AI-Deep Learning 2023(Final) Exam time (9:20-11:20)

1. (30%) Given an LSTM model as shown in Figure 1, point out (i.e., write down the symbols) which is the input gate, output gate, and forget gate, respectively (10%). Then explain the function of each gate (10%). Suppose that we have an application that needs to predict an output y for a sequence of three inputs (x1, x2, x3) (i.e., three time steps); draw an unfolded figure of LSTM for this application. (10%)

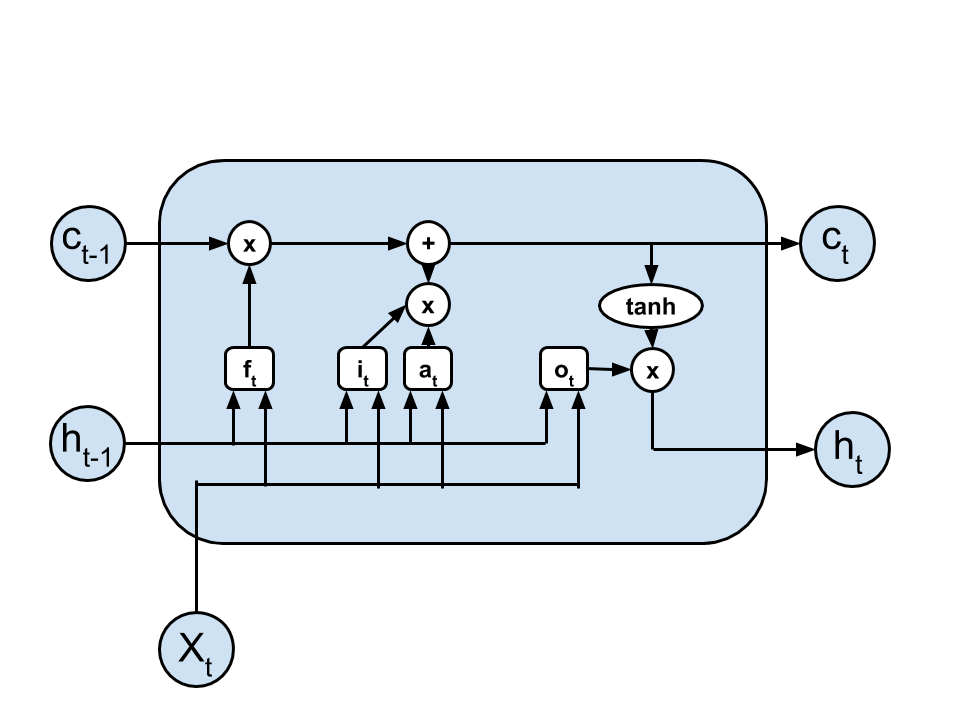
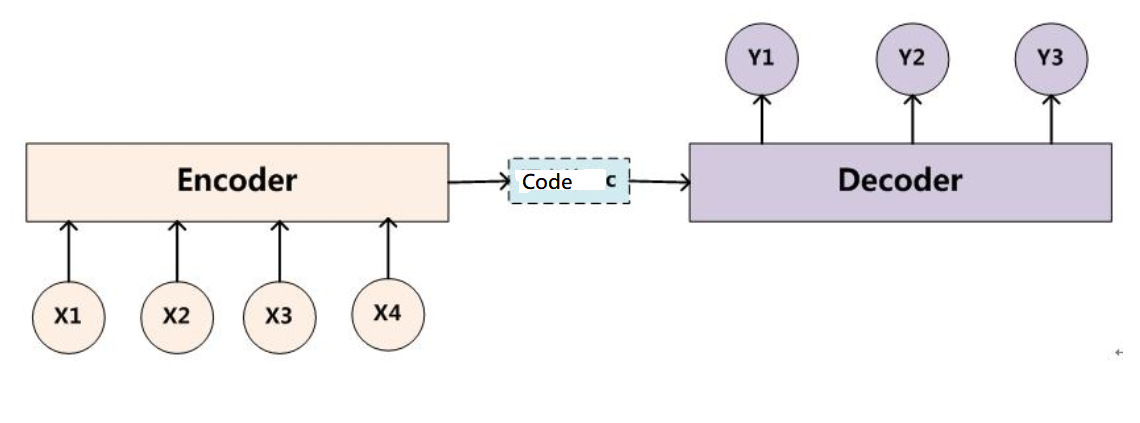


Figure 1. LSTM

1. (10%) What is the problem with a simple (or vanilla) RNN? How to solve it?
2. (10%) The following diagrams show a sequence-to-sequence translation application of an RNN with or without using attention. Point out which (Fig.2(**a)** or Fig.2(**b**) is the RNN with attention model and which is not. Explain what is the difference between them. Fig. 2(a)

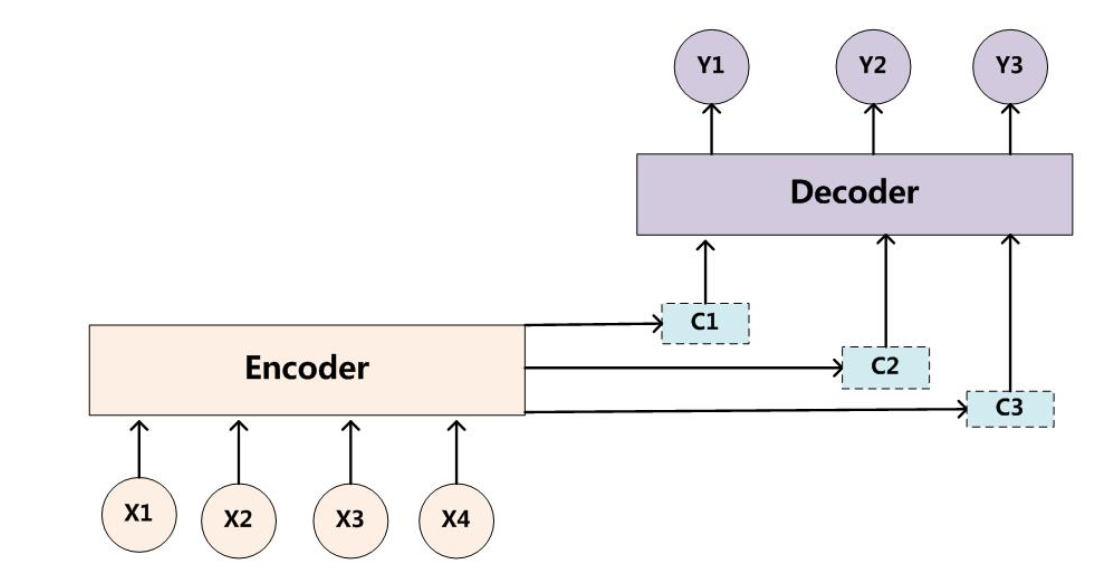


Fig. 2(b)

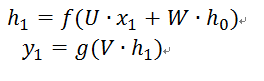
4. (20%) A simple RNN with an initial hidden state of h0=[0.0, 0.0], U=[0.5,0.6], V=[1.0,2.0], Hidden layer bias=[1.0,-1.0], Output bias=[0.1]

W = [0.1,0.2]

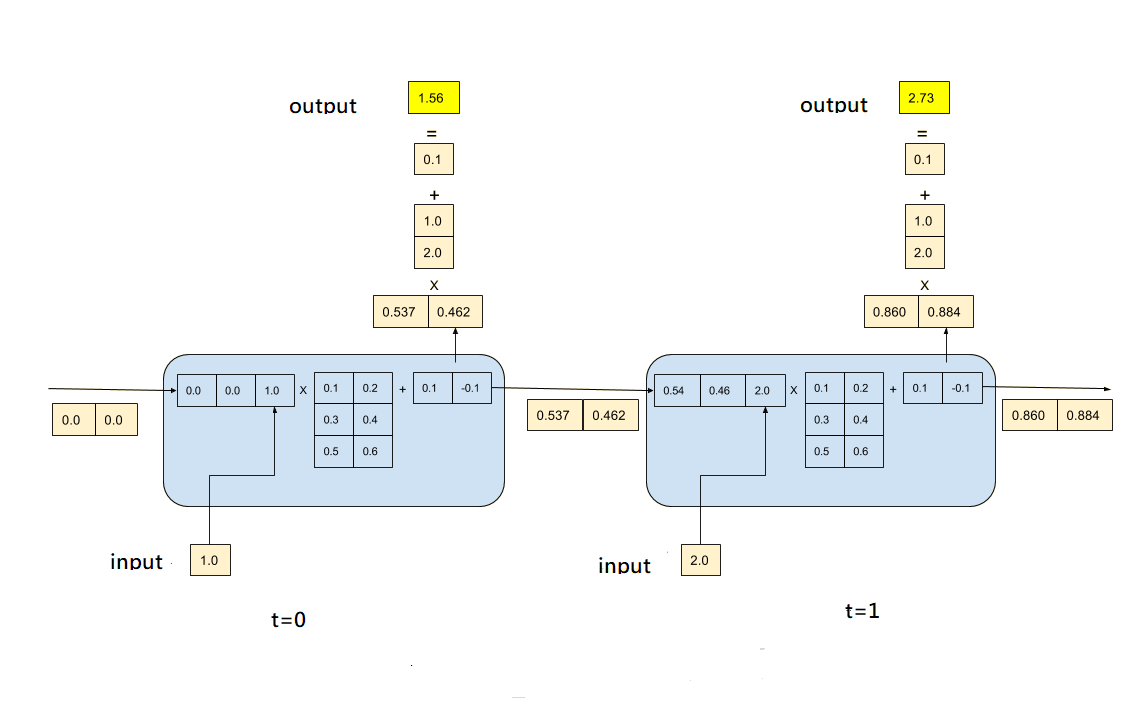
    [0.3,0.4]

**Given the input vector [2, 3], calculates the corresponding output of y.**

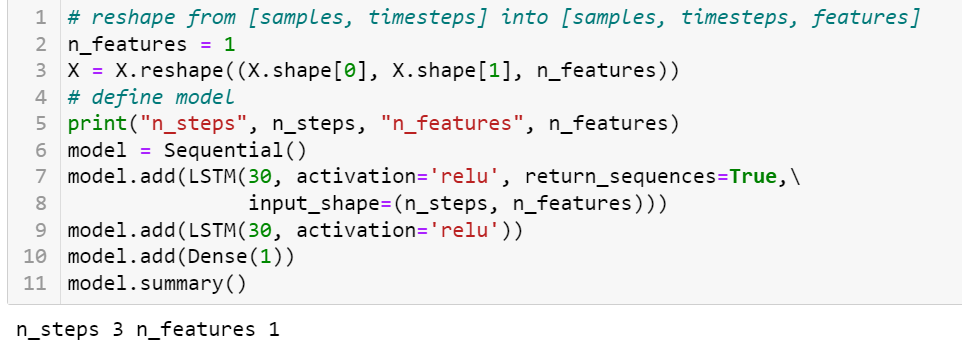
*For your reference, note that the activation function f is tanh().*

**

An example of calculation is shown in the following figure.



5. (20%)The following piece of code defines a deep LSTM model for a time series prediction. It uses three time steps of historical time series values to predict the value of the next time step.



Answer the following questions:

1. How many parameters are there in the first LSTM layer?
2. How many parameters are there in the second LSTM layer?

*Hint: The input of the first layer is just a real number, and the input of the second layer is the hidden vector of the first layer, which has a dimension of 30.*

6. (10%) Transformer is the foundation of the many recent large pre-trained language models, such as BERT and ChatGPT. The concept of self-attention is the core of the Transformer. Based on the following figure, please explain the relationships among query, key, value and the resulting representation Z. Please answer this question by considering only one-head attention. *Specifically, please answer how Z's first row is derived from v1 and v2.*

