

Last Name: \_\_\_\_\_

First Name: \_\_\_\_\_

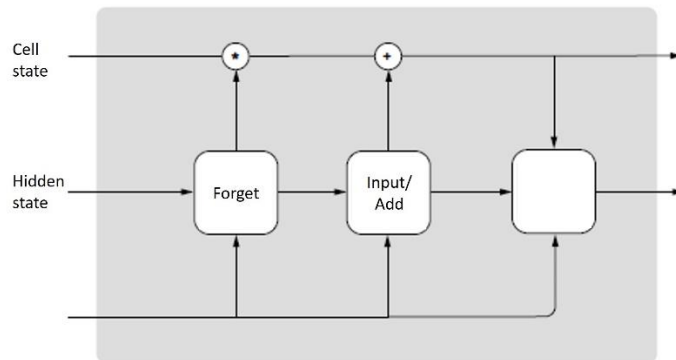
Course: ☐ 423 ☐ 723

**Midterm Exam**  
**Fall 2023 - COMPSCI 423/723**  
**October 30, 2023: 2:30-3:45 pm**  
**Natural Language Processing**  
**Total Points: 25**  
**Total Questions: 8**

*Closed book.*

*No computers or electronic devices are allowed for use during the exam.*

1. [3 points] Following is a schematic diagram of an LSTM cell.



- (i) Depict the cell state and the hidden state in the above diagram.
- (ii) Show the forget gate and the input/add gate in the above diagram.
- (iii) What is the purpose of the forget gate in LSTM? Answer in one or two sentences.

To learn to forget some information from the cell state which is not relevant for the rest of the task.

2. [3 points]

(i) The following sentence is encountered while training a trigram language model.

“The dog slept on the couch.”

This sentence will contribute towards estimation of which probabilities in the model?

$P(\text{slept}|\text{The,dog})$ ,  $P(\text{on}|\text{dog,slept})$ ,  $P(\text{the}|\text{slept,on})$ ,  $P(\text{couch}|\text{on,the})$

(ii) Which probability/probabilities will be used to predict the next word in the following sentence out of the given choices if a trigram language model is being used?

“The tree fell over the \_\_\_\_\_.” (house/street/car)

$P(\text{house}|\text{over,the})$ ,  $P(\text{street}|\text{over,the})$   $P(\text{car}|\text{over,the})$

(iii) Will a trigram language model give a low probability to the following ungrammatical sentence? Explain your answer in a sentence.

“The cars on the busy highway was going very fast.”

No. The sentence has a long-distance dependency between “cars” and “was” separated by 4 words, a trigram model cannot capture dependencies beyond the distance of three words.

3. [4 points] Answer each of the following in one sentence.

(i) What is the purpose of the skip-gram model?

To learn embeddings for words.

(ii) What prediction task is used to train a skip-gram model?

Predict previous and next  $n$  words given the current word in a sentence, where  $n$  is a parameter.

(iii) Given a corpus, how will you generate a training dataset to train a skip-gram model?

Use every word of every sentence as input and use the previous and next words as targets.

(iv) What is learned in the hidden layer of a skip-gram model?

The word embeddings.

4. [3 points] Answer each of the following in one sentence.

(i) What is the NLP task of part-of-speech tagging?

To label each word of a sentence with its part-of-speech tags, such as nouns, verbs, adjectives, prepositions, etc.

- (ii) Why is part-of-speech tagging not a trivial task?

The same word could have different part-of-speech tags in different sentences based on the context, hence one cannot trivially associate a part-of-speech tag with every word.

- (iii) Name any one machine learning model that can be trained to do part-of-speech tagging.

Recurrent neural networks.

5.

- (i) [2 points] A sequence labeling model is being used to extract person and location entities from sentences. Give the model's output labels for the following sentence if the model correctly extracts the entities.

“Ever since John Smith left Los Angeles, he has been living in the New York City.”

“Ever/O since/O John/B-PER Smith/I-PER left/O Los/B-LOC Angeles/I-LOC, he/O has/O been/O living/O in/O the/O New/B-LOC York/I-LOC City/I-LOC.”

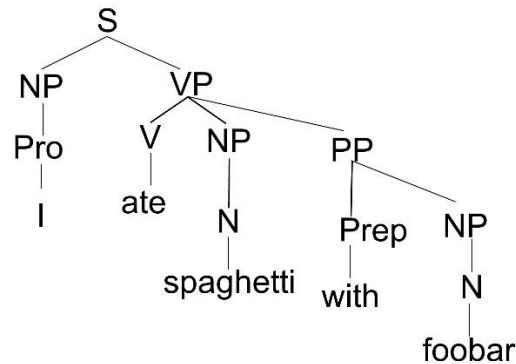
- (ii) [1 point] If a bi-directional RNN is being used for the task in Part (i), then information about which other words will it take into account while making prediction for the word “has”?

All other words of the sentence.

- (iii) [1 point] Name the evaluation measure(s) which is/are used to evaluate named entity recognition performance. No explanation or definition is required.

Precision, recall and F-measure.

6. The following is a parse tree for the sentence “I ate spaghetti with foobar.”



- (i) [1 point] Write down the productions involving “VP” in this parse.

$S \rightarrow NP VP$

$VP \rightarrow V NP PP$

- (ii) [2 points] Based on the parse, tell whether “foobar” is an edible item (e.g. sauce) or an instrument used for eating (e.g. fork). Explain your answer in one sentence.

Based on the parse, “foobar” should be an instrument of eating. This is because the prepositional phrase “with foobar” is attached to the verb phrase corresponding to the verb “ate” indicating that eating happened “with foobar”.

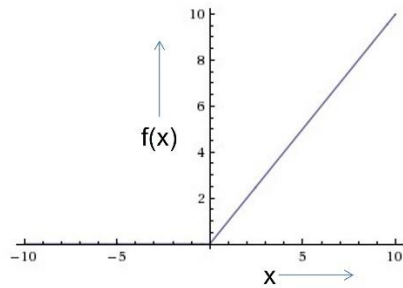
(Note: If it was an edible item then the prepositional phrase “with foobar” would have been attached to the noun phrase corresponding to “spaghetti”.)

7.

- (i) [1 point] What enables neural networks to learn non-linear relationship between inputs and outputs? Answer in a sentence.

Non-linear activation function.

- (ii) [1 point] Identify the following activation function by its name. Mention one of its advantages.



ReLU activation function. Its gradient does not diminish when used in back-propagation over several layers.

8.

- (i) [2 point] Why are word embeddings preferable over one-hot encoding to numerically represent words? Answer in one or two sentences.

They take fewer dimensions, and are not sparse, making them computationally more suitable. Additionally, similar words typically have similar word embeddings which helps in generalizing over words, on the other hand, one-hot encodings have no similarity between any two words.

- (ii) [1 point] Why do most methods learn similar word embeddings for antonyms (words with opposite meanings)? Answer in one sentence.

Antonyms tend to occur in similar contexts, for example, wherever “good” can be used “bad” can also be used, and embedding for a word is learned based on the contexts in which it occurs.