CS280 Fall 2021 Assignment 3
Part A

RNN, LSTM and GRU

November 22, 2021

Name:

Student ID:

1. Parity-check network (10 points)

Note that the initial parity bit is 1, what's the relation between each input and the previous parity bit? Determine the relation between the parity and inputs and complete the parity bits(p_1, p_2, p_3, p_4) and design and draw a RNN to predict parity.

Parity bits :	1	0	0	0	1	0	1	p_1	p_2	p_3	p_4	$\rightarrow$
Input :		0	1	1	0	0	0	1	1	0	0	

① The relation between input and previous parity bit is XNOR

② $[p_1, p_2, p_3, p_4]$ should be $[1, 1, 0, 1]$

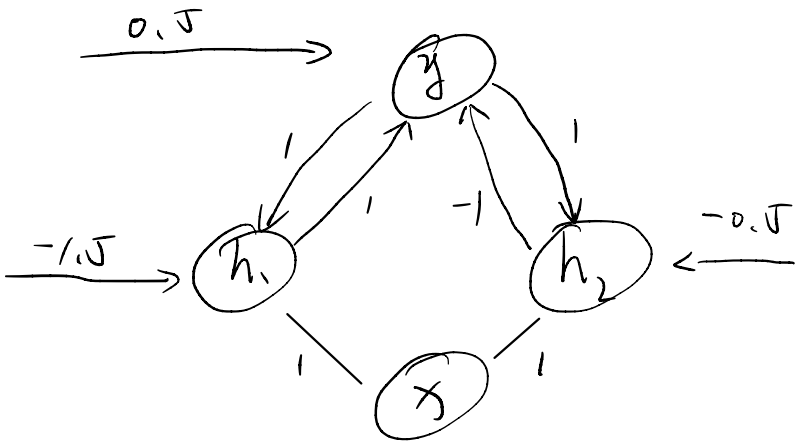
③ Denote the input bits and parity bits as x and y.

Denote the two hidden units as h_1, h_2.

Also look at the output strategy

$$\text{output} = \begin{cases} 1, & \text{state} \geq 0 \\ 0, & \text{state} < 0 \end{cases}$$

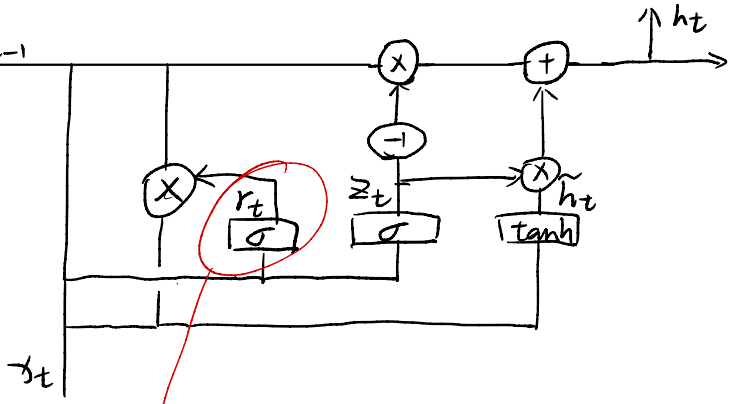
$y^{(t-1)}$	$x^{(t)}$	$h_1^{(t)}$	$h_2^{(t)}$	$y^{(t)}$
0	0	0	0	1
1	0	0	1	0
0	1	0	1	0
1	1	1	1	1



2. GRU (5 points)

1. Draw the diagram of GRU, describe the gates (where? What is the role of each gate?), and point out the differences between GRU and LSTM in the design of gates.
2. In what situations(s) is LSTM/GRU used respectively? Explain your reason.

1. ① h_{t-1}



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

② Reset gate : $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$

To control how much information used from the current input

③ update gate : $\begin{cases} z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases}$

The combination of the forget gate and the input gate in LSTM.

To control forgetting information and remembering information.

2. GRU takes less training data and training cost than LSTM, which is easy to train

Therefore, if there is enough data and resourses, LSTM performs better.

Otherwise GRU is better.