LNMIIT/B.Tech./CSE/PE/2018-19/ODD/DL/ET

LNMIIT
The LNM Institute of Information Technology

The LNM Institute of Information Technology

Computer Science and Engineering

Deep Learning (DL)

End Term

Time: 3 hours | Date: December 06, 2018 | Max. Marks: 80

Read the following instructions carefully:

- *There are 12 questions printed on both sides of the paper.*
- **No marks** *for providing just expressions/answers unless accompanied with correct justification and/or derivation.*
- *In case of any doubt, make your assumption, write it clearly and continue.*
- *Your answers have to be mathematically sound rather than having lots of English text.*

1. (a) In a convolutional neural network (CNN), if the filter size is $f \times f$, the size of the image is $n \times n$, padding used is p and a stride of s is also applied. Compute the dimensions of the output.

   (b) Find out the appropriate values of the '?' in the following:

   i. $28 \times 28 \times 6 \xrightarrow{f=3,p=0,s=2,max-pools=2} ?$

   ii. $28 \times 28 \times 6 \xrightarrow{?} 28 \times 28 \times 34$

   4+(2+2) Marks

2. In practice, the performance of neural networks decrease beyond a certain depth. Why? Residual networks do not suffer from such a problem. Make a case for residual networks why they can afford to have many layers.

   2+4 Marks

3. Assume that we have a vocabulary of 10,000 words. We also know that deep learning works only with numbers. So each word is given a one hot encoding. Answer the following questions:

   (a) If the word *elephant* appears at place 1031 in the vocabulary, then what is its one hot encoding?

   (b) One day someone thinks of getting rid of one-hot encodings and tries to come up with another encoding for words. The new encoding is 300 dimensional for each word. Almost all the words in our vocabulary have non-zero values in the new 300 dimensional encoding. How would you compare it with the one-hot encoding in terms of performance of the neural network?

   2+4 Marks

4. Justify the equation for computing softmax and discuss its significance. How is it connected to sigmoid activation function?

   4+2 Marks

5. You are provided a previous layer activation volume with dimensions $28 \times 28 \times 192$. Your goal is to convert this volume to another volume of size $28 \times 28 \times 256$ using operations (1×1 conv, 3×3 conv, 5×5 conv, Max-pool) at least once and all at once for this layer. Also provide the count of parameters for this layer.

   6+2 Marks

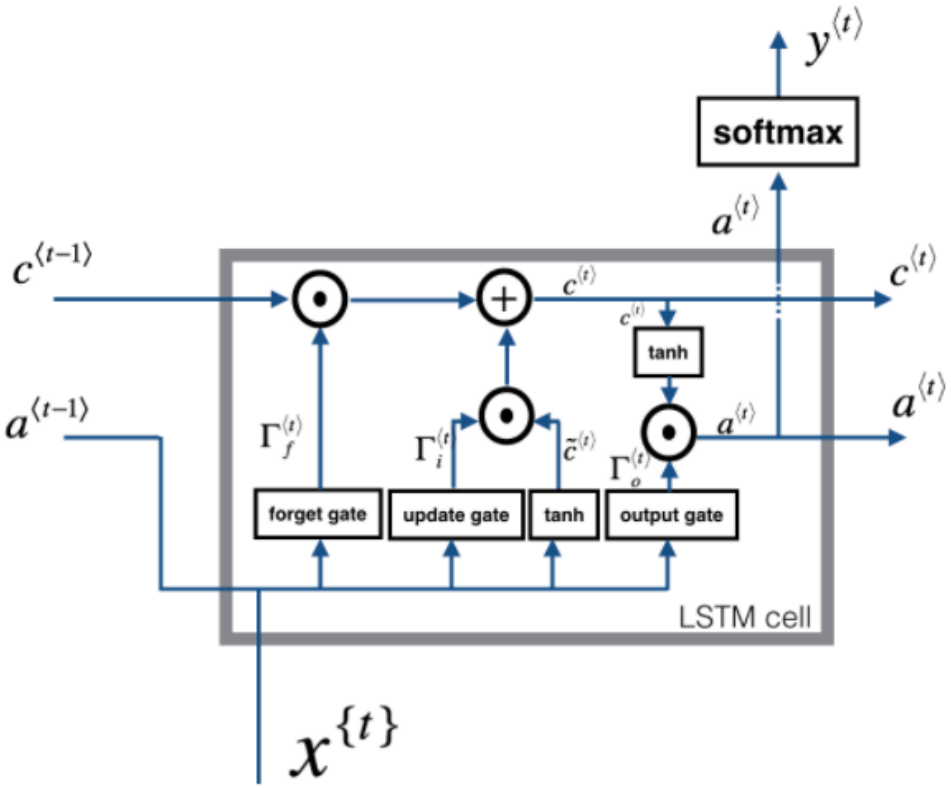

Figure 1: LSTM for question 10.

6. Discuss why a standard neural network cannot be used for sequence data. Provide four such examples of different kinds of RNNs along with their applications.

2+4 Marks

7. Draw a neat and clean Bi-directional RNN. Write the equations governing it by clearly labeling the diagram with the variables used in the equations.

2+4 Marks

8. Write the equation of cost function with the terms of L1 and L2 regularization. Explain the role of both these kinds of regularization in reducing overfitting?

(1+1)+2+2 Marks

9. Write batch gradient decent algorithm. Derive the equation of gradient decent with momentum. How does it improve learning rate?

2+4+2 Marks

10. You are provided a sketch of an LSTM cell in Figure 1. Derive five equations related to it. (2*5 Marks)

11. What is the exploding and vanishing gradients problem? What are the ways to overcome this problem?

5 Marks

12. Write the backpropagation algorithm. How do you perform backpropagation when using dropout as a regularizer.

5 Marks