CS280 Fall 2021 Assignment 3
Part A

RNN, LSTM and GRU

December 11, 2021

Name: FanYuxin

Student ID: 2020233216

1. Parity-check network (16 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.

$$
\begin{array}{lccccccccccc}
\text{Parity bits}: & 0 & 0 & 0 & 1 & 0 & 1 & p_1 & p_2 & p_3 & p_4 & \rightarrow \\
\text{Input}: & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 &
\end{array}
$$

- $(p1, p2, p3, p4) = (1, 1, 0, 1)$.

- let input as In_i and parity bit as P_i. $P_i = (P_{i-1} == \text{In}_i)$.

- RNN: $\mathbf{H}_t = \phi(\mathbf{X}_t \mathbf{W}_{xh} + \mathbf{H}_{t-1} \mathbf{W}_{hh} + \mathbf{b}_h)$.
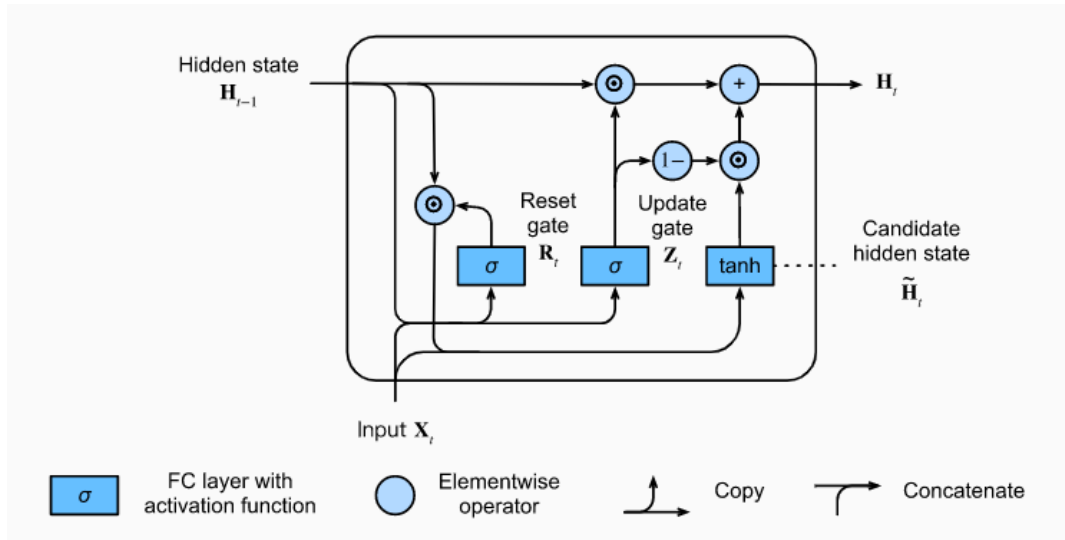 $\mathbf{O}_t = \mathbf{H}_t \mathbf{W}_{hq} + \mathbf{b}_q$

  Figure as :
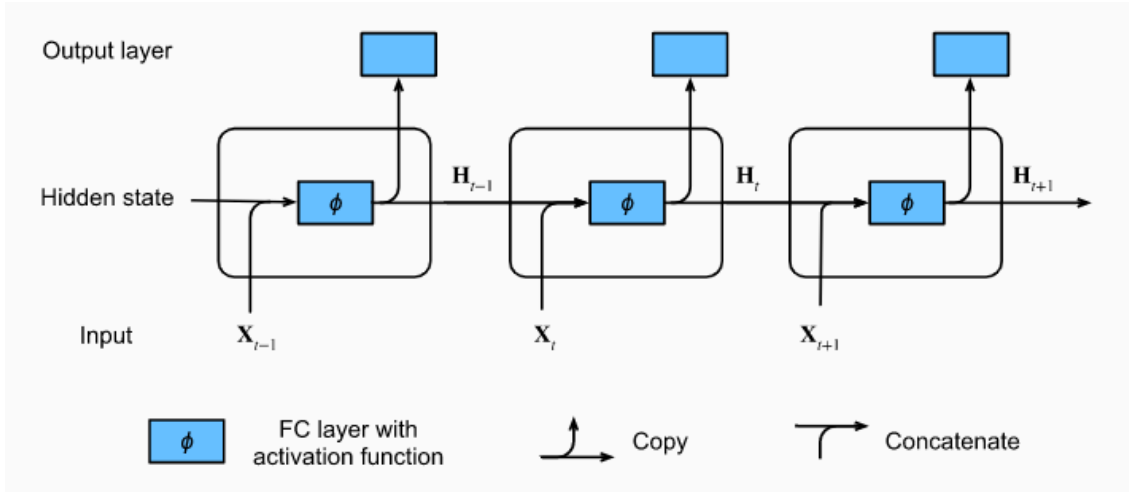
  with input demision = 1, hidden size = 2

2

2. GRU (17 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. GRU diagram

    Reset gates help capture short-term dependencies in sequences.

    Update gates help capture long-term dependencies in sequences.

    LSTMs have three types of gates: input gates, forget gates, and output gates that control the flow of information. The positions of gates of LSTM and GRU are also different.

    2. The number of parameters of GRU is smaller than LSTM. So GRU converges faster than LSTM.

    When dataset size is small, GRU and LSTM have similar performace, but GRU converges faster. GRU is better.

    When the dataset size is large, LSTM performs better than GRU.