Neural Network and Deep Learning

Recurrent Neural Network

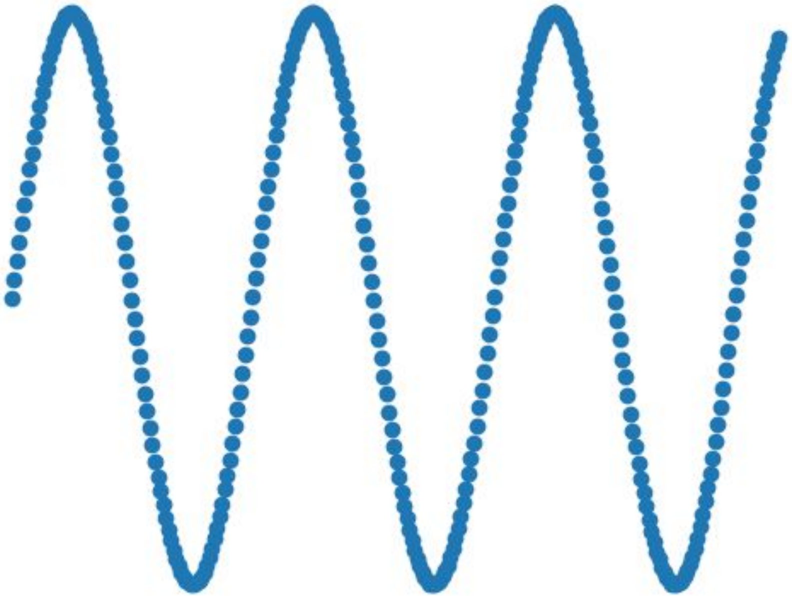
YANIKA KONGSOROT

Sequential Data

Sequential data refers to data where the order of the elements matters.

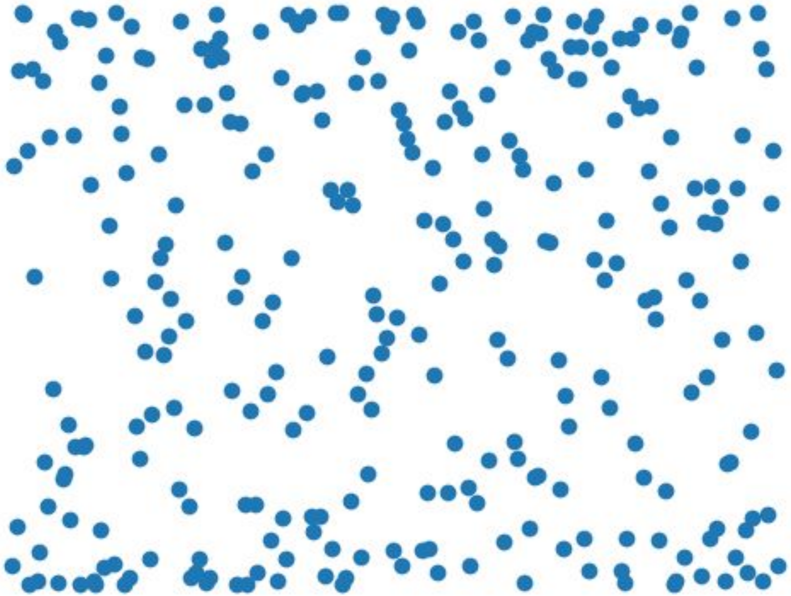
- **Time series data**: Stock prices, temperature readings over time, where the order of the data points indicates the progression of time.
- **Natural language**: Words or sentences, where the order determines the meaning (e.g., “The cat sat” is different from “Sat the cat”).
- **Video data**: A series of frames, where each frame is connected to the previous and next to form a coherent motion.

Sequential Data



Sequential Data

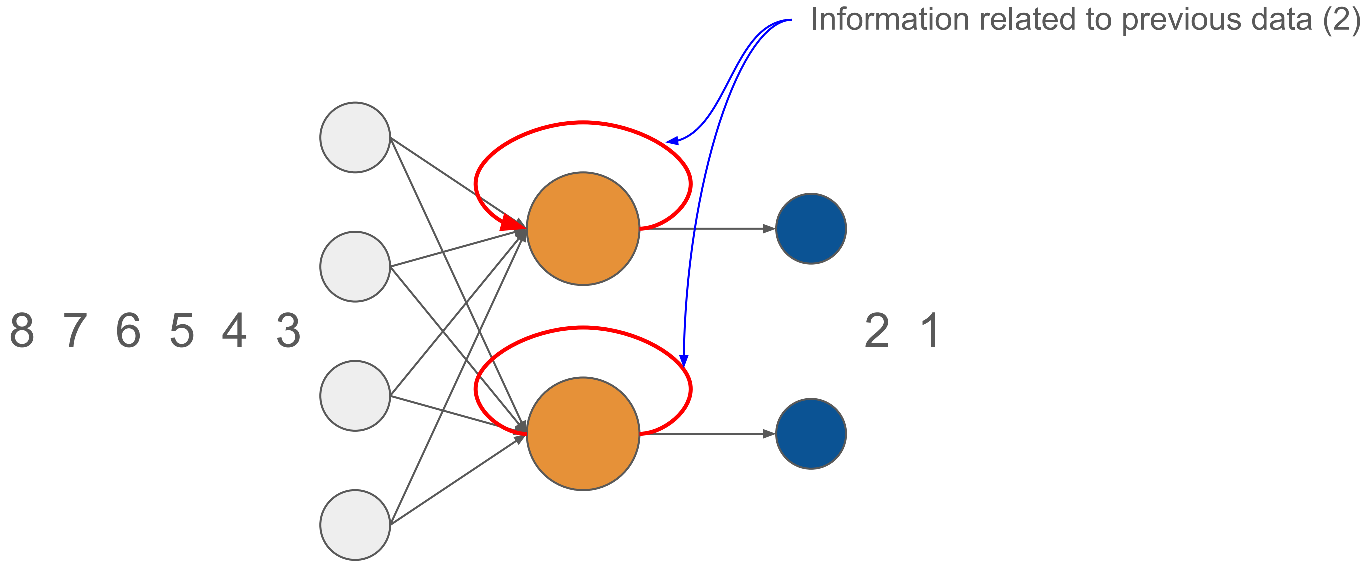
VS



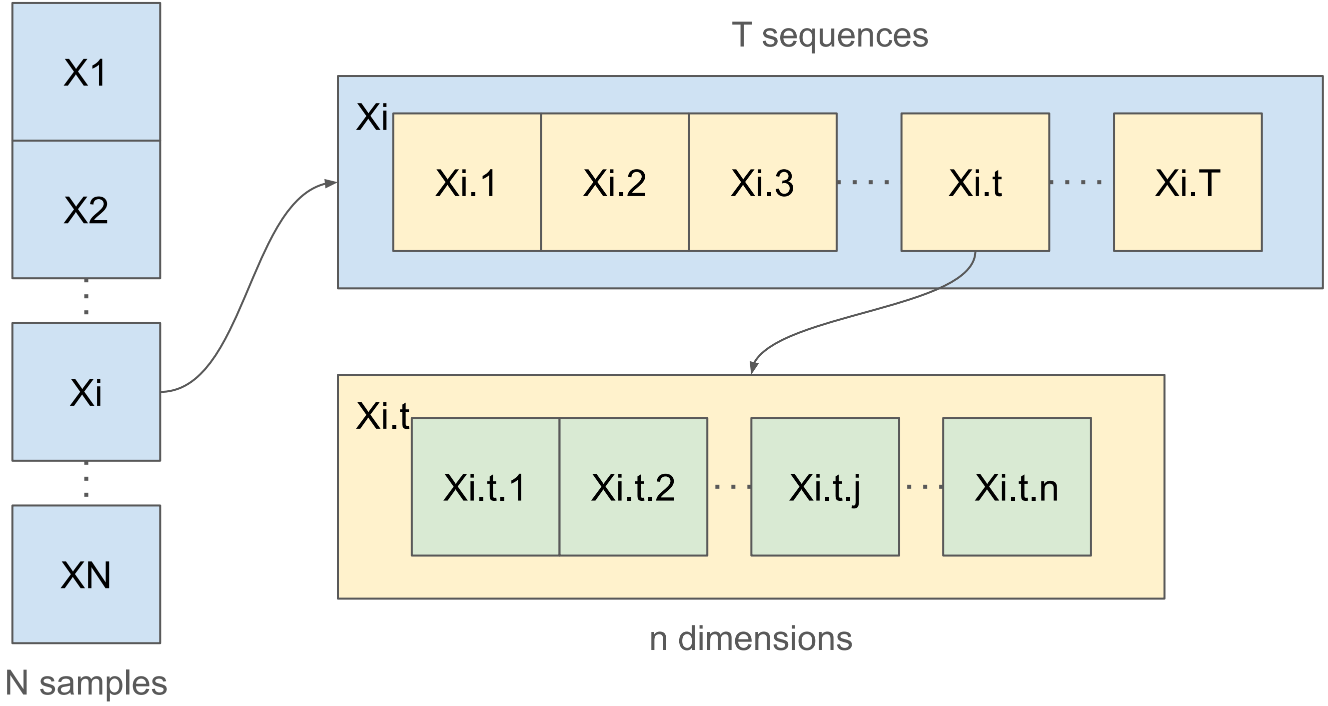
Shuffled Data

Recurrent Neural Network

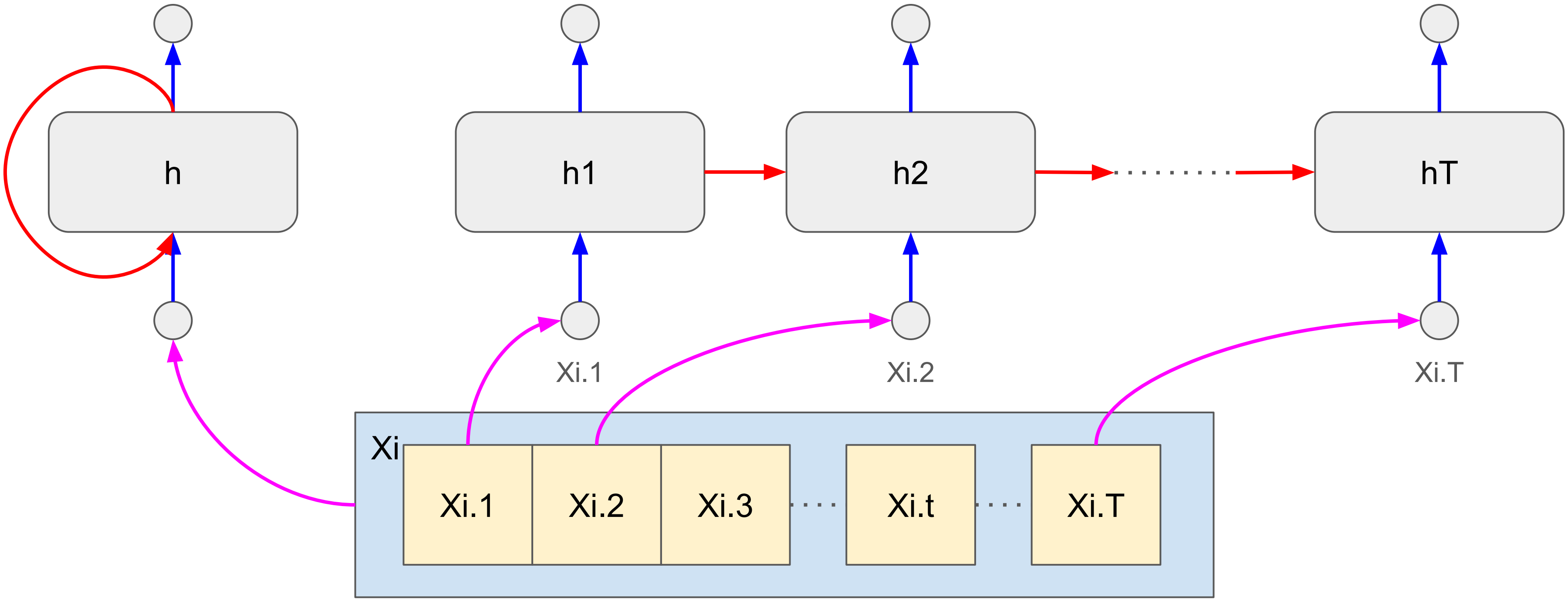
Recurrent Neural Network



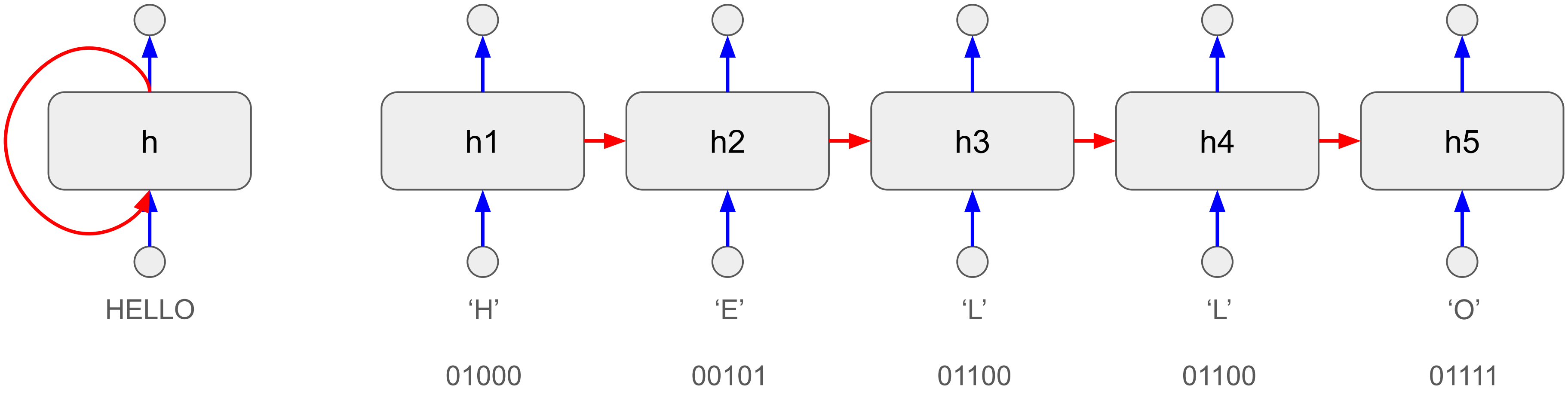
Recurrent Neural Network – Unfolding



Recurrent Neural Network – Unfolding

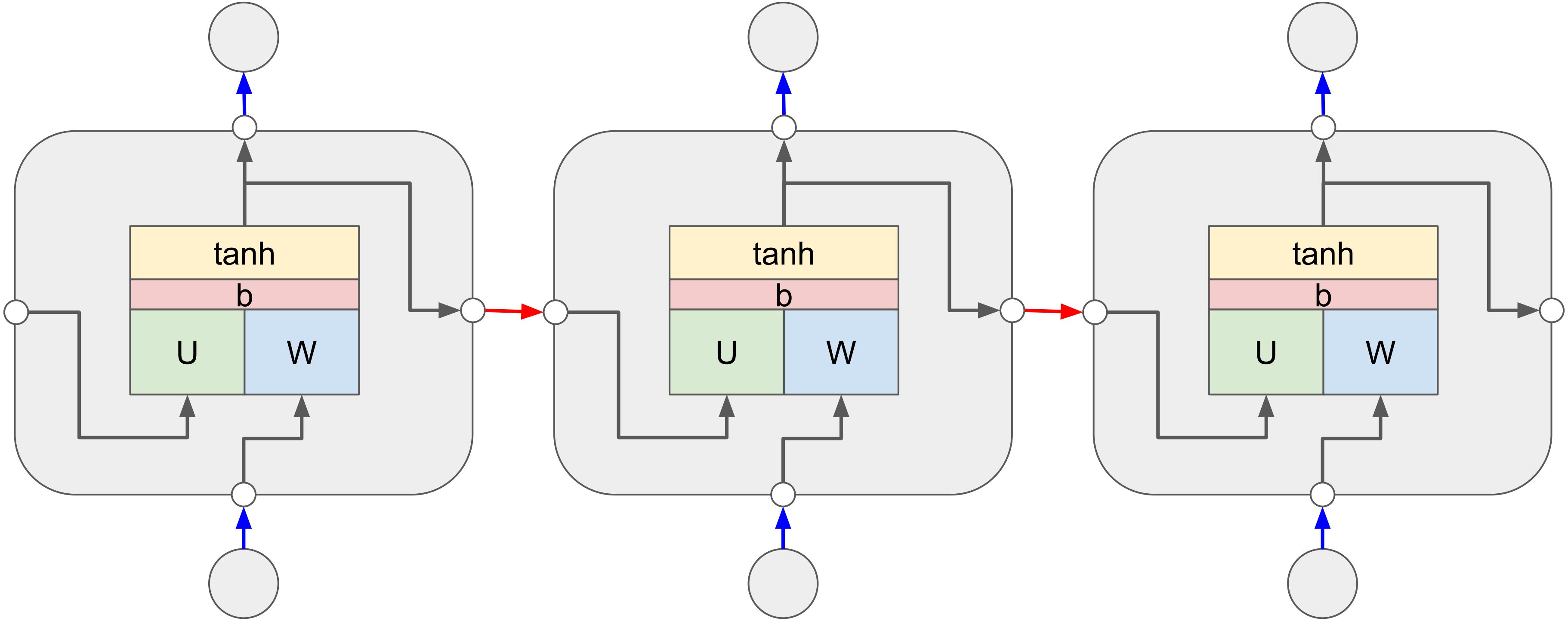


Recurrent Neural Network



Sequence length = 5
Dimension = 5

Recurrent Neural Network



Recurrent Neural Network – Example

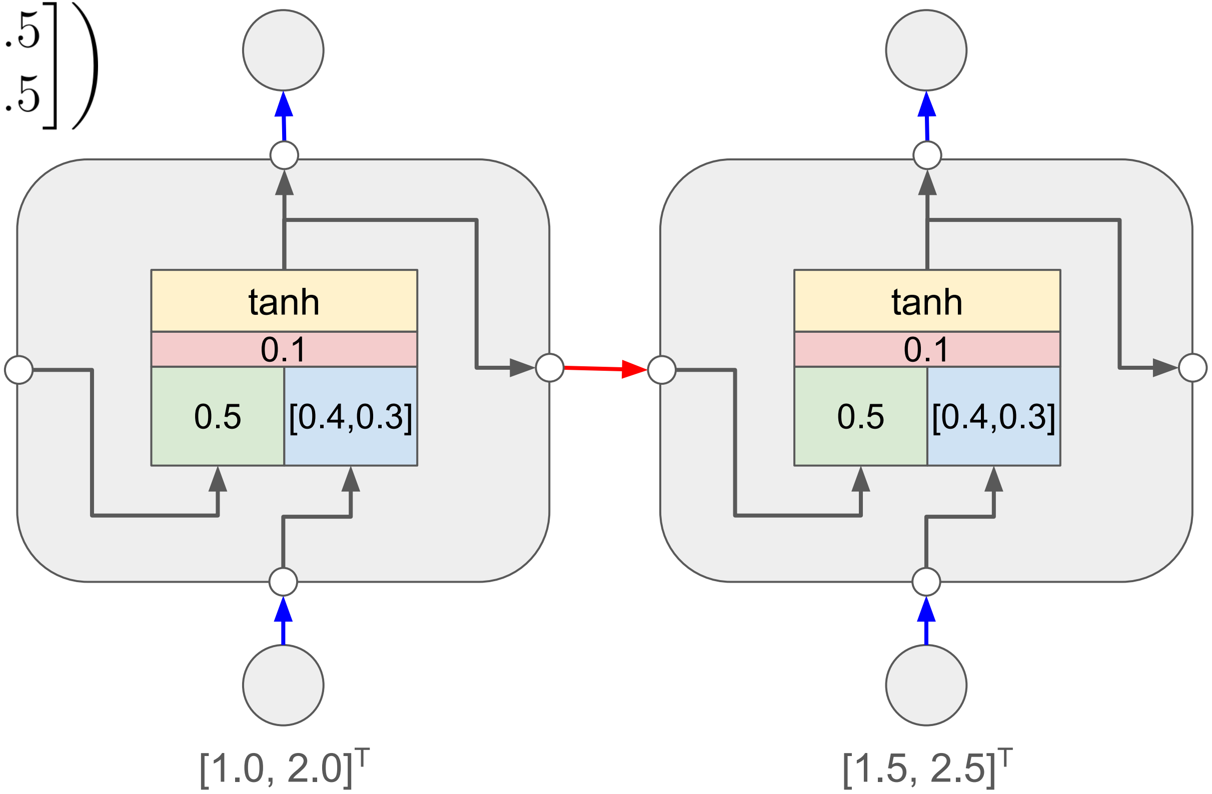
$$X = (x_1, x_2) = \left(\begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix}, \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} \right)$$

$$W = [0.4, 0.3]$$

$$U = 0.5$$

$$b = 0.1$$

$$h_0 = 0$$



Recurrent Neural Network – Example

$$X = (x_1, x_2) = \left(\begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix}, \begin{bmatrix} 1.5 \\ 2.5 \end{bmatrix} \right)$$

$$W = [0.4, 0.3]$$

$$U = 0.5$$

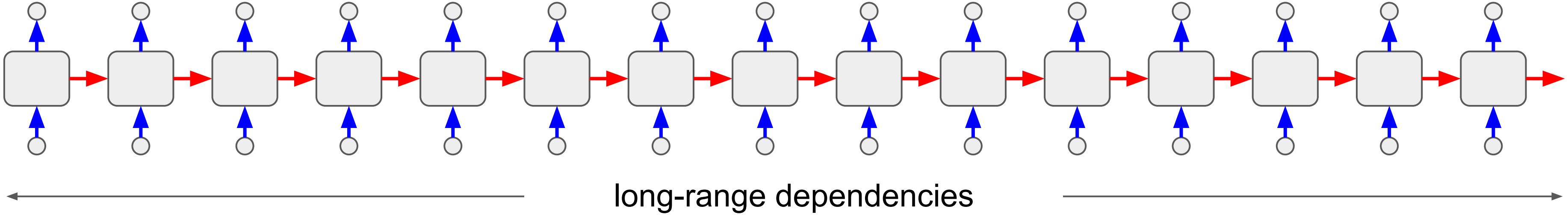
$$b = 0.1$$

$$h_0 = 0$$

$$\begin{aligned} h_1 &= \tanh(Uh_0 + Wx_1 + b) \\ &= \tanh(0.5 \cdot 0 + [0.4, 0.3] \cdot [1.0, 2.0]^\top + 0.1) \\ &= \tanh(1.1) \approx 0.8005 \end{aligned}$$

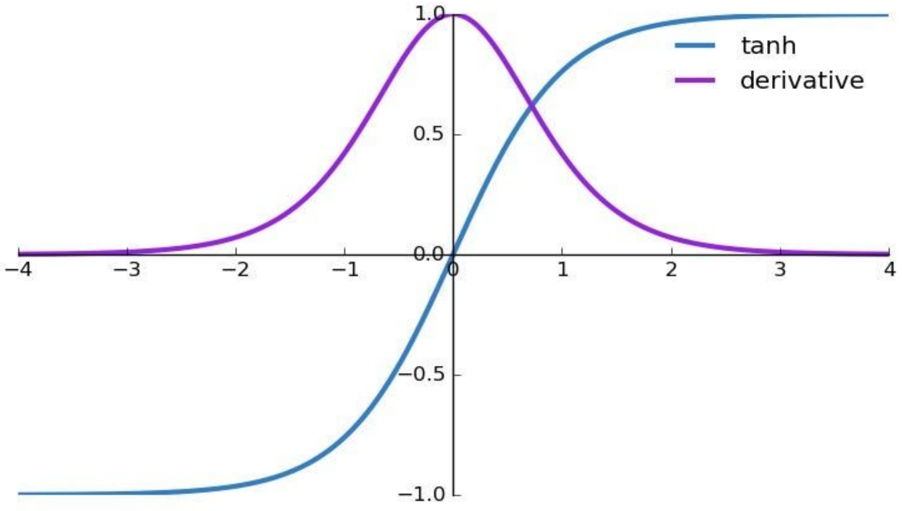
$$\begin{aligned} h_2 &= \tanh(Uh_1 + Wx_2 + b) \\ &= \tanh(0.5 \cdot 0.8005 + [0.4, 0.3] \cdot [1.5, 2.5]^\top + 0.1) \\ &= \tanh(1.85) \approx 0.9516 \end{aligned}$$

Recurrent Neural Network – Disadvantage



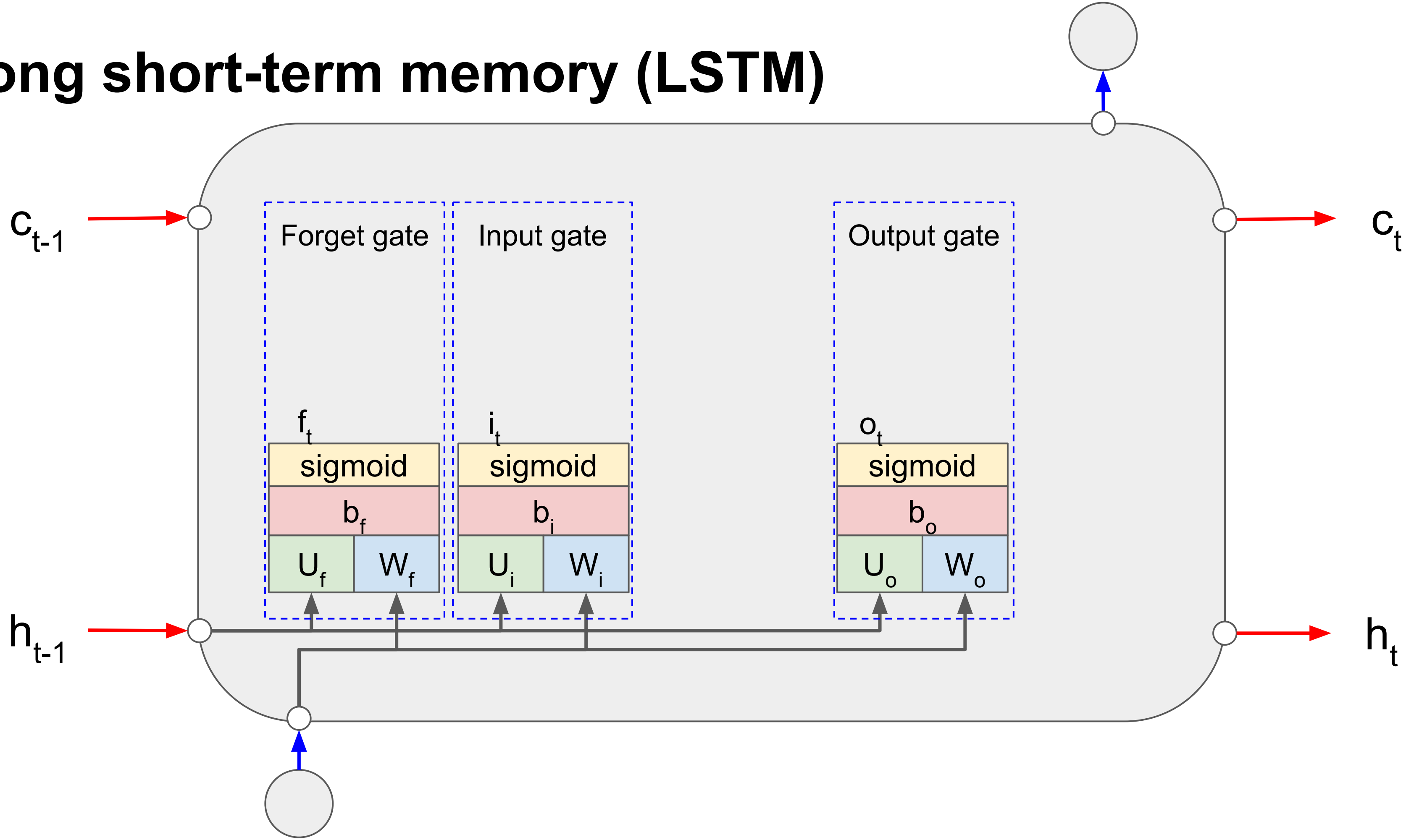
Definition: Long-range dependencies refer to relationships or patterns in data that span over many time steps.

The vanishing gradient problem is particularly pronounced in RNNs when dealing with long-range dependencies.

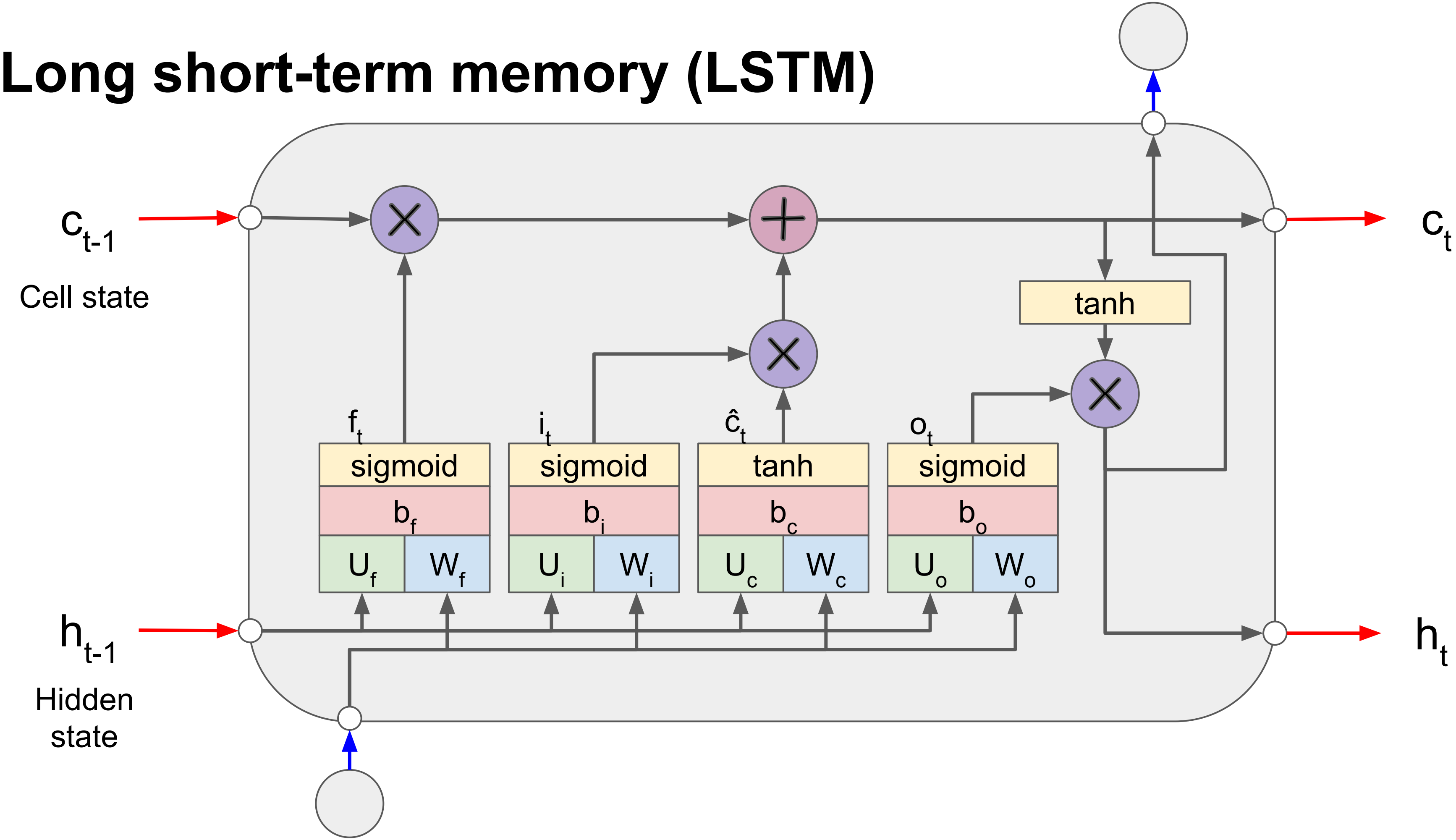


Long short-term memory (LSTM)

Long short-term memory (LSTM)



Long short-term memory (LSTM)



Long short-term memory (LSTM)

Forget gate: Determines which information from the previous cell state should be discarded to forget irrelevant data.

Input gate: Controls how much new information from the current input and previous hidden state is added to the cell state.

Output gate: Regulates what part of the cell state is output as the hidden state for the next time step.

Long short-term memory (LSTM)

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$ **Forget gate**

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$ **Input gate**

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$ **Output gate**

$$\hat{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$ **Cell input node**

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t$$ **Cell state**

$$h_t = o_t \odot \tanh(c_t)$$ **Hidden state**

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

$$\begin{aligned} f_1 &= \sigma(W_f x_1 + U_f h_0 + b_f) \\ &= \sigma([0.2, 0.4] \cdot [0.1, 0.2] + 0.5 \cdot 0 + 0.1) \\ &= \sigma(0.02 + 0.08 + 0.1) \\ &= \sigma(0.2) \end{aligned}$$

$$f_1 \approx 0.55$$

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$\begin{aligned} h_0 &= 0, \quad c_0 = 0 \\ W_f &= [0.2, 0.4], \quad U_f = [0.5] \\ W_i &= [0.3, 0.7], \quad U_i = [0.6] \\ W_o &= [0.6, 0.9], \quad U_o = [0.4] \\ W_c &= [0.5, 0.8], \quad U_c = [0.3] \\ b_f &= 0.1, \quad b_i = 0.2 \\ b_o &= 0.3, \quad b_c = 0.0 \end{aligned}$$

$$\begin{aligned} i_1 &= \sigma(W_i x_1 + U_i h_0 + b_i) \\ &= \sigma([0.3, 0.7] \cdot [0.1, 0.2] + 0.6 \cdot 0 + 0.2) \\ &= \sigma(0.03 + 0.14 + 0.2) \\ &= \sigma(0.37) \\ i_1 &\approx 0.59 \end{aligned}$$

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

$$\begin{aligned} o_1 &= \sigma(W_o x_1 + U_o h_0 + b_o) \\ &= \sigma([0.6, 0.9] \cdot [0.1, 0.2] + 0.4 \cdot 0 + 0.3) \\ &= \sigma(0.06 + 0.18 + 0.3) \\ &= \sigma(0.54) \\ o_1 &\approx 0.63 \end{aligned}$$

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

$$\begin{aligned}
\hat{c}_1 &= \tanh(W_c x_1 + U_c h_0 + b_c) \\
&= \tanh([0.5, 0.8] \cdot [0.1, 0.2] + 0.3 \cdot 0 + 0.0) \\
&= \tanh(0.05 + 0.16) \\
&= \tanh(0.21)
\end{aligned}$$

$$\hat{c}_1 \approx 0.21$$

LSTM – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

$$c_1 = f_1 \odot c_0 + i_1 \odot \hat{c}_1$$

$$= 0.55 \odot 0 + 0.59 \odot 0.21$$

$$c_1 \approx 0.12$$

$$h_1 = o_1 \odot \tanh(c_1)$$

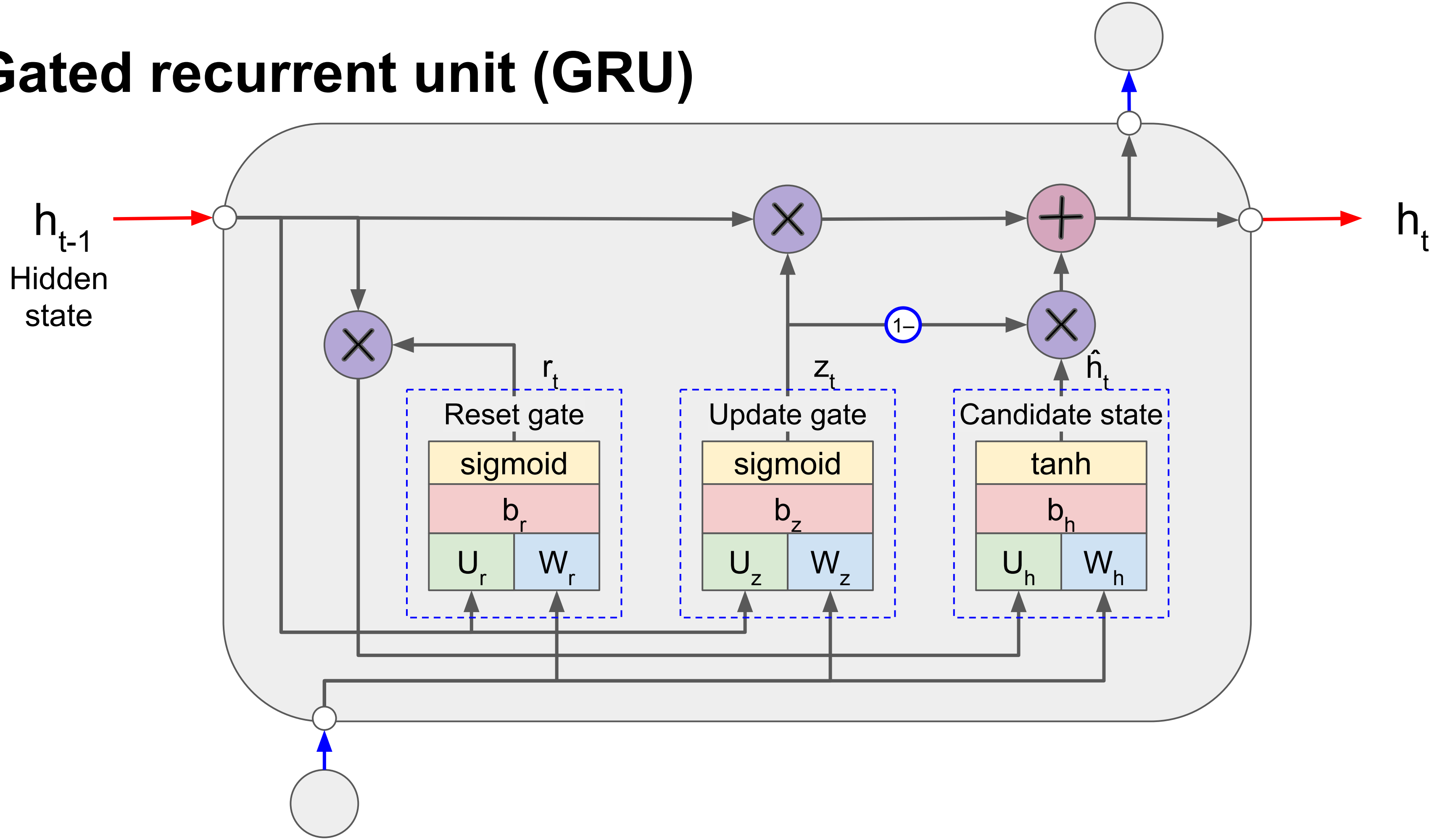
$$= 0.63 \odot \tanh(0.12)$$

$$= 0.63 \odot 0.12$$

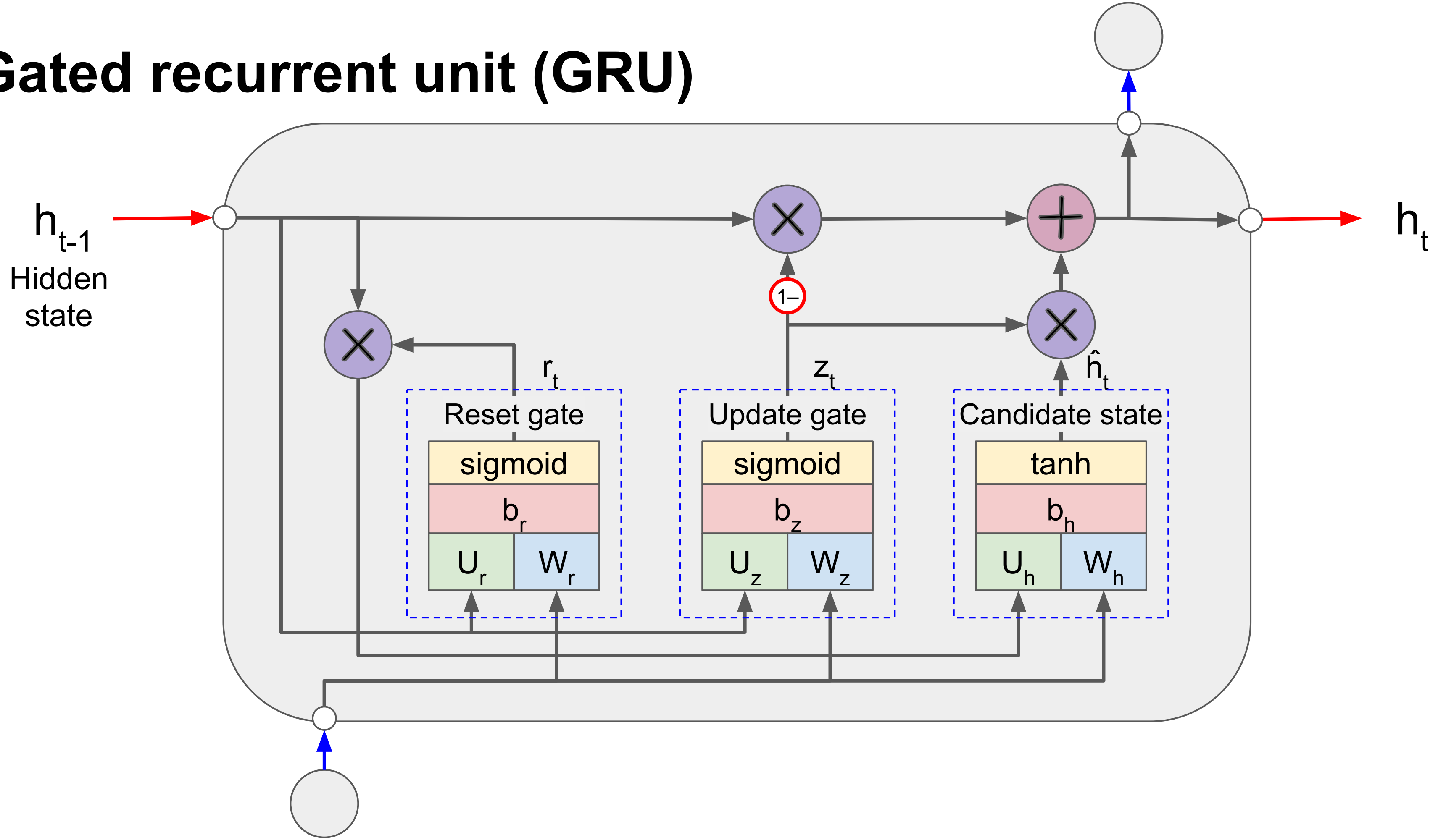
$$h_1 \approx 0.076$$

Gated recurrent unit (GRU)

Gated recurrent unit (GRU)



Gated recurrent unit (GRU)



Gated recurrent unit (GRU)

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

Update gate

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

Reset gate

$$\hat{h}_t = \phi(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

Candidate state

$$\left.\begin{array}{l} h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h}_t \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \end{array}\right\} \text{Hidden state}$$

GRU – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0$$

$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$

$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$

$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$

$$b_r = 0.1, \quad b_z = 0.2$$

$$b_h = 0.3$$

GRU – Example

$$x_1 = [0.1, 0.2]$$
$$h_0 = 0$$
$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$
$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$
$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$
$$b_r = 0.1, \quad b_z = 0.2$$
$$b_h = 0.3$$

$$\begin{aligned} r_1 &= \sigma(W_r \cdot x_1 + U_r \cdot h_0 + b_r) \\ &= \sigma([0.2, 0.4] \cdot [0.1, 0.2] + 0.5 \cdot 0 + 0.1) \\ &= \sigma(0.02 + 0.08 + 0.1) \end{aligned}$$
$$r_1 \approx 0.55$$

GRU – Example

$$x_1 = [0.1, 0.2]$$

$$h_0 = 0$$

$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$

$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$

$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$

$$b_r = 0.1, \quad b_z = 0.2$$

$$b_h = 0.3$$

$$\begin{aligned} z_1 &= \sigma(W_z \cdot x_1 + U_z \cdot h_0 + b_z) \\ &= \sigma([0.3, 0.7] \cdot [0.1, 0.2] + 0.6 \cdot 0 + 0.2) \\ &= \sigma(0.03 + 0.14 + 0.2) \end{aligned}$$

$$z_1 = 0.59$$

GRU – Example

$$x_1 = [0.1, 0.2]$$
$$h_0 = 0$$
$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$
$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$
$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$
$$b_r = 0.1, \quad b_z = 0.2$$
$$b_h = 0.3$$

$$\begin{aligned} \hat{h}_1 &= \tanh(W_h \cdot x_1 + U_h \cdot (r_1 \odot h_0) + b_h) \\ &= \tanh([0.6, 0.9] \cdot [0.1, 0.2] \\ &\qquad + 0.4 \cdot (0.5498 \odot 0) + 0.3) \\ &= \tanh(0.06 + 0.18 + 0.3) \end{aligned}$$
$$\hat{h}_1 \approx 0.49$$

GRU – Example

$$x_1 = [0.1, 0.2]$$
$$h_0 = 0$$
$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$
$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$
$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$
$$b_r = 0.1, \quad b_z = 0.2$$
$$b_h = 0.3$$

$$\begin{aligned} h_1 &= z_1 \odot h_0 + (1 - z_1) \odot \hat{h}_1 \\ &= 0.5914 \odot 0 + (1 - 0.5914) \odot 0.4910 \\ &= 0 + (1 - 0.5914) \odot 0.4910 \end{aligned}$$
$$h_1 \approx 0.20$$

GRU – Example

$$x_1 = [0.1, 0.2]$$
$$h_0 = 0$$
$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$
$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$
$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$
$$b_r = 0.1, \quad b_z = 0.2$$
$$b_h = 0.3$$

$$\begin{aligned} h_1 &= (1 - z_1) \odot h_0 + z_1 \odot \hat{h}_1 \\ &= (1 - 0.5914) \odot 0 + 0.5914 \odot 0.4910 \\ &= 0 + 0.5914 \odot 0.4910 \end{aligned}$$
$$h_1 \approx 0.29$$

Hands On

Compute the **RNN** cell manually using the equations provided, with **2-dimensional input x for 3 time-step**

$$X = (x_1, x_2, x_3) = \left(\begin{bmatrix} -1.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} 1.0 \\ -1.0 \end{bmatrix}, \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \right)$$

$$W = [0.4, 0.3]$$

$$U = 0.5$$

$$b = 0.1$$

$$h_0 = 0$$

Hands On

Compute the **LSTM** cell manually using the equations provided, with **2-dimensional input x for 2 time-step**

$$X = (x_1, x_2) = \left(\begin{bmatrix} -1.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} 1.0 \\ -1.0 \end{bmatrix}, \right)$$

$$h_0 = 0, \quad c_0 = 0$$

$$W_f = [0.2, 0.4], \quad U_f = [0.5]$$

$$W_i = [0.3, 0.7], \quad U_i = [0.6]$$

$$W_o = [0.6, 0.9], \quad U_o = [0.4]$$

$$W_c = [0.5, 0.8], \quad U_c = [0.3]$$

$$b_f = 0.1, \quad b_i = 0.2$$

$$b_o = 0.3, \quad b_c = 0.0$$

Hands On

Compute the **GRU** cell manually using the equations provided, with **2-dimensional input x for 2 time-step**

$$X = (x_1, x_2) = \left(\begin{bmatrix} -1.0 \\ 1.0 \end{bmatrix}, \begin{bmatrix} 1.0 \\ -1.0 \end{bmatrix}, \right)$$

$$h_0 = 0$$

$$W_r = [0.2, 0.4], \quad U_r = [0.5]$$

$$W_z = [0.3, 0.7], \quad U_z = [0.6]$$

$$W_h = [0.6, 0.9], \quad U_h = [0.4]$$

$$b_r = 0.1, \quad b_z = 0.2$$

$$b_h = 0.3$$