Mid Semester Examination (2024-25)
CS411: Deep Learning for Natural Language Processing

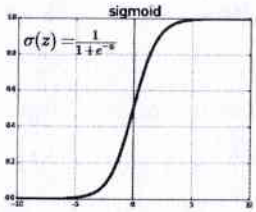
Time: 02 Hrs. Max. Marks: 30

Attempt all the questions. Please be brief and concrete in your answers. Q 6. has two parts (a) and (b), attempt any one.

Q1. (a) What is the ambiguity in parsing the sentence "There's an awful cost to getting a PhD that no one talks about"? (write the answer in 2-4 lines) **[Marks 1+1=2]**

(b) Which kind of ambiguity the following sentence "No smoking areas allow hookas inside, except the one in Hotel Grand" consists of? (discuss it in 2-4 lines)

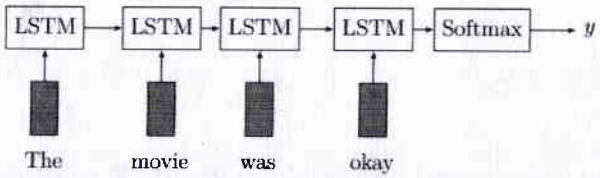
Q2. Consider the sigmoid activation function below:



What would the gradient of the sigmoid be with respect to an input that is very large?

[Marks 2]

Q3. A popular model used for sentiment classification is an LSTM model:



This model inputs word vectors to the LSTM model at each time step and uses the last hidden state vector to predict the sentiment label (y). If we use a simple "bag-of-vectors" model for sentiment classification and take average of all the word vectors in a sentence to predict the sentiment label, Name at least two benefit of the LSTM model over the bag-of-vectors model.

[Marks 3]

Q4. If we chose to update our word vectors when training the LSTM model on sentiment classification data, how would these word vectors differ from ones not updated during training? Explain with an example. Assume that the word vectors of the LSTM model were initialized using GloVe or word2vec. **[Marks 3]**

Q5. For each of the following tasks, state how you would run an RNN to do that task. In particular, specify how the RNN would be used at test time (not training time), and specify

a) how many outputs i.e. number of times the softmax ŷ (t) is called from your RNN. If the number of outputs is not fixed, state it as arbitrary.
b) what each ŷ (t) is a probability distribution over (e.g. distributed over all species of cats)

c) which inputs are fed at each time step to produce each output The inputs are specified below.

1. **Named-Entity Recognition:** For each word in a sentence, classify that word as either a person, organization, location, or none. Inputs: A sentence containing n words. **[Marks 3]**
2. **Sentiment Analysis:** Classify the sentiment of a sentence ranging from negative to positive (integer values from 0 to 4). Inputs: A sentence containing n words. **[Marks 3]**

Q6 (a) You observe that your model predicts very positive sentiment for the following passage:

"Yesterday turned out to be a terrible day. I overslept my alarm clock, and to make matters worse, my dog ate my homework. At least my dog seems happy..."

Why might the model misclassify the appropriate sentiment for this sentence? **[Marks 2]**

OR

(b) If we use term-term co-occurrence metrics with SVD techniques to obtain the word vector representation. What would be the limitation of this method? **[Marks 2]**

Q7. word vectors can alternatively be learned via cooccurrence count-based methods. How does Word2Vec compare with these methods? Please briefly explain one advantage and one disadvantage of the Word2Vec model. **[Marks 2]**

Q8. Word2Vec represents a family of embedding algorithms that are commonly used in a variety of contexts. Suppose in a recommender system for online shopping, we have information about co-purchase records for items x_1, x_2, ..., x_n (for example, item xi is commonly bought together with item x_j). Explain how you would use ideas similar to Word2Vec to recommend similar items to users who have shown interest in any one of the items. **[Marks 3]**

Q9. Give two examples of how we can evaluate word vectors. For each example, please indicate whether it is intrinsic or extrinsic. **[Marks 3]**

Q10. Explain the vanishing gradient problem. How does LSTM address this issue compared to traditional RNNs? **[Marks 4]**