Programming Assignment 2: Learning Word Representations.
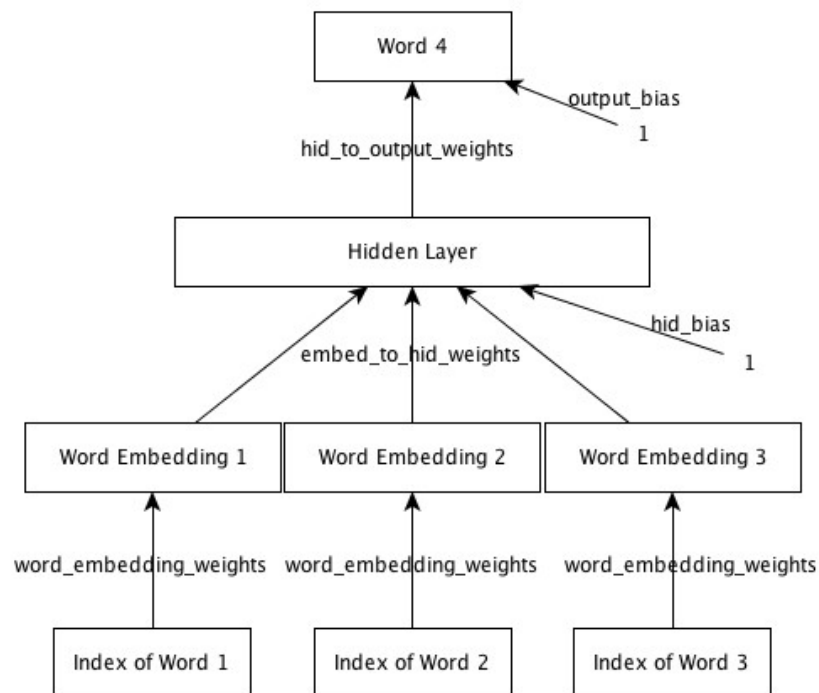
Quiz, 13 questions

<div style="border:1px solid">1
point</div>

1.

We are now ready to start using neural nets for solving real problems!

In this assignment we will design a neural net language model. The model will learn to predict the next word given the previous three words. The network looks like this:

To get started, download any one of the following archives.

assignment2.tar.gz

Or

assignment2.zip

Programming Assignment 2: Learning Word Representations.

Quiz, 13 questions

containing each file individually:

- README.txt

- train.m

- raw_sentences.txt

- fprop.m

- word_distance.m

- display_nearest_words.m

- predict_next_word.m

- load_data.m

- data.mat

The starter code implements a basic framework for training neural nets with mini-batch gradient descent. Your job is to write code to complete the implementation of forward and back propagation. See the README file for a description of the dataset, starter code and how to run it.

This sample_output shows you what output to expect once everything is implemented correctly.

Once you have implemented the required code and have the model running, answer the following questions.

Ready to start? (Please select a response. This is a reflective question and choosing one answer over the other will not count against this quizzes' grade.)

○ Yes

○ No

4
points

2.

Train a model with 50 dimensional embedding space, 200 dimensional hidden layer and default setting of all other hyperparameters. What is average validation set cross entropy as reported by the training program after 10 epochs ? Please provide a numeric answer (three decimal places). [4 points]

Enter answer here

3
points

3.

Train a model for 10 epochs with a 50 dimensional embedding space, 200 dimensional hidden layer, a learning rate of 100.0 and default setting of all other hyperparameters. What do you observe ? [3 points]

- [ ] Cross Entropy on the training set fluctuates around a large value.

- [ ] Cross Entropy on the training set decreases smoothly but fluctuates around a large value on the validation set.

- [ ] Cross Entropy on the validation set fluctuates around a large value.

- [ ] Cross Entropy on the training set fluctuates wildly and eventually diverges.

3
points

4.

If all weights and biases in this network were set to zero and no training was performed, what will be the average cross entropy on the validation set ? Please provide a numeric answer (three decimal places). [3 points]

Programming Assignment 2: Learning Word Representations.

Quiz, 13 questions

<div style="border:1px solid #000">

1

point

</div>

5.

Train three models each with 50 dimensional embedding space, 200 dimensional hidden layer.

- Model A: Learning rate = 0.001,
- Model B: Learning rate = 0.1
- Model C: Learning rate = 10.0.

Use a momentum of 0.5 and default settings for all other hyperparameters. Which model gives the lowest training set cross entropy after 1 epoch ? [3 points]

○ Model A

○ Model B

○ Model C

<div style="border:1px solid #000">

2

points

</div>

6.

In the models trained in Question 5, which one gives the lowest training set cross entropy after 10 epochs ? [2 points]

○ Model A

○ Model B

○ Model C

3
points

7.

Train each of following models:

- Model A: 5 dimensional embedding, 100 dimensional hidden layer

- Model B: 50 dimensional embedding, 10 dimensional hidden layer

- Model C: 50 dimensional embedding, 200 dimensional hidden layer

- Model D: 100 dimensional embedding, 5 dimensional hidden layer

Use default values for all other hyperparameters.

Which model gives the best training set cross entropy after 10 epochs of training ? [3 points]

○ Model C

○ Model A

○ Model B

○ Model D

2
points

8.

In the models trained in Question 7, which one gives the best validation set cross entropy after 10 epochs of training ? [2 points]

○ Model A

○ Model B

○ Model D

○ Model C

3
points

9.

Train three models each with 50 dimensional embedding space, 200 dimensional hidden layer.

- Model A: Momentum = 0.0

- Model B: Momentum = 0.5

- Model C: Momentum = 0.9

Use the default settings for all other hyperparameters. Which model gives the lowest training set cross entropy after 5 epochs ? [3 points]

- ○ Model B

- ○ Model A

- ○ Model C

2
points

10.

Train a model with 50 dimensional embedding layer and 200 dimensional hidden layer for 10 epochs. Use default values for all other hyperparameters.

Which words are among the 10 closest words to the word 'could'. [2 points]

- ☐ 'should'

- ☐ 'might'

- ☐ 'some'

- ☐ 'house'

Programming Assignment 2: Learning Word Representations.

11.

In the model trained in Question 10, why is the word 'percent' close to 'dr.' even though they have very different contexts and are not expected to be close in word embedding space? [2 points]

○ Both words occur too frequently.

○ Both words occur very rarely, so their embedding weights get updated very few times and remain close to their initialization.

○ The model is not capable of separating them in embedding space, even if it got a much larger training set.

○ We trained the model with too large a learning rate.

12.

In the model trained in Question 10, why is 'he' close to 'she' even though they refer to completely different genders? [2 points]

○ They differ by only one letter.

○ They often occur close by in sentences.

○ Both words occur very rarely, so their embedding weights get updated very few times and remain close to their initialization.

○ The model does not care about gender. It puts them close because if 'he' occurs in a 4-gram, it is very likely that substituting it by 'she' will also make a sensible 4-gram.

Programming Assignment 2: Learning Word Representations.

13.

In conclusion, what kind of words does the model put close to each other in embedding space. Choose the **most** appropriate answer. [3 points]

○ Words that are such that if one word occurs in a 4-gram replacing it with the other also creates a sensible 4-gram.

○ Words that occur close to each other (within three words to the left or right) in many sentences.

○ Words that belong to similar topics. A topic is a semantic categorization (like 'sports', 'art', 'business', 'computers' etc).

○ Words that have a lot of letters in common.

☐ I understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera's Honor Code

> cy

Submit Quiz