



INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

End-Autumn Semester Examination 2022-23

Date of Examination: 24/11/2022 Session: FN Duration: 3 hours Full Marks: 60

Subject No.: CS60057

Subject: Natural Language Processing

Department: Computer Science & Engineering

Special Instructions :

1. All parts of a question must be answered together.

2. Each question must begin on a new page.

.....

1. Consider the following language models:

M1 ngram model ($n = 4$)

M2 Neural window based language model: A neural network whose input is the previous three words and the aim is to predict the next word. Consider using a neural network with one hidden layer having 200 units. Words are input as 1-hot vectors.

M3 RNN based language model with a single RNN layer. Assume that the hidden unit has 500 neurons.

(a) Assume that the Vocabulary size is 1,00,000. For each of the above models, briefly describe how the model works or sketch the model. Find out the total number of parameters / values to be stored for each of these models. (9)

(b) How is a language model evaluated? Give a quantitative measure with respect to a test corpus. (3)

2. (a) Suppose that you have a dataset comprising the writings of 4 authors (A1, A2, A3 and A4). You are required to design a classifier to predict the author of a given new article which has been written by one of the 4 authors.

i. Under the following conditions, state the value of the cross-entropy loss on a single example of the classifier. You are given that the example article has been actually authored by A1. (3)

i Your classifier predicts each author with equal probability (0.25, 0.25, 0.25, 0.25).

ii Your classifier predicts the following probabilities for the authors (0.7, 0.1, 0.1, 0.1)

iii Your classifier predicts the following probabilities for the authors (0.7, 0.3, 0, 0)

ii. Suggest a good model that you may use for this task of authorship attribution. (3)

(b) Consider a LSTM based model that you have designed for the sentiment classification task. The model takes as input pre-trained word2vec vectors of the words, and predicts the sentiment label of a given review text (positive or negative).

i. State why the LSTM model may be better than a basic RNN model. (2)

- ii. Suppose that you start with the words encoded by pre-trained word2vec, but you update the word vector while training the LSTM. How would these word vectors differ from ones not updated during training? Explain with an example. (3)
3. Consider a self-attention layer of a neural network containing 4 attention heads. The input is a sequence of nodes of length 100 taken as x_1, \dots, x_n . Each node x_i is a 50-dimensional vector. State any other hyper-parameters that are needed and write down the value(s) you have chosen for them.
- (a) Write down the expressions used to compute the values of y_1, y_2, \dots, y_n of the next layer. (4)
- (b) How many parameters are required by the above model? (This may depend on the hyperparameters you have chosen.) (3)
4. Suppose that you wish to use the noisy channel model based MT to translate from Tamil (source) to Bengali (target).
- (a) Given a sentence in Tamil, show how the model is formulated to find the optimum Bengali sentence. (4)
- (b) What do each of the two components of the above model correspond to? What are the resources (data) required for the components? (4)
- (c) Name the algorithms that are used for the two components (name only) (2)
5. (a) i. Suppose that you wish to use a BERT model for the tasks below. You have some training examples of each task. For each task below, show how BERT can be used for the task. The entire procedure must be clearly explained. If the task cannot be effectively modelled using BERT, explain why. (10)
- T1 Sentiment Classification
- T2 Part of Speech Tagging
- T3 Machine Translation
- T4 Question Answering
- T5 Text Summarization
- ii. State the basic difference between the BERT and BART model (2 sentences) (2)
- (b) i. Describe in brief a method to learn multilingual word embeddings that work with 4 different languages. You are given monolingual corpora in all these languages. Your model may use some additional resource which you must clearly state. (4)
- ii. How is multilingual BERT trained? How can it be used for crosslingual transfer? (4)