

Name: Chakrya Ros

Section 3: Neural Models (15 points)

1. How many weights must be trained for a Feed-forward Neural Network that does the sentiment classification task from question 8 in the Statistical Models section?

We have 4 features: $(d_h * (d_x + 1)) + (d_h * (d_x + 1)) + ((d_h + 1) * d_o)$

$\Rightarrow \text{Weights} = (2 * (4+1)) + (3 * (2+1)) + ((3+1)*1) = 23$

2. Explain how words are represented when used in neural language models (2 – 3 sentences) **and** how these representations are produced using the skip-gram algorithm (2 - 3 sentences).

A word embedding is to represent each word as vectors when used in neural language models. It's used the features learning techniques in NLP where words from vocabulary are mapped to vectors of real numbers. For example, if we have 4 words like cat, dog, tiger and lion in the vocabulary then we can represent them as word vectors as cat [1,0,0,0], dog [0,1,0,0], tiger [0,0,1,0], lion [0,0,0,1]. Finally, we use these vectors to train the neural language models.

By using the skip-gram, the goal is to learn the weights of the hidden layer that are actually the “word vectors” that we try to learn. They learn two separate embeddings for each word, one for the word as a target word and one for the word as a context word. To producing the word embedding for Skip-gram model, given a word, we will try to predict its neighboring words. The contexts are immediate neighbors of the target and are retrieved using a window of size n by capturing n words to the right of the target and n words to its left.

3. Explain how an RNN with 1 hidden layer that uses “regular” hidden units (hidden units with a single non-linearity) can perform POS tagging. Make sure to address all of the following:

- a) **Weight matrices are 3 (weight U, V, and W).**
- b) **Inputs at each time step are pre-training word embeddings corresponding to the input tokens.**
- c) **Given the inputs, the hidden layer computes those inputs with activation value of the hidden layer from the preceding point in time. The hidden layer from the previous time step provides a form of memory or context and give information to be made at later points in time.**

- d) Output layer are tag probabilities generated by a softmax layer over the tag set.
- e) POS tags are assigned to a given sentence using sequence model to build a local classifier that classifies each word left to right and find the best tag for the first word in the sentence, then find the best tag second word and so on. In the RNN model, we apply softmax of each hidden layer on each classifier of tag and pick the one that have high probability of classifier for that word.

Multiple Choice

- 4. Which of the following is a Feed-forward Neural Network most similar to?
(Choose one)
 - a) N-gram Language Model
 - b) Naïve Bayes Classifier
 - c) HMM
 - d) Logistic Regression Classifier
 - e) MEMM
- 5. **True or False:** an RNN performing POS tagging of the last word in a sentence is influenced by the first word in the sentence.
False
- 6. For skip-gram word embeddings, the dimensionality is _____ (**dependent on**/independent from) the size of the vocabulary of the training data.

Citations:

- Chapter 8, 9, textbook
- Hw4
- Quiz 7, 8