CSC321: Assignment 1

Due on February 7, 2018

Wenxin Chen - 1002157676

February 8, 2018

1 Neural Network

1.1 Trainable Parameters

The weights from input to embedding layer contain 16×250 parameters, as each word maps to a 16-dimension vector.

The embedded layer outputs a 3×16 matrix which is fed into a 128 unit hidden layer, requiring $3 \times 16 \times 128$ weights plus 128 bias parameters.

Lastly the hidden layer outputs its 128 units to each of the 250 units in the output layer, each unit corresponding to a word in the vocabulary. This requires 128×250 weights. A bias parameter with 250 values is also present.

The total number of trainable parameters is then

$$N = 16 \times 250 + 3 \times 16 \times 128 + 128 + 128 \times 250 + 250 = 42\,522$$

1.2 N-Gram Model

If the N-Gram Model was used, and the counts of every possible 4-gram were stored explicitly, there would be

$$N = 250^4 = 3\,906\,250\,000$$

counts, corresponding to all of the possible 4-gram combinations.

2 Part 2

2.1 Output of print_gradients

```
loss_derivative[2, 5] 0.0013789153741
loss_derivative[2, 121] -0.999459885968
loss_derivative[5, 33] 0.000391942483563
loss_derivative[5, 31] -0.708749715825

param_gradient.word_embedding_weights[27, 2] -0.298510438589
param_gradient.word_embedding_weights[43, 3] -1.13004162742
param_gradient.word_embedding_weights[22, 4] -0.211118814492
param_gradient.word_embedding_weights[2, 5] 0.0

param_gradient.embed_to_hid_weights[10, 2] -0.0128399532941
param_gradient.embed_to_hid_weights[15, 3] 0.0937808780803
param_gradient.embed_to_hid_weights[30, 9] -0.16837240452
param_gradient.embed_to_hid_weights[35, 21] 0.0619595914046

param_gradient.hid_bias[10] -0.125907091215
param_gradient.hid_bias[20] -0.389817847348

param_gradient.output_bias[0] -2.23233392034
param_gradient.output_bias[1] 0.0333102255428
param_gradient.output_bias[2] -0.743090094025
param_gradient.output_bias[3] 0.162372657748
```

3 Part 3

3.1 Model Predictions

Below are some examples of predictions made by the model:

```
government of united ? Prob: 0.21415
government of united states Prob: 0.10061
government of united own Prob: 0.04361
government of united said Prob: 0.03280
government of united life Prob: 0.02852
government of united work Prob: 0.02605
government of united . Prob: 0.02530
government of united do Prob: 0.02258
government of united did Prob: 0.01991
government of united left Prob: 0.01930

city of new york Prob: 0.96524
city of new , Prob: 0.00729
city of new . Prob: 0.00635
city of new year Prob: 0.00308
city of new season Prob: 0.00179
city of new world Prob: 0.00153
city of new life Prob: 0.00140
city of new home Prob: 0.00087
city of new ? Prob: 0.00083
city of new one Prob: 0.00069

life in the world Prob: 0.12110
life in the united Prob: 0.07956
life in the end Prob: 0.07401
life in the street Prob: 0.05815
life in the game Prob: 0.04333
life in the country Prob: 0.04119
life in the first Prob: 0.03422
life in the house Prob: 0.03021
life in the right Prob: 0.02993
life in the last Prob: 0.02247

he is the best Prob: 0.17525
he is the same Prob: 0.12968
he is the only Prob: 0.08094
he is the first Prob: 0.05676
he is the end Prob: 0.04158
he is the money Prob: 0.03516
he is the last Prob: 0.03272
he is the right Prob: 0.03045
he is the president Prob: 0.02512
```

```
he is the team Prob: 0.02216

what is our world Prob: 0.09128
what is our money Prob: 0.07374
what is our family Prob: 0.05450
what is our life Prob: 0.04622
what is our way Prob: 0.04582
what is our first Prob: 0.03664
what is our country Prob: 0.03338
what is our business Prob: 0.03269
what is our best Prob: 0.02926
what is our day Prob: 0.02465

how do you do Prob: 0.18523
how do you know Prob: 0.16254
how do you get Prob: 0.15087
how do you want Prob: 0.07165
how do you make Prob: 0.06881
how do you see Prob: 0.04685
how do you ? Prob: 0.04126
how do you have Prob: 0.03285
how do you work Prob: 0.03112
how do you think Prob: 0.03038

what a good time Prob: 0.08644
what a good team Prob: 0.07867
what a good one Prob: 0.05375
what a good ? Prob: 0.04282
what a good year Prob: 0.03724
what a good , Prob: 0.03252
what a good for Prob: 0.02811
what a good . Prob: 0.02611
what a good is Prob: 0.02542
what a good man Prob: 0.02505

i like the best Prob: 0.12860
i like the people Prob: 0.12509
i like the money Prob: 0.05799
i like the children Prob: 0.04692
i like the other Prob: 0.04362
i like the same Prob: 0.02910
i like the team Prob: 0.02467
i like the world Prob: 0.02337
i like the police Prob: 0.02318
i like the united Prob: 0.02180
```

"What is our" was followed by only the word "business" one time in the training set.

"What a good" was followed by only "team" (3 times) and "time" (1 time) in the training set.

3.2 Word Plot

A cluster can be seen containing the words "my", "our", "yours", "his", "theirs", and "its". All of these words are related as they are all possessive terms.

Another cluster can be seen containing the words "day", "night", "week", "year", and "days", which all describe periods of time.

Another cluster can be seen containing the words "few", "several", "many", "three", "four", "five", "two", which all describe quantities of objects.

Lastly, a cluster can be seen containing the words "should", "could", "would", "can", "might", "will", "may", which are all modal verbs.

3.3 New York

The words "new" and "york" are fairly separated in the learned representation (d = 3.53). Compare this with words such as "should" and "could" (d = 1.67) or "he" and "she" (d = 0.55). Even the seemingly unrelated pair "year" and "these" are separated by a smaller distance (d = 3.46). This is due to the nature of the learned representation: as observed previously, it groups words that are similar in usage- i.e. can be switched out for one another in a sentence. "new" and "york" clearly cannot be switched out for one another in a normal sentence, and thus are located quite far away in the learned representation, even though they appear together quite often in regular sentences.

3.4 Government

The word "university" is closer to "government" (d=1.06) than the word "political" is (d=1.41). This is likely due to the nature of the representation once again- "government" can likely be replaced with "university" and result in an acceptable sentence in the majority of cases. However, "political" is an adjective and thus would not make sense when swapped with "government", a noun, in most sentences. Thus, "university" is closer to "government" in the representation.