Your Name: _____

COMP 394 (NLP), Fall 2024

Practice Exam 2

Suhas Arehalli

Notes about this exam:

- Please write legibly and large enough so I can read your answers.
- You **may and should** ask me for clarification if you have any questions.
- You may use your handwritten notecard you have created.
- This is an individual assignment **DO NOT** discuss it with anyone.

Classifiers

Suppose you've trained a binary logistic regression model to predict whether a sentence was written by author A (label 0) or author B (label 1). You have hand-designed the following set of features: f_0: the presence of punctuation (0 or 1), f_1: the length of the sentence in words, f_2: the number of times the word "dog" appears in the sentence, and f_3: the number of times the word "cassowary" appears in the sentence.

Note that we feed the input into the model as the feature vector

$$x = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix}.$$

You train the model with SGD (as we discussed in class) and we learn the weight vector

$$w = \begin{bmatrix} -3.20 \\ 0.35 \\ 1.05 \\ -0.35 \end{bmatrix}$$

(a) Is the input *the dog and the cassowary ran away* more likely to be written by author A or B?

(b) According to your model, does the presence of punctuation (independent of other features) make a sentence more likely to be written by author A or B? Are longer sentences more likely to be written by author A or B? Which author is more likely to be identified by the use of the word *dog*? What about *cassowary*?

Word Embeddings

(a) Which of the following are likely to be *sparse* vector representations? Circle the letters corresponding to these choices.

a. TF-IDF document vectors
b. word2vec word vectors
c. PPMI word vectors
d. a one-hot vector

(b) Which of the following most accurately describes the skip-gram task?

a. Predict a word's meaning given the word
b. Predict a target word given surrounding context words.
c. Predict context words given a target word.
d. Predict a label given a sequence of tokens.

(c) Briefly describe what is meant by the *distributional hypothesis* and how it motivates the training of dense word vectors.

Feedforward NNs

A colleague is building a sentiment analysis model on movie reviews, and they begin by using the simplest model you can think of — a perceptron trained on hand constructed features.

(a) They run into a problem with 2 specific features: One indexes the presence of the word *don't* and the other the presence of the word *not*. These features are such that you classify the sentences that have one of these features correctly, but the model makes errors when both are activated.

To help them diagnose the issue, you look at some of the misclassified examples and see the sentence *I don't mean it's not good — it is!*.

What does this problem indicate? How might you modify the model to make it work? Briefly defend your suggestion, using the language and examples we've developed in class.

(b) Your colleague later decides that they need a model that allows for neutral labels. They approach this by training a model that produces 3 scores, one for each label, that indicates how well the input matches that label. For now, they are classifying based on which model gets the highest score, but they would like to also provide a probability distribution over the 3 labels for each input.

What would you suggest they do? What tools might be helpful to them?

Modern NN Architectures

(a) We briefly discussed how n-gram models can fail to capture phenomena like agreement. For example, a 4-gram models cannot, in principle, reliably predict the number of the verb that can come after *the dog near the gardens.*. You can construct such an example for an n-gram model for any choice of n, including a neural n-gram model.

Briefly explain why you cannot construct such in-principle counterexamples for an RNN model.

(b) Is the same true for a transformer model? Explain.