Programming Assignment 2: Learning Word Representations.

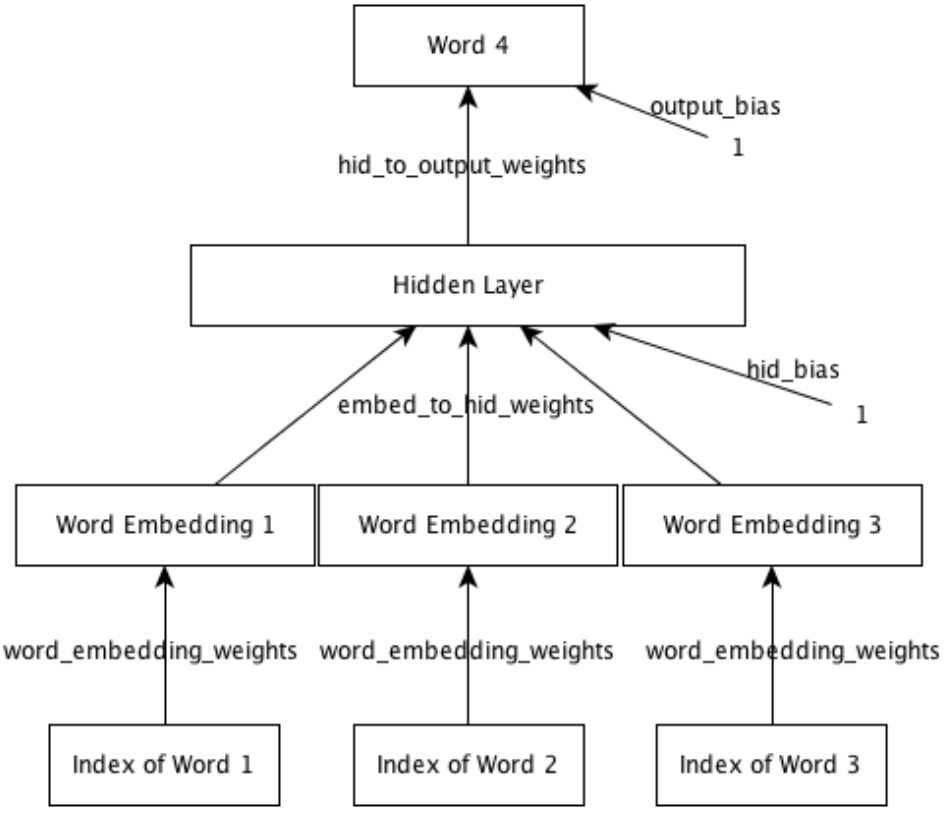
Quiz, 13 questions

1 point

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

Or each file individually:

Programming Assignment 2: Learning Word Representations.

Quiz, 13 questions

- 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 training 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

3.959

Programming Assignment 2: Learning Word Representations.

Quiz, 13 questions

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 validation set fluctuates wildly and eventually diverges.
- [ ] Cross Entropy on the training set fluctuates wildly and eventually diverges.
- [x] Cross Entropy on the validation set fluctuates around a large value.
- [x] Cross Entropy on the training set fluctuates around a large value.

Average Training CE 27.878
Finished Training.
Final Training CE 27.878
Final Validation CE 24.787
Final Test CE 24.813
Training took 579.48 seconds

---

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]

Enter answer here

5.521
for everything

---

1 point

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

A: Finished Training.
Final Training CE 4.432
Final Validation CE 4.386
Final Test CE 4.393
Training took 43.83 seconds

B: Finished Training.
Final Training CE 3.963
Final Validation CE 3.315
Final Test CE 3.315
Training took 43.16 seconds

C: Finished Training.
Final Training CE 4.701
Final Validation CE 4.662
Final Test CE 4.668
Training took 42.10 seconds

Model A

Model C

Model B

---

2 points

A: Final Training CE 4.378
Final Validation CE 4.380
Final Test CE 4.386
Training took 405.74 seconds

B: Finished Training.
Final Training CE 2.535
Final Validation CE 2.603
Final Test CE 2.616
Training took 423.42 seconds

C: Finished Training.
Final Training CE 4.665
Final Validation CE 4.662
Final Test CE 4.668
Training took 421.68 seconds

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

A: Final Training CE 2.811
Final Validation CE 2.831
Final Test CE 2.838
Training took 230.74 seconds

B: Final Training CE 3.020
Final Validation CE 3.033
Final Test CE 3.027
Training took 190.12 seconds

C: Final Training CE 2.536
Final Validation CE 2.603
Final Test CE 2.616
Training took 436.80 seconds

7.

Train each of following models:

D: Final Training CE 3.229
Final Validation CE 3.242
Final Test CE 3.234
Training took 224.02 seconds

- 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 B

-> Model C

Model D

Model A

Programming Assignment 2: Learning Word Representations.

Quiz, 13 questions

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 B
- -> Model C
- Model A
- Model D

---

3 points

A: Final Training CE 3.984
Final Validation CE 3.941
Final Test CE 3.945
Training took 218.80 seconds

B: Final Training CE 3.318
Final Validation CE 3.249
Final Test CE 3.247
Training took 205.79 seconds

C: Final Training CE 2.715
Final Validation CE 2.714
Final Test CE 2.722
Training took 205.65 seconds

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 A
- -> Model C

C: Final Training CE 2.715
Final Validation CE 2.714
Final Test CE 2.722
Training took 205.65 seconds

- Model B

---

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 'day'. [2 points]

- [x] 'week'
- [ ] 'today'
- [ ] 'during'
- [x] 'night'

```
octave:41> word_distance('day', 'week', model)
ans =  2.1149
octave:42> word_distance('day', 'today', model)
ans =  3.5172
octave:43> word_distance('day', 'during', model)
ans =  4.1623
octave:44> word_distance('day', 'night', model)
ans =  2.2594

octave:45> display_nearest_words('day', model, 10)
week 2.11
night 2.26
days 2.32
year 2.51
case 2.99
ago 3.10
season 3.12
street 3.14
school 3.19
center 3.20
```

2 points

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]

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

2 points

12.

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

- -> 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.

- 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.
- They differ by only one letter.

---

3 points

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 belong to similar topics. A topic is a semantic categorization (like 'sports', 'art', 'business', 'computers' etc).
- -> Words that can be substituted for one another and still make up a sensible 4-gram.
- Words that occur close to each other (within three words to the left or right) in many sentences.
- Words that occur close in an alphabetical sort.

I, **Peter Myer Nore**, 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

Submit Quiz