

csc321 a1

Guanxiong Liu

February 2018

Part 1:

1.

The total number of trainable parameters is $250 \times 16 + 128 \times 3 \times 16 + 128 + 250 \times 128 + 250 = 42522$. `hid_to_output_weights` have the largest number of trainable parameters.

2.

This table would have $250^3 * 250 = 3.90625 * 10^9$ entries.

Part 2:

```
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
```

Part 3:

1.

model.predict_next_word("government","of","united") predicts the next word, house with the highest probability. It is not sensible.
 model.predict_next_word("city","of","new") predicts the next word, york with the highest probability. It is very sensible.
 model.predict_next_word("life","in","the") predicts the next word, world with the highest probability. It is very sensible. Also, the other 9 words make sense as well.
 model.predict_next_word("he","is","the") predicts the next word, best with the highest probability. It is sensible.

"home is the best" is not occurred in the dataset but model.predict_next_word("home","is","the") predicts the next word, best with the highest probability. It is very plausible.

2.

The words in each cluster have the same part of speech.

3.

```
model.display_nearest_words("new")
old: 2.5535126462
big: 2.55945271961
white: 2.73984378399
political: 2.88484265523
your: 3.006437714
several: 3.01906192705
next: 3.02475033797
national: 3.10155784802
good: 3.13521094039
its: 3.14471278895
```

```
model.display_nearest_words("york")
city: 1.124827517
general: 1.15772178628
?: 1.16935145872
company: 1.17931845857
state: 1.18097730112
university: 1.18888567523
national: 1.19510942647
former: 1.20973434783
public: 1.22121500361
west: 1.22634066971
```

```
model.word_distance("new","york")
Out[35]: 3.604173247205491
```

They are not close together. The closer the distance, the more common they are. "york" are more likely the consecutive word of "new", so they are not

consider to be close together.

4.

```
model.word_distance("government","political")  
Out[36]: 1.4257880895493982
```

```
model.word_distance("government","university")  
Out[37]: 1.1542781681862873
```

("government","university") is closer because "government" and "university" are both noun while "political" is adjective. Thus, "political" will not belong to the same cluster as "government".