



JKLU/B.Tech./CSE/Elective/2021-22/ODD/NLP/ET

JK Lakshmipat University Jaipur
Computer Science and Engineering
Natural Language Processing (CS2203)

End Term

Time: 2 hours

Date: December 20, 2021

Max. Marks: 50

Read the following instructions carefully:

- There are 7 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.
- Exam is **open hand-written** notes, i.e., you are allowed to bring handwritten notes but no photocopies or printouts or any other printed material is allowed.

1. Assume a huge text corpus whose vocabulary size is 10,000 words.

(a) How will you represent each word using one hot vectors?

- (b) How will the representation change if the corpus has 2,00,000 words in the vocabulary?
- (c) State two prominent disadvantages of one hot vector representation.
- (d) What are word embeddings?
- (e) How do word embeddings solve the problems mentioned by you in part (c)?

2+2+2+2+2 Marks

2. (a) Plain RNN unit is almost never used in real applications. What is actually used is gated recurrent units like GRUs and LSTMs. Do you agree with these statements? State reasons for your answer.
- (b) Draw bi-directional LSTM network.
- (c) Provide two examples where bi-directional LSTMs will be more effective than plain RNN units.
- (d) How RNNs help making a language model?

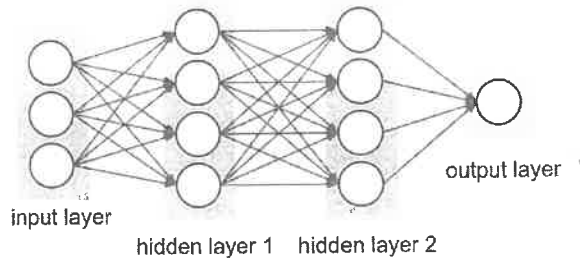
2+2+2+2 Marks

3. Assume your neural network has only two features and two parameters w_1 and w_2 . You are supposed to do binary classification.

- Draw a potential network subject to the above constraints.
- Write down stochastic gradient descent algorithm for learning the above parameters.
- Give ideas on how can we use a better learning rate?

2+2+2 Marks

4. Consider a neural network as shown in the figure.



- Provide the dimensions of parameter matrices learnt.
- Show that initializing all the parameters to zero is a bad idea.
- Redraw the network with minimum changes required, if it is designed to classify three distinct classes.
- Show that the softmax activation computes a probability distribution in part (c).

2+2+2+2 Marks

5. Show how the attention mechanism works in the case of neural machine translation using RNNs. You need to discuss query and values. How does the use of attention mechanism improve the model that just uses RNNs?

2+2+2 Marks

6. What is language modeling problem? Write and explain three applications where it may be used as a subroutine. Describe the n-gram approach for solving language modeling problem.

2+~~2~~+2 Marks

7. Differentiate between the skip gram model and the continuous bag of words model for word2vec.

4 Marks