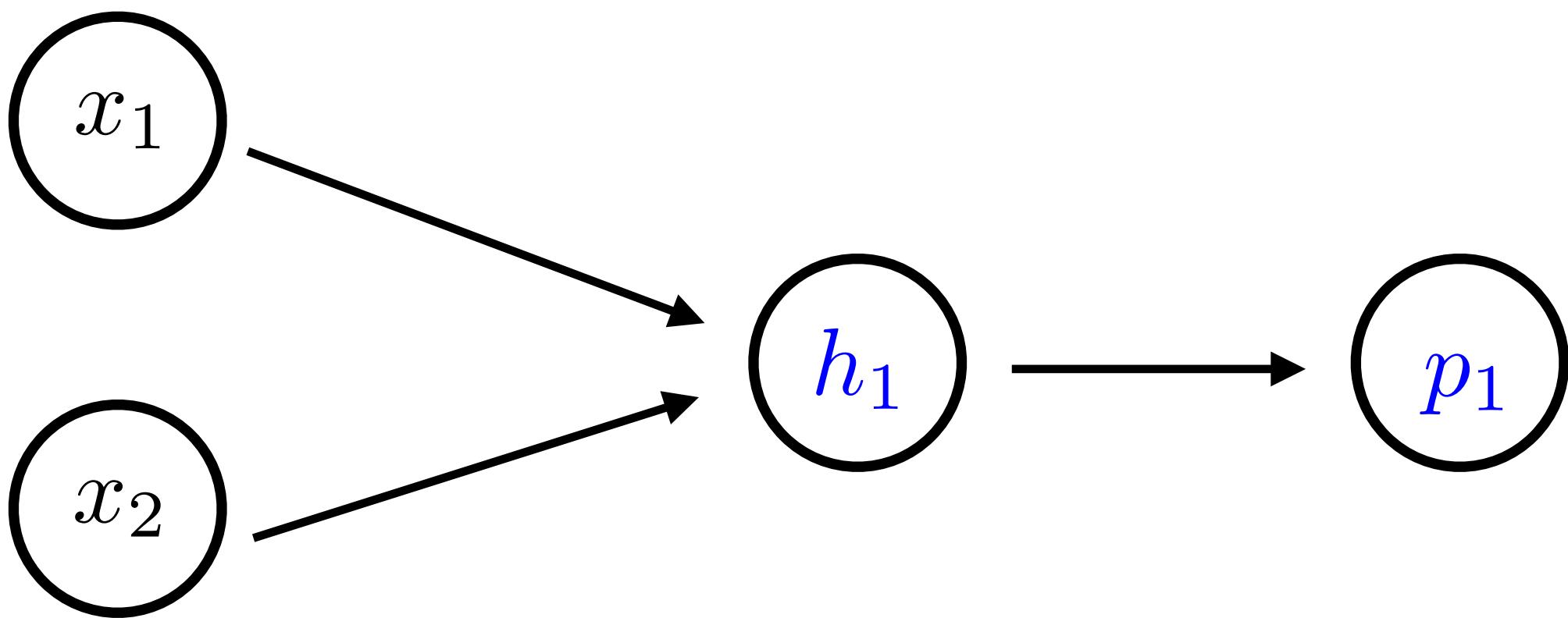
# The Computational Graphs

COMP3361 – Week 3

Lingpeng Kong

Department of Computer Science, The University of Hong Kong  
Some materials from Stanford University CS224n with special thanks!

# Logistic Regression

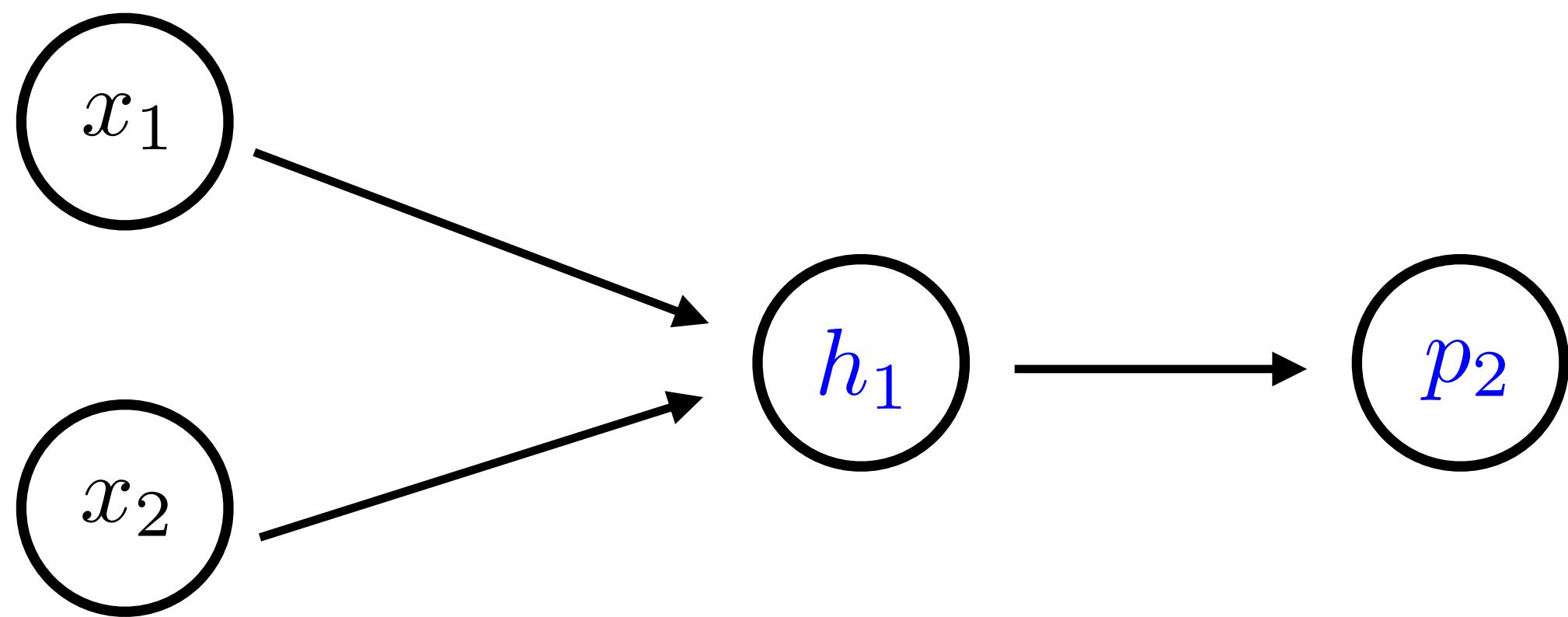


$$p_1 := p(y = 1 \mid x_1, x_2)$$

$$h_1 = w_1 x_1 + w_2 x_2 + b$$

$$p_1 = \frac{1}{1 + \exp(-h_1)}$$

# Logistic Regression



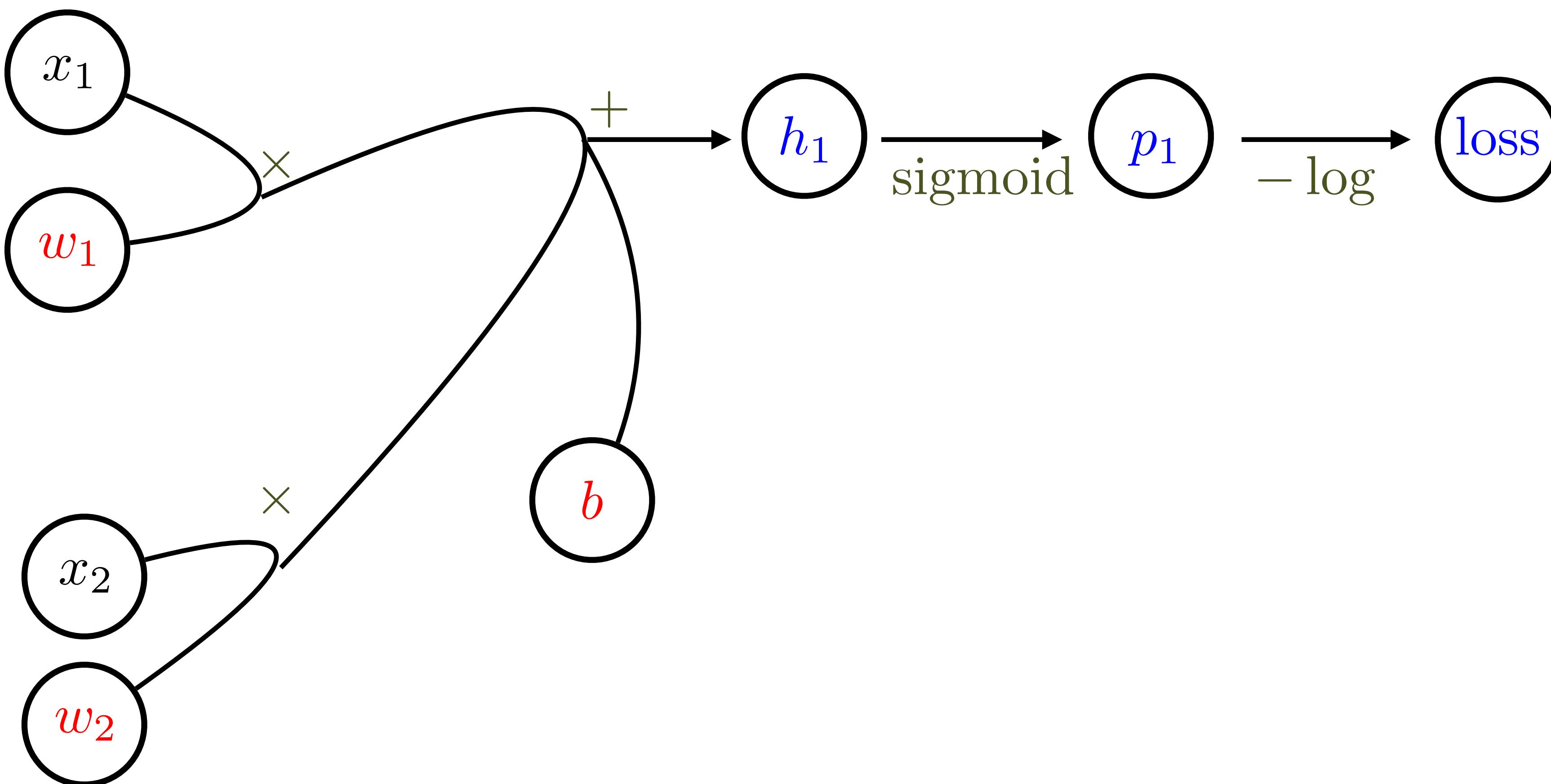
$$p_2 := p(y = 0 \mid x_1, x_2)$$

$$h_1 = w_1 x_1 + w_2 x_2 + b$$

$$p_2 = 1 - \frac{1}{1 + \exp(-h_1)} = \frac{\exp(-h_1)}{1 + \exp(-h_1)}$$

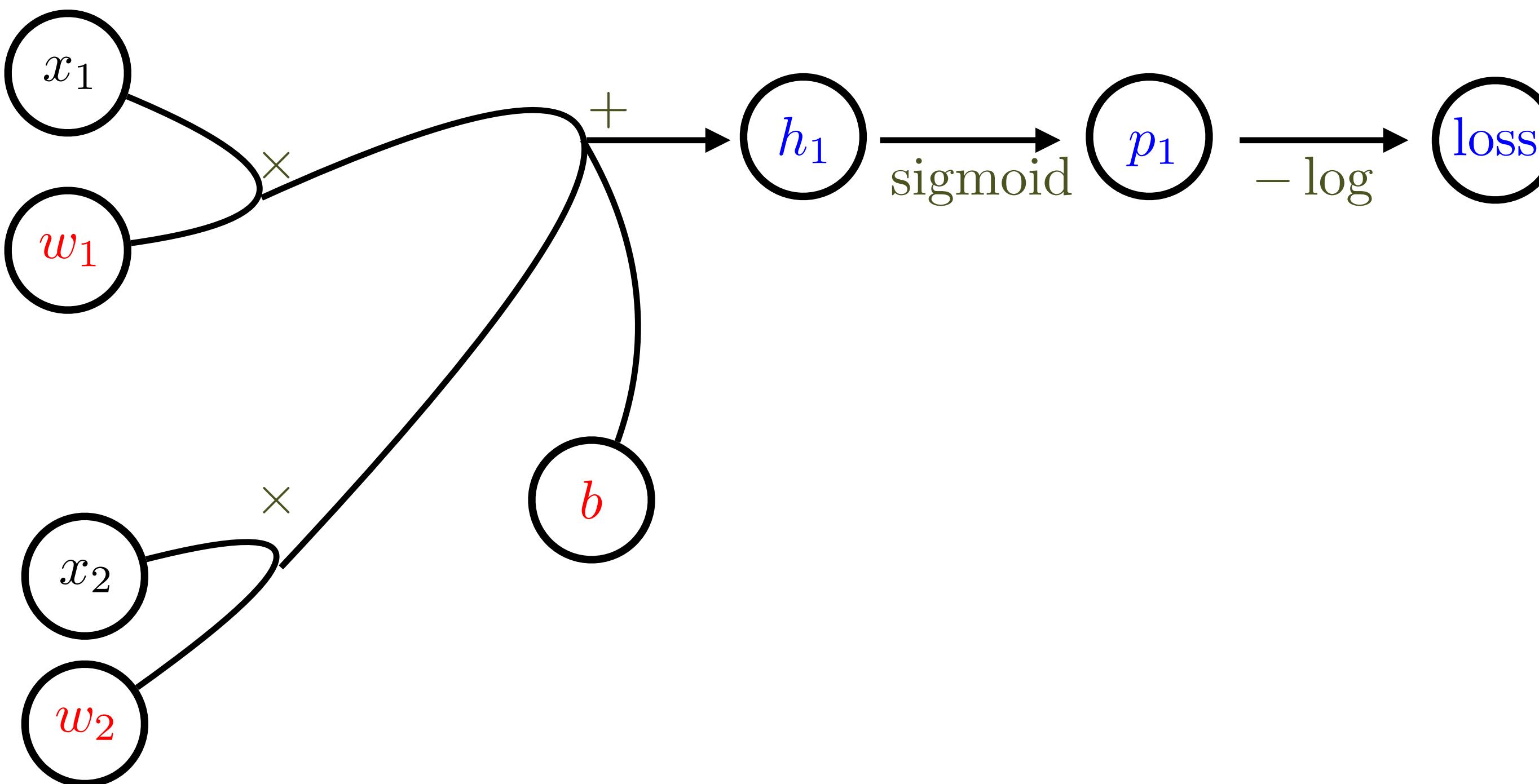
# Loss Function

case  $y = 1$  :



# Computational Graphs

Input	$x_1$	$x_2$	
Parameter	$w_1$	$w_2$	
Expression	$h_1$	$p_1$	loss
Operation	$\times$	$+$	sigmoid – log

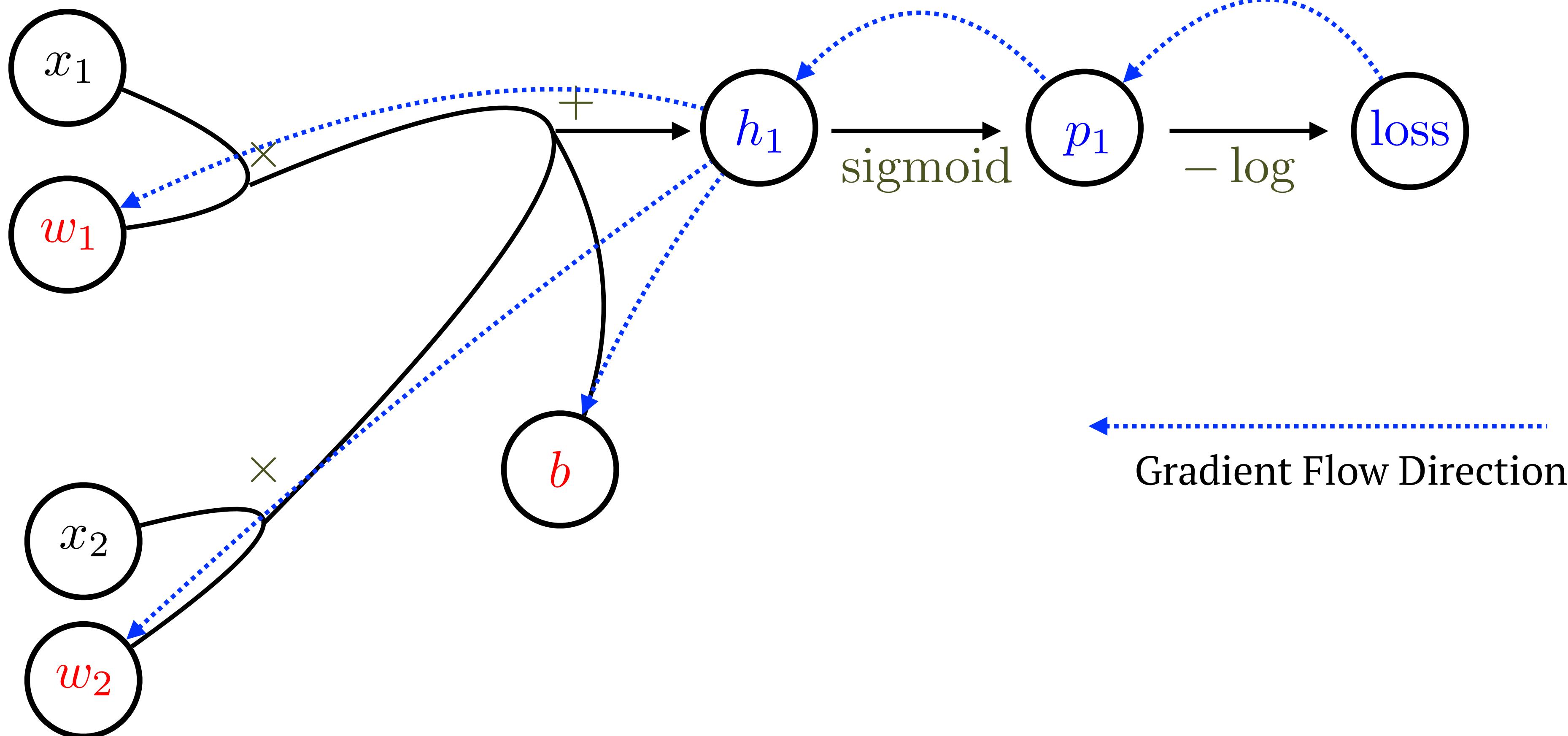


# How to minimize? (Automatic Differentiation)

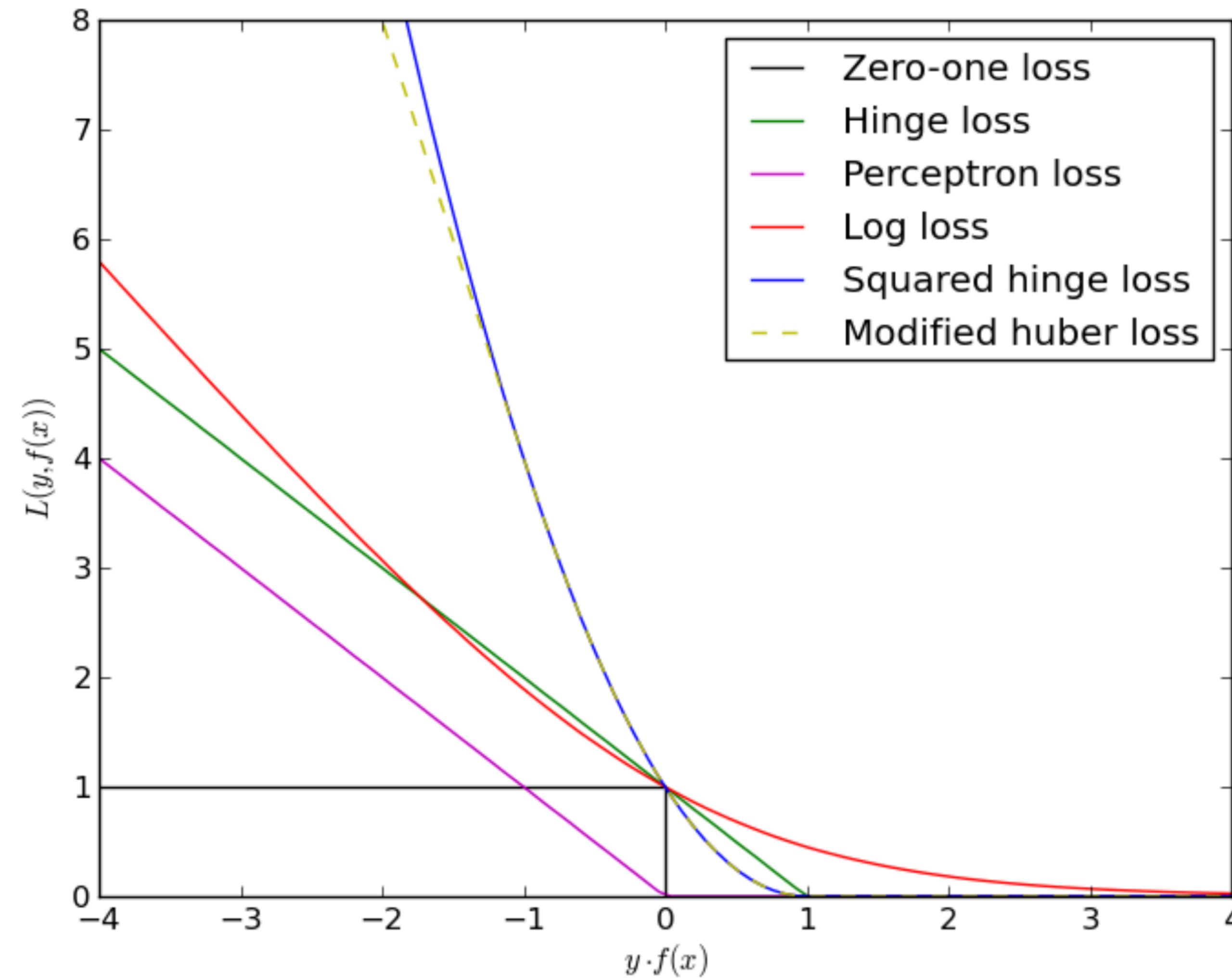
$$h = f(z)$$
$$\text{Chain rule} \quad \frac{\partial s}{\partial z} = \frac{\partial s}{\partial h} \frac{\partial h}{\partial z}$$

Downstream Gradient      Local Gradient      Upstream Gradient

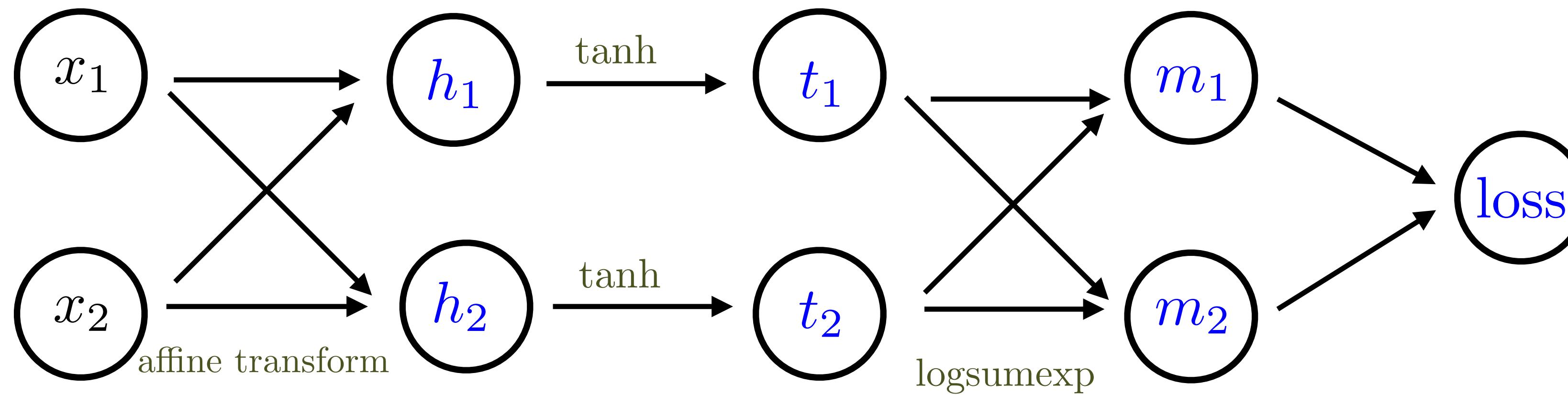
# How to minimize? (Automatic Differentiation)



# Other Loss Function?

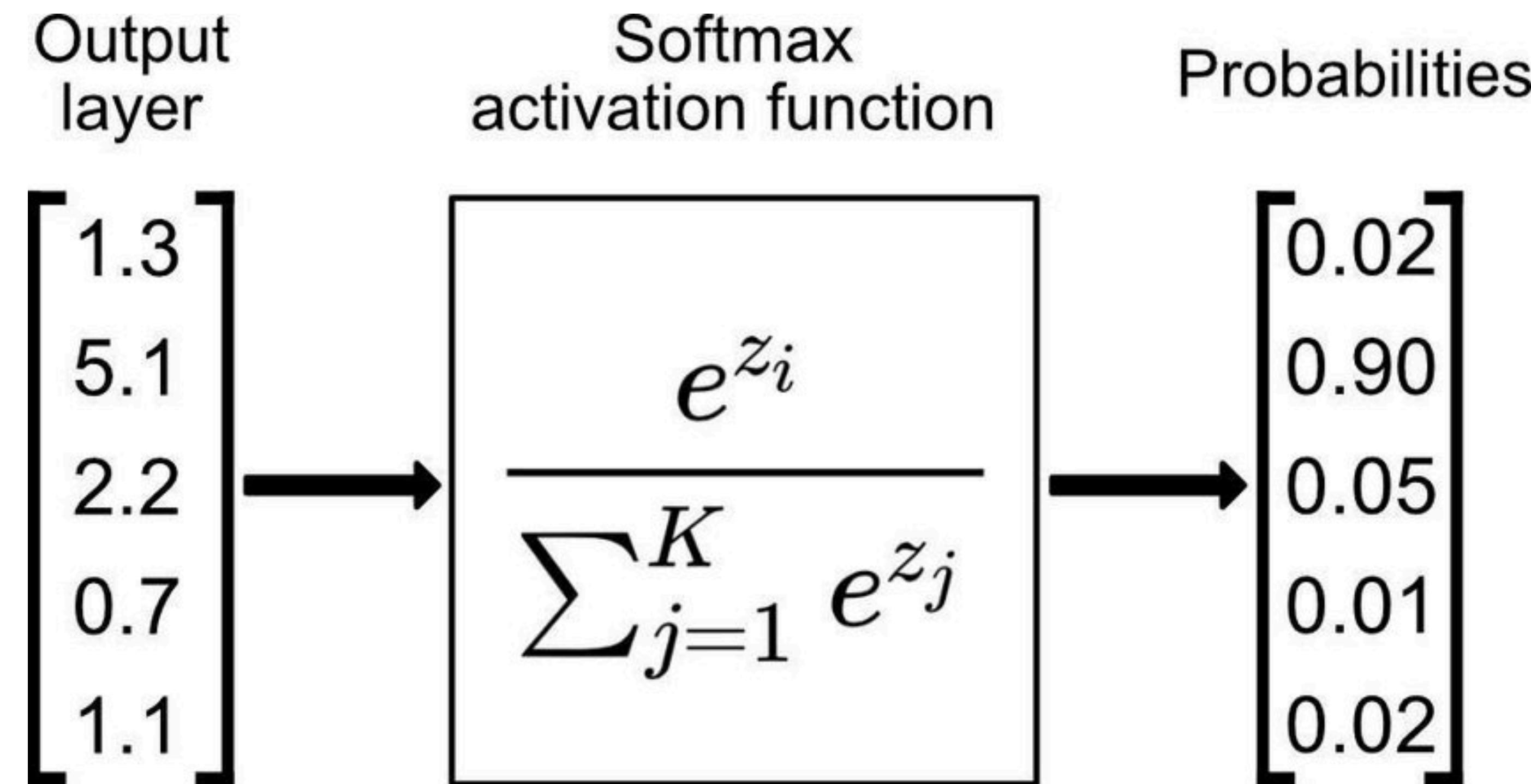


# “Deeper” Neural Network

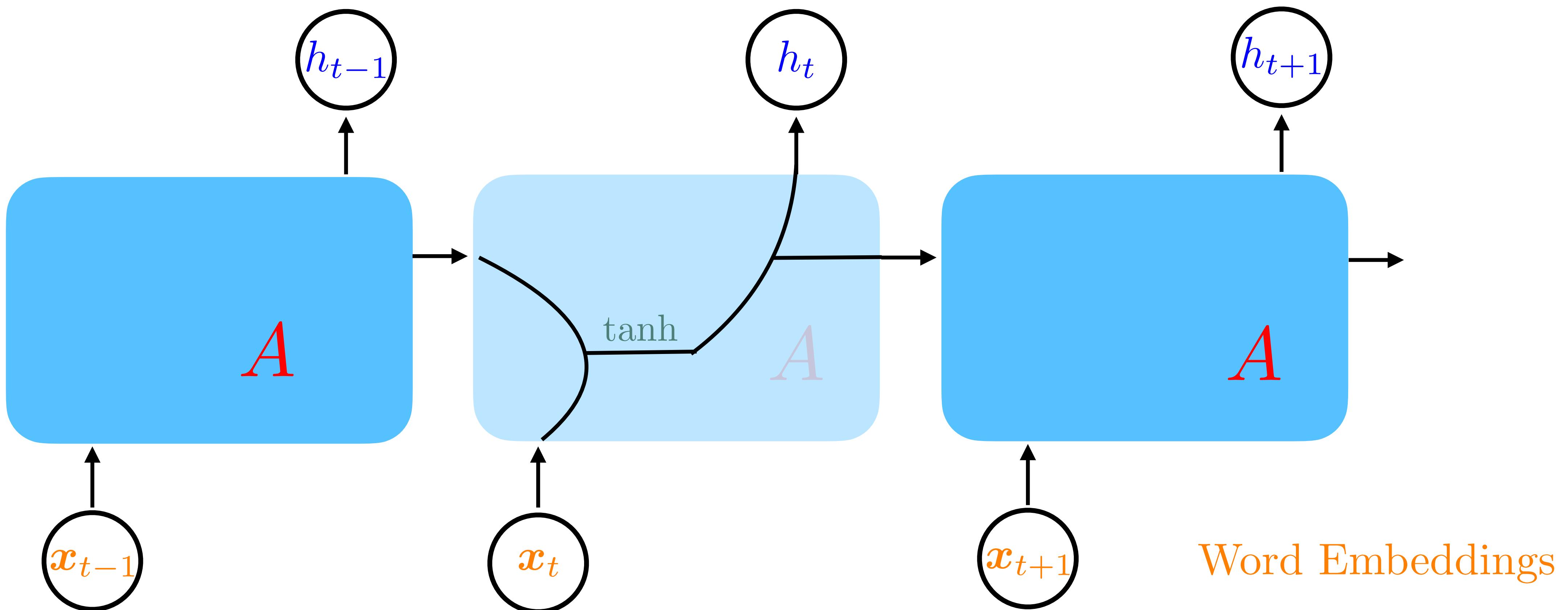


$$m_1 = \log\left(\frac{\exp(t_1)}{\exp(t_1) + \exp(t_2)}\right)$$

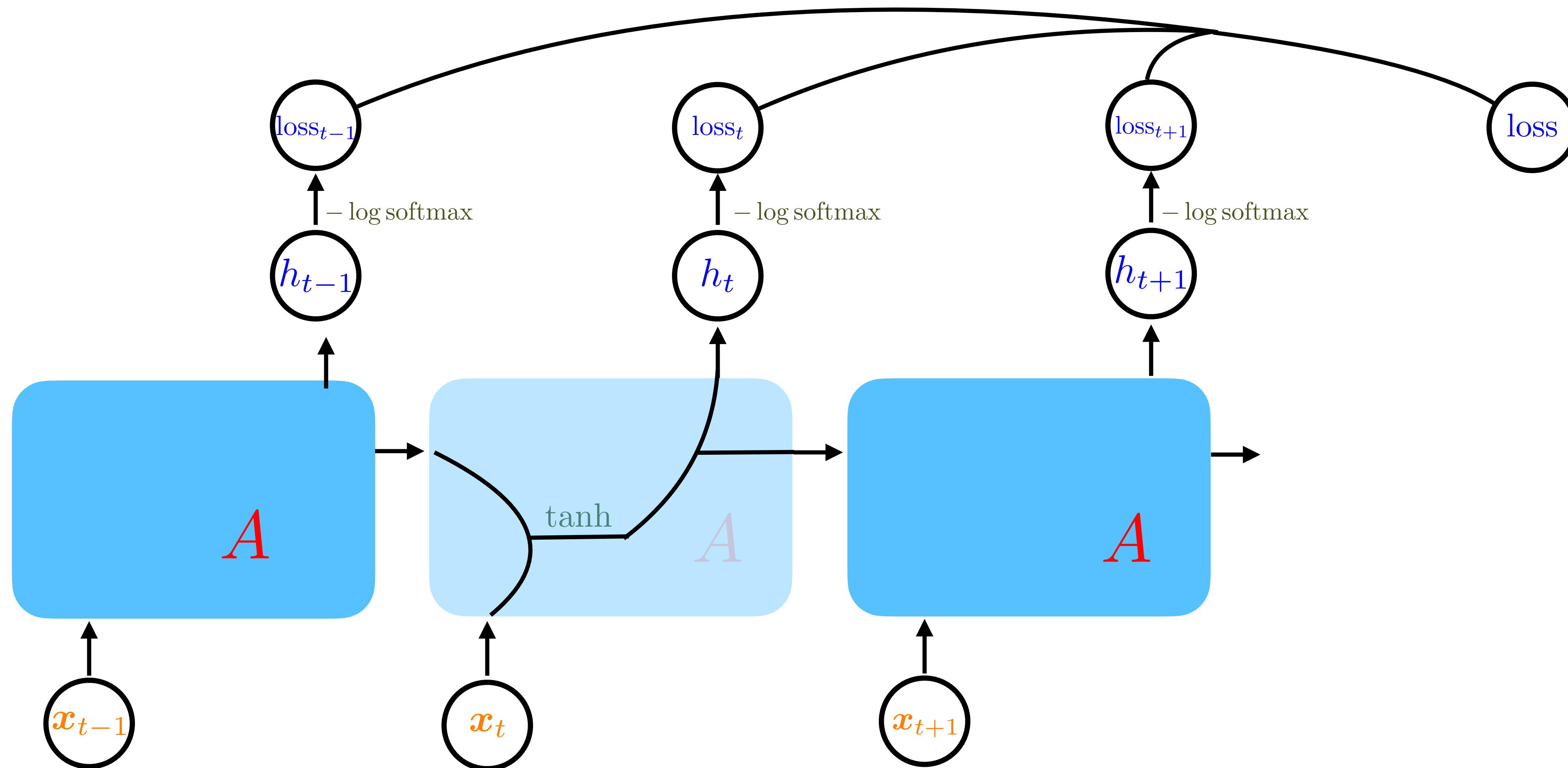
# Softmax Function



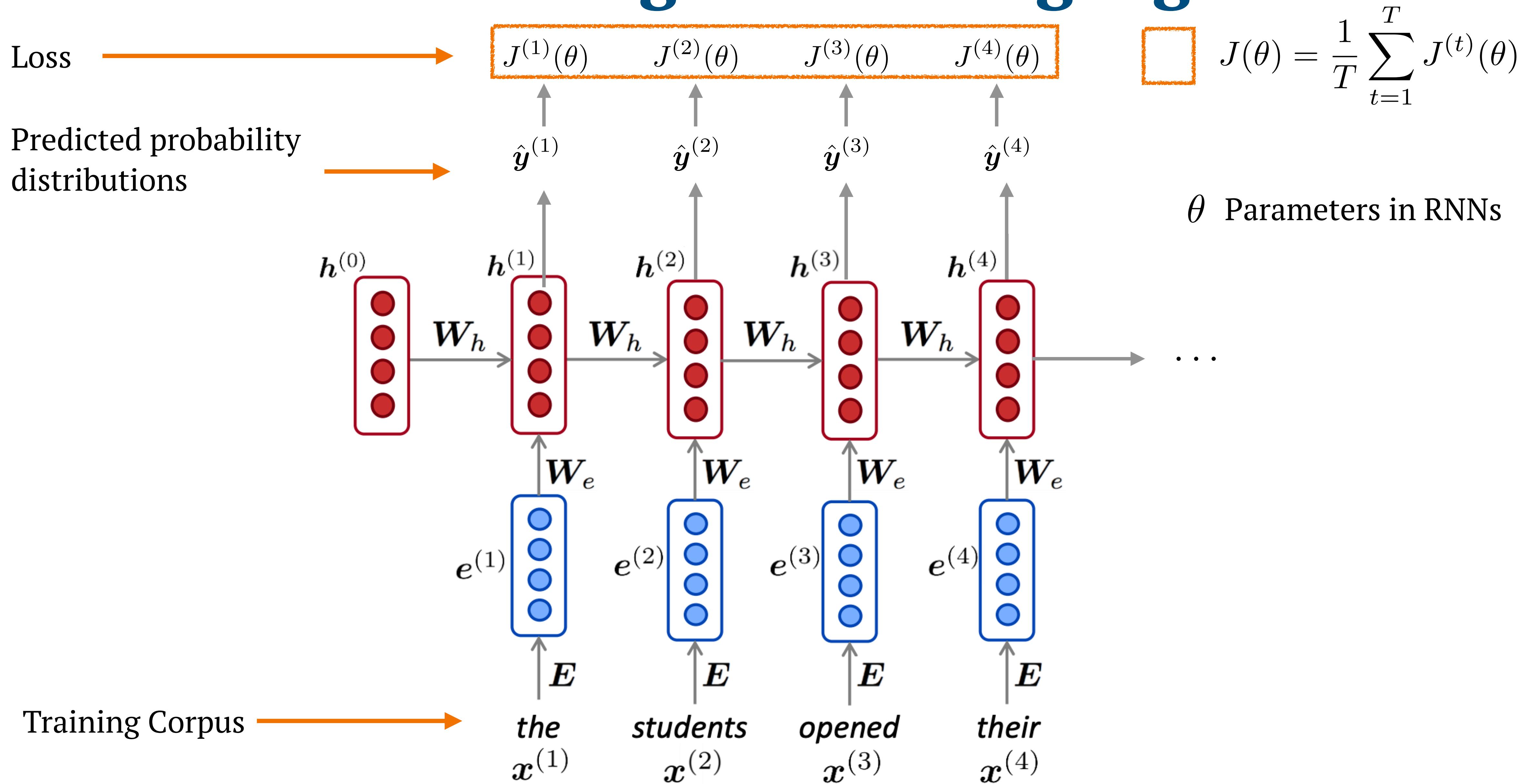
# Recurrent Neural Network



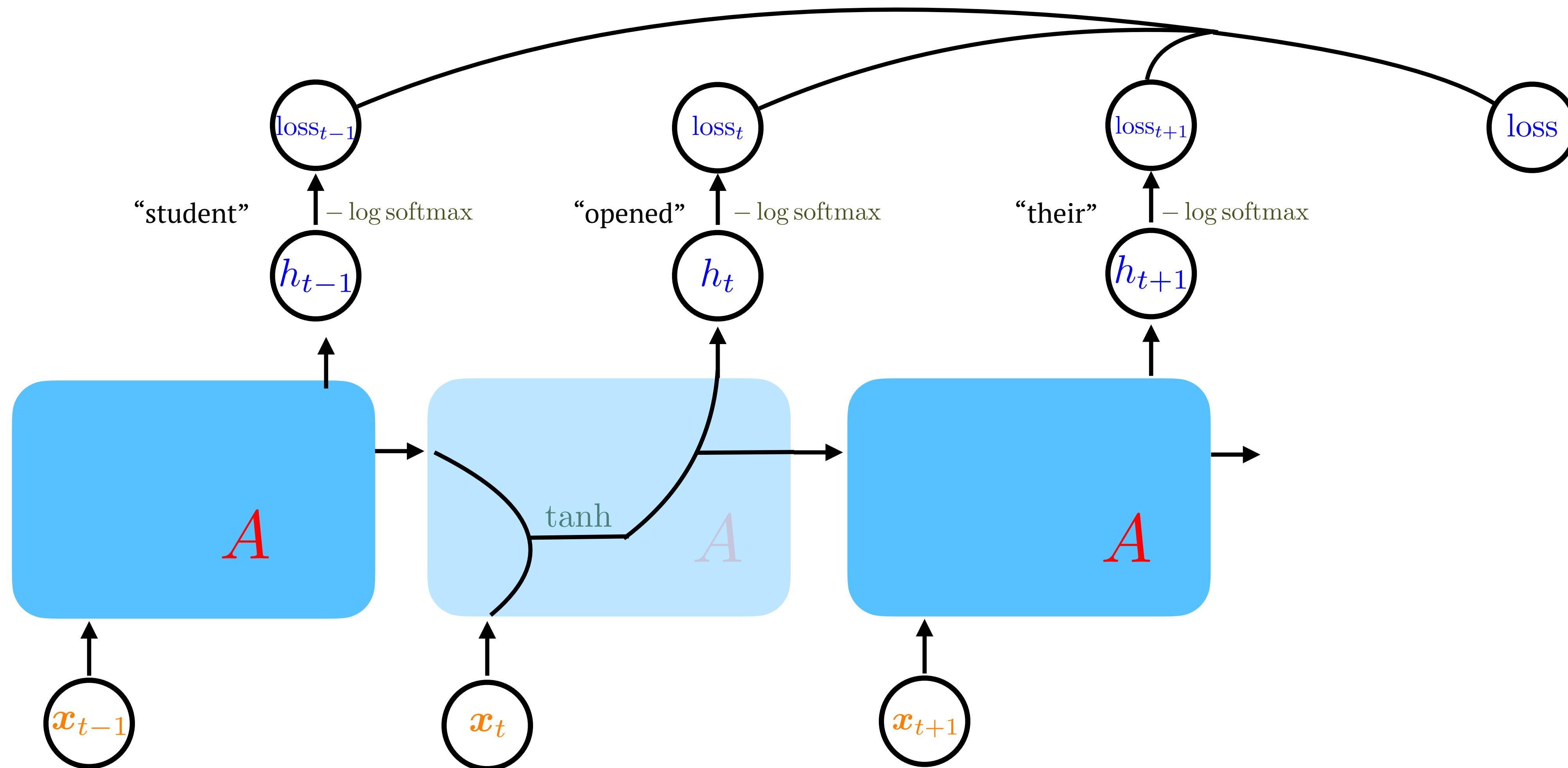
# Recurrent Neural Network (Language Model)



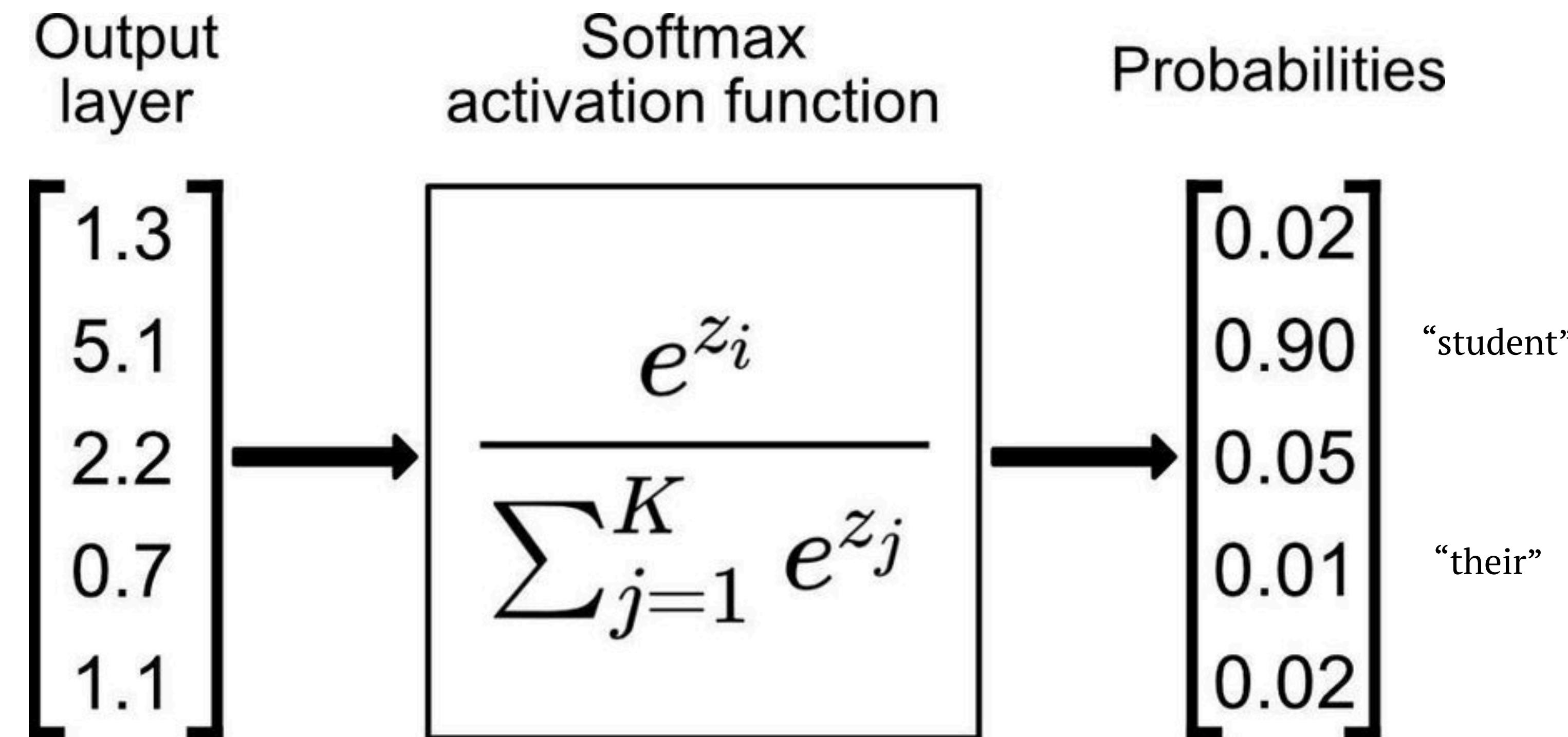
# Flashback – Training a RNN Language Model



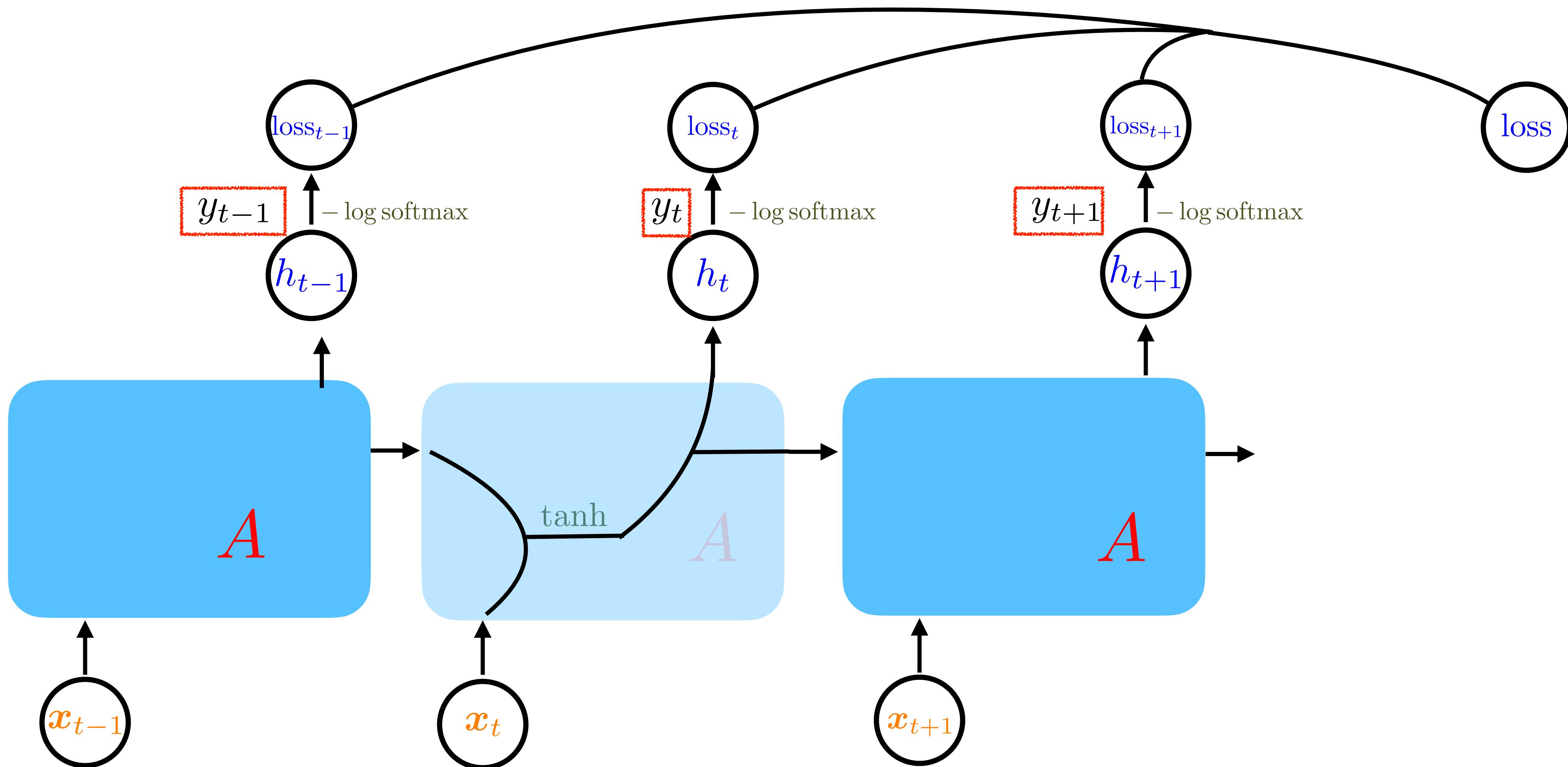
# Recurrent Neural Network (Language Model)



# Softmax Function



# RNNs for Tagging



# Part-of-Speech Tagging

## INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

## OUTPUT:

Profits/N soared/V at/P Boeing/N Co./N ,/, easily/ADV topping/V forecasts/N on/P Wall/N Street/N ,/, as/P their/POSS CEO/N Alan/N Mulally/N announced/V first/ADJ quarter/N results/N ./.

N = Noun

V = Verb

P = Preposition

Adv = Adverb

Adj = Adjective

...

# Named Entity Recognition (NER)

**INPUT:** Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

**OUTPUT:** Profits soared at [Company Boeing Co.], easily topping forecasts on [Location Wall Street], as their CEO [Person Alan Mulally] announced first quarter results.

# Named Entity Recognition (NER)

## INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

## OUTPUT:

Profits/NA soared/NA at/NA Boeing/SC Co./CC ,/NA easily/NA  
topping/NA forecasts/NA on/NA Wall/SL Street/CL ,/NA as/NA  
their/NA CEO/NA Alan/SP Mulally/CP announced/NA first/NA  
quarter/NA results/NA ./NA

NA = No entity

SC = Start Company

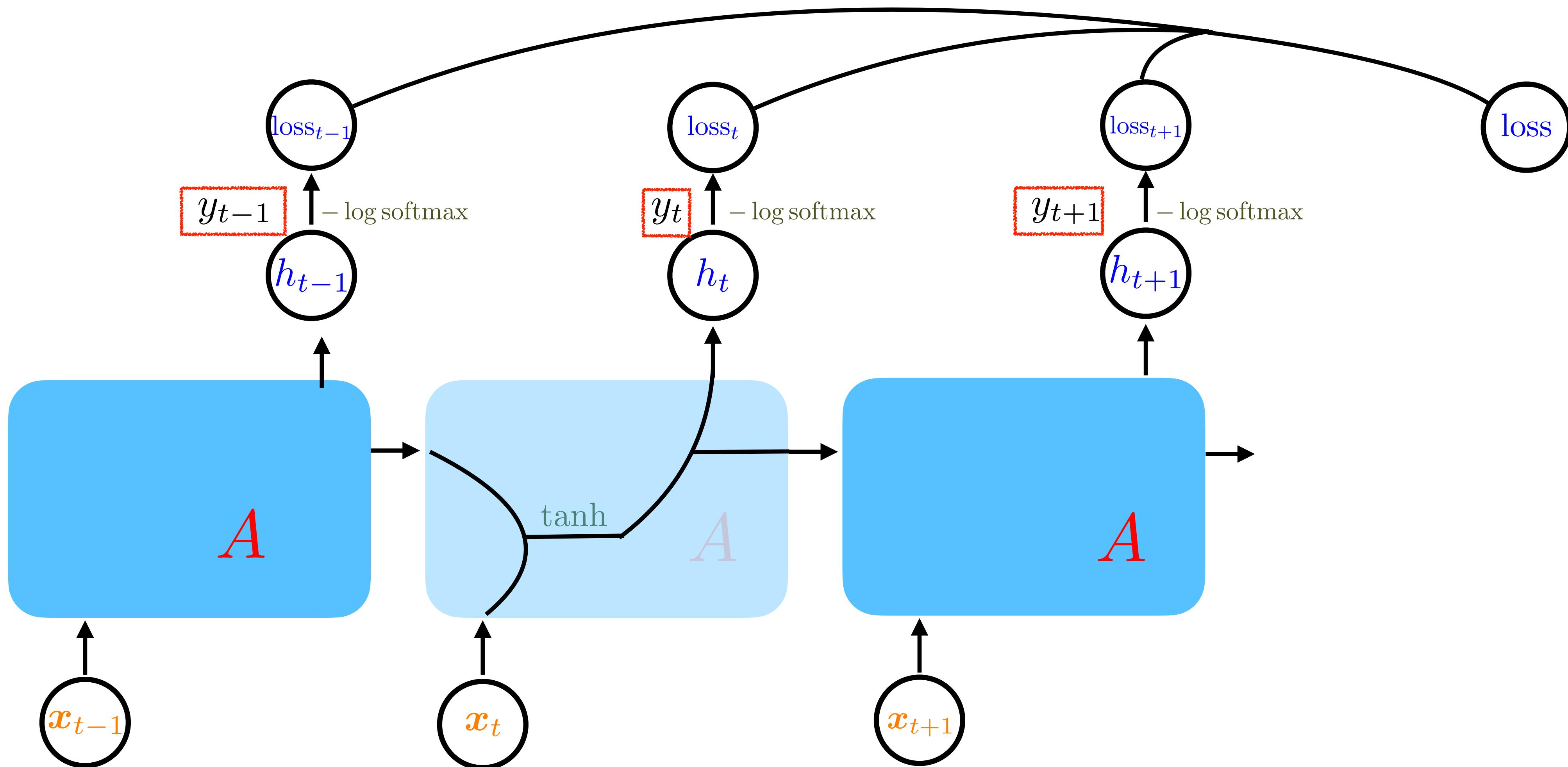
CC = Continue Company

SL = Start Location

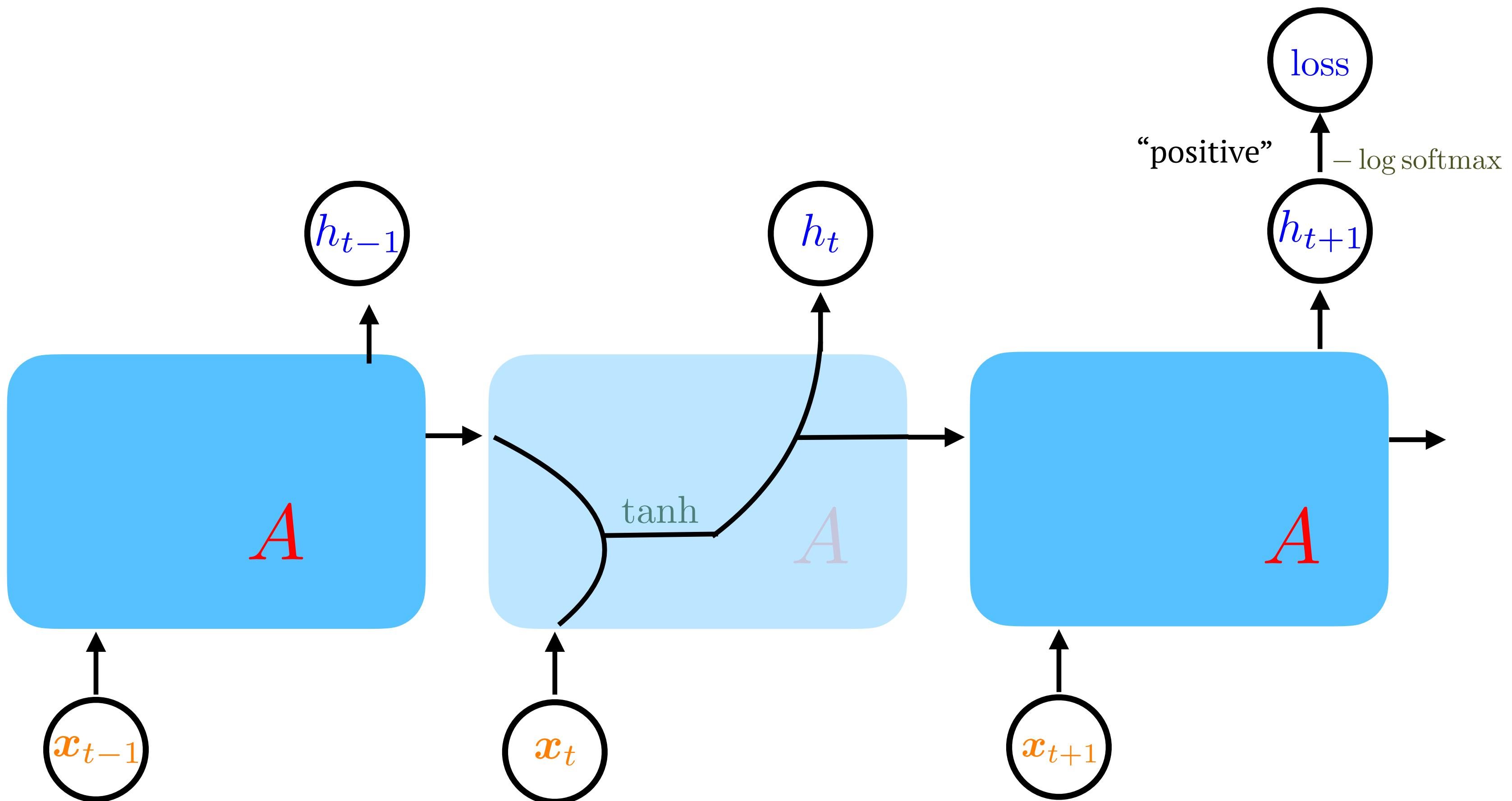
CL = Continue Location

...

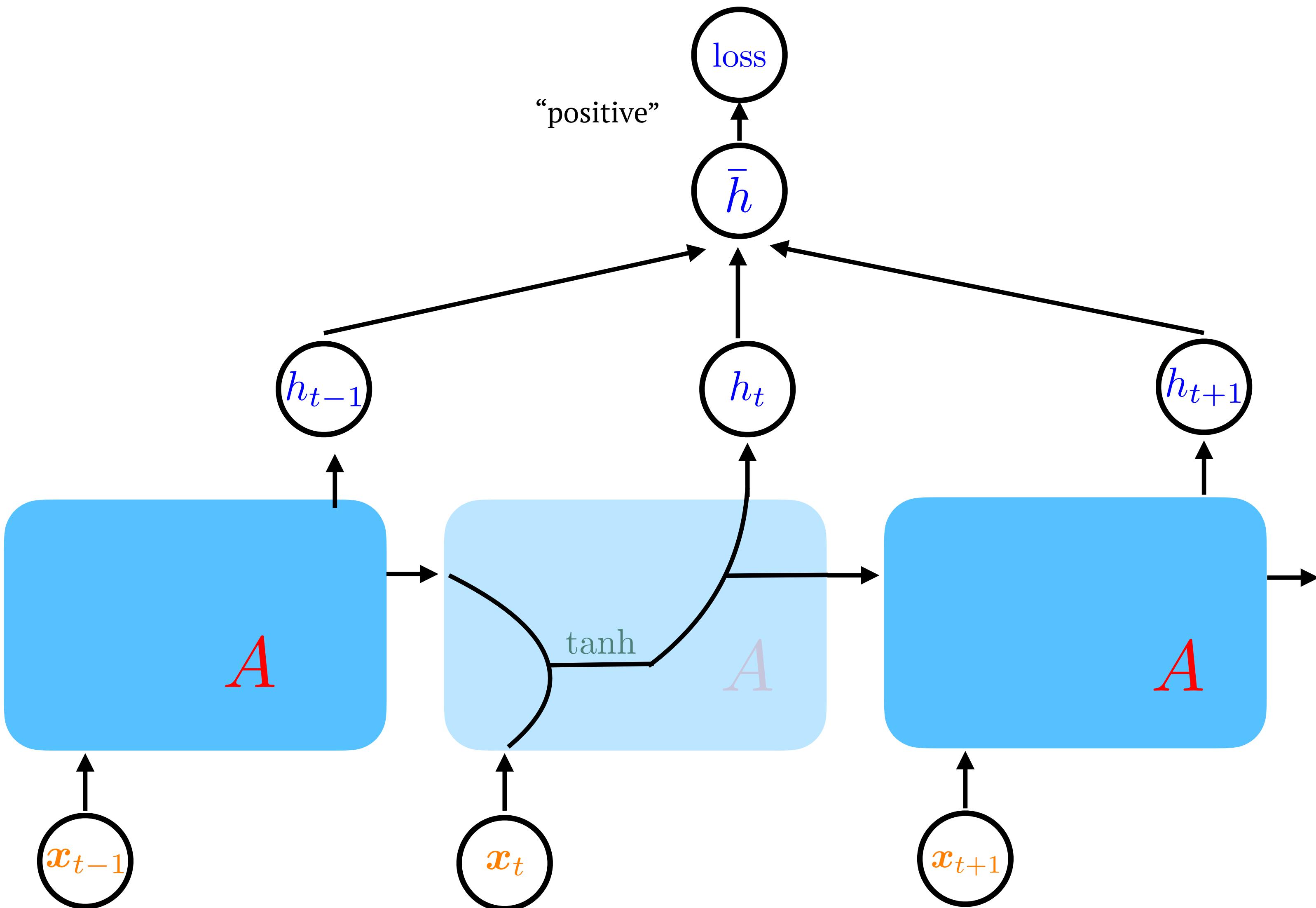
# RNNs for Tagging



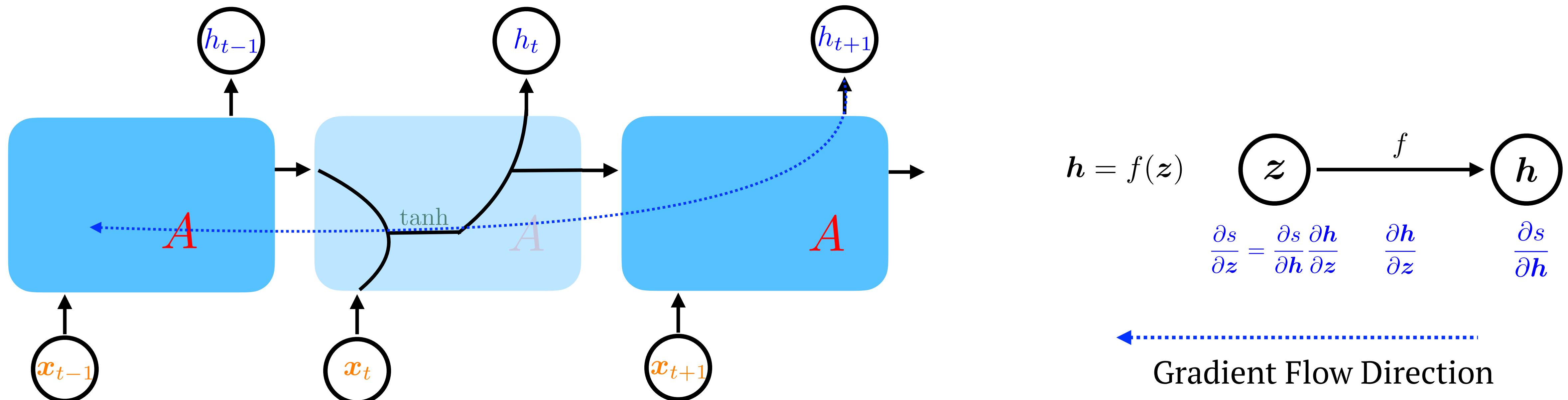
# RNNs for Sentence Classification



# RNNs for Sentence Classification



# Vanishing Gradient in RNNs



In general, the longer the path, the smaller the gradient signal.

# Long Short-Term Memory (LSTMs)

