CS 224N Assignment #1

# 1 Softmax

## (a) prove:

For each ,

Hence,

## (b) python code

# 2 Neural Network Basics

## (a) derive:

Denote

We have

## (b) derive:

We have already known that

Assume that, for and ,

When ,

When ,

hence,

## (c) derive:

We have already known that

From 1 (a), we have

From 2 (a), we have

From 2 (b), we have

From 2 (c), we have

Assume

Here, we denote as dot product and denote as element-wise product.

Since is an element-wise function, hence before is an element-wise product.

Check the dimension by analysis.

Since is a scalar, must have the same dimension as

Assume the input is , output is , hidden units is , then

* has dimension
* has dimension
* has dimension
* has dimension
* has dimension

Hence, the dimension of has dimension

## (d) answer:

The parameters we need are .

Total number is

## (e) python code

## (f) python code

## (g) python code

# 3 word2vec

## (a) derive

When ,

The gradient of , w.r.t.

Assume is a column vector, and is a row vector,

## (b) derive

The gradient of , w.r.t.

* When ,
* When ,

## (c) derive:

The gradient of , w.r.t.

Where

Hence

The gradient of , w.r.t.

* When ,

Where

Hence,

* When ,

Where

Hence,

* When ,

## (d) derive:

For , we have

For , we have

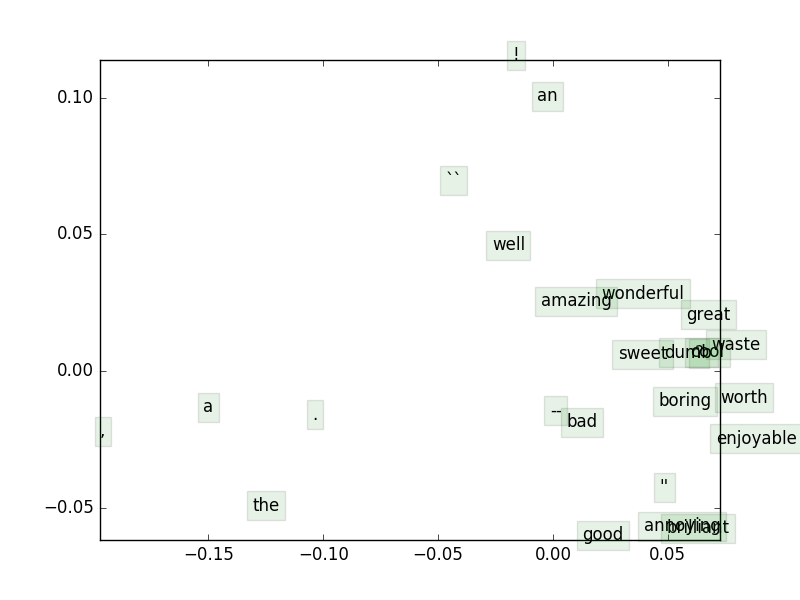
## (e) python code

## (f) python code

## (g) answer:

From the following picture, we can observe:

* Words with similar meaning cluster together.
* Words with different meaning are far from each other.
* Words(Strings) with different part of speech are far from each other.



## (h) python code

# 4 Sentiment Analysis

## (a) python code

## (b) answer:

Regularization prevents model from overfitting to the training set.

## (c) python code

maxVal = results[0]["dev"]

idx = 0

for i in xrange(len(results)):

if results[i]["dev"] > maxVal:

maxVal = results[i]["dev"]

idx = i

bestResult = results[idx]

## (d) answer:

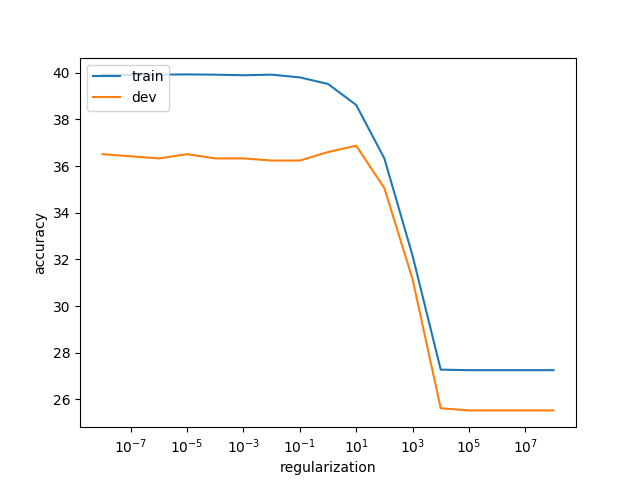
Why does the pretrained vectors perform better?

* The sample set is larger.
* The sample set comes from Wikipedia, which is more standard.
* GloVe algorithm Can capture complex patterns beyond word similarity.

|  |  |
| --- | --- |
| --yourvectors | --pretrained |
| Reg Train Dev Test  1.00E-08 31.016 32.516 30.452  1.00E-07 31.016 32.516 30.407  1.00E-06 31.016 32.516 30.407  1.00E-05 31.016 32.516 30.452  1.00E-04 31.016 32.698 30.362  1.00E-03 31.156 32.698 30.271  1.00E-02 30.946 32.334 29.910  1.00E-01 30.290 31.880 29.819  1.00E+00 28.897 29.609 27.149  1.00E+01 27.247 25.522 23.077  1.00E+02 27.247 25.522 23.032  1.00E+03 27.247 25.522 23.032  1.00E+04 27.247 25.522 23.032  1.00E+05 27.247 25.522 23.032  1.00E+06 27.247 25.522 23.032  1.00E+07 27.247 25.522 23.032  1.00E+08 27.247 25.522 23.032 | Reg Train Dev Test  1.00E-08 39.923 36.512 37.014  1.00E-07 39.911 36.421 36.968  1.00E-06 39.923 36.331 36.968  1.00E-05 39.934 36.512 37.014  1.00E-04 39.923 36.331 37.014  1.00E-03 39.899 36.331 37.104  1.00E-02 39.923 36.240 37.195  1.00E-01 39.806 36.240 37.149  1.00E+00 39.525 36.603 37.330  1.00E+01 38.624 36.876 37.692  1.00E+02 36.330 35.059 35.701  1.00E+03 32.163 31.153 30.588  1.00E+04 27.271 25.613 23.122  1.00E+05 27.247 25.522 23.032  1.00E+06 27.247 25.522 23.032  1.00E+07 27.247 25.522 23.032  1.00E+08 27.247 25.522 23.032 |
| Best regularization value: 1.00E-04  Test accuracy (%): 30.361991 | Best regularization value: 1.00E+01  Test accuracy (%): 37.692308 |

## (e) answer:

With the increment of regularization value, the accuracy increases slightly. After regularization value passes 10, the accuracy decreases significantly. The overly large value of regularization takes over the model and reduces the effectiveness of the model.

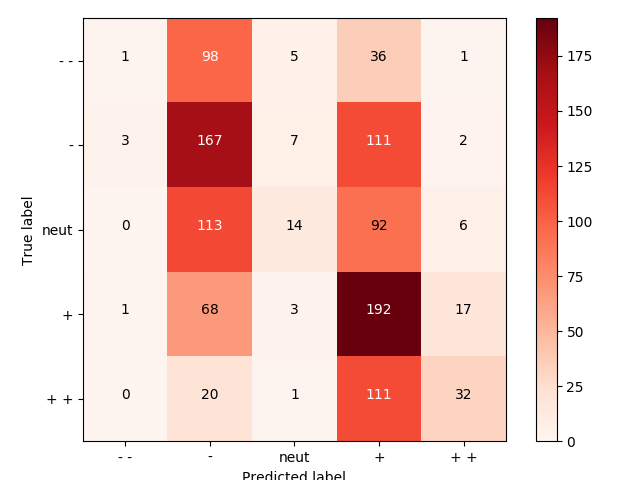


## (f) answer:

The predicted result is not absolutely accurate.

The prediction of (++) and (--) are relatively more accurate.

The prediction of (--), (neut) and (++) are not so accurate.



## (g) answer:

|  |  |  |  |
| --- | --- | --- | --- |
|  | True | Predicted | Sentence |
| (1) | 3 | 1 | we know the plot 's a little crazy, but it held my interest from start to finish. |
| (2) | 4 | 1 | manages to transcend the sex, drugs and show-tunes plot into something far richer. |
| (3) | 1 | 3 | a subject like this should inspire reaction in its audience; the pianist does not. |

(1) The word “but” in this sentence change the meaning of the first half.

(2) The second sentence maybe doesn’t take the verb “transcend” into consideration.

(3) The latter half of the third sentence turns over the sentiment of the former half.