# Homework 3: Multi-Layer & Language Modeling

CS 1470/2470

Due October 23, 2019 at 11:59pm

## 1 Conceptual Questions

1. What are the dimensions of an embedding matrix? What do they represent?

2. Given the following sentences, plot reasonable embeddings in 2d for Sam, Came, Theater, Store, and Went. *(Hint: A simple graph with some clusters is fine.)*

   Sam went to the store.
   Then Sam came to the theater.
   I went to the theater.

3. What are some benefits to using RNNs over trigrams (or n-grams generally speaking?)

4. What are LSTM cells? How are they different from Vanilla RNNS, and why are they able to 'remember' information for longer timeframes than vanilla RNNs? *(Hint: Your answer should, at minimum, address the concepts of gates and gradients.)*

5. (Optional) Have feedback for this assignment? Found something confusing? We'd love to hear from you!

## 2 Ethical Implications

1. Since 2014, Amazon has been automating parts of its job application process. As part of this effort, they created a language model that took in resumes as an input and returned a score that represented hireability. In 2015, it was discovered that this model discriminated against women, and by 2018, the model had been scrapped because it was too discriminatory. Read more about the model here:

(a) How can language models inherit bias from the world around them? What mistakes did Amazon make with this model that made it discriminatory?

(b) What strategies did Amazon use to attempt to correct for the bias in their model? In which ways did those strategies succeed and/or fail?

2. In February 2019, OpenAI created a transformer-based language-generator model called GPT-2. The text that it is capable of generating nears human-level writing ability. The computer scientists who made the model believed was irresponsible to release the entire thing, as it would be capable of auto-generating fake news. Instead they have released several substantially smaller model architectures. Read more about GPT-2 and the rationale behind its (non)release here.

   (a) What positive effects (if any) come from the deployment of this algorithm?

   (b) What negative effects (if any) come from the deployment of this algorithm?

   (c) What negative effects (if any) come from the mistakes made by this algorithm?

   (d) This past August, two Brown University graduate students recreated GPT-2 from scratch and are now delivering auto-generated text online. Do you agree more with OpenAI's decision to not release the model or the Brown graduate students' decision to make the model accessible to the public? Why?

# 3 CS2470-only Questions

1. The Gated Recurrent Unit (GRU) is another recurrent network cell that can, like the LSTM, retain information over long sequences. How is it able to do this? Describe its architecture, and compare it with that of the LSTM.

2. While we have studied Convolutional Neural Networks (CNNs) in the context of 2D images, CNNs can also be used for 1D sequence modeling tasks, such as language modeling. Look up some papers that have attempted this, and that compare CNN language models to RNN language models (cite which papers you read). What appears to be the general consensus on the pros and cons of the two approaches?