

## Week2 Assignment

Ma Yubo

### Part One

$$(1) - \sum_w y_w \log(\hat{y}_w) = - \sum_{y_w \neq 0} y_w \log(\hat{y}_w) = - y_0 \log(\hat{y}_0) = - \log p(0|c).$$

$$(2) J = - \log p(0|c) = - \log \frac{\exp(u_0^T V_c)}{\sum_w \exp(u_w^T V_c)}$$

$$= -u_0^T V_c + \log \left[ \sum_w \exp(u_w^T V_c) \right].$$

$$\therefore \frac{\partial J}{\partial V_c} = -u_0 + \frac{\sum_w u_w \cdot \exp(u_w^T V_c)}{\sum_w \exp(u_w^T V_c)}.$$

$$= -u_0 + \sum_w \frac{u_w \exp(u_w^T V_c)}{\sum_w \exp(u_w^T V_c)}.$$

$$= -u_0 + \sum_w u_w \log p(w|c).$$

$$= -Uy + U\hat{y} = U(\hat{y} - y).$$

$$(3) \frac{\partial J}{\partial u_w} = -I(w=0) V_c + \frac{\exp(u_w^T V_c) \cdot V_c}{\sum_k \exp(u_k^T V_c)}$$

$$= -y^{(w)} V_c + \hat{y}^{(w)} V_c.$$

$$\Rightarrow \frac{\partial J}{\partial U} = (\hat{y} - y)^T \otimes V_c, \text{ where } \otimes \text{ is the Kronecker product.}$$

$$(4) \sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})x(1 + e^{-x})}, \text{ where } x \text{ is the element-wise multiplication}$$

$$= \sigma(x) \times (1 - \sigma(x))$$

## Part Two

The source codes have uploaded on Github and here's the running result:

```
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

尝试新的跨平台 PowerShell https://aka.ms/powershell

PS D:\Research\Yu Lab\NLP学习班\作业\第二次作业\Week 2 Assignment> python word2vec.py
==== Gradient check for skip-gram with naiveSoftmaxLossAndGradient ====
Gradient check passed!
==== Gradient check for skip-gram with negSamplingLossAndGradient ====
Gradient check passed!

=== Results ===
Skip-Gram with naiveSoftmaxLossAndGradient
Your Result:
Loss: 11.16610900153398
Gradient wrt Center Vectors (dJ/dV):
[[ 0.          0.          0.          ]
 [ 0.          0.          0.          ]
 [-1.26947339 -1.36873189  2.45158957]
 [ 0.          0.          0.          ]
 [ 0.          0.          0.          ]]
Gradient wrt Outside Vectors (dJ/dU):
[[-0.41045956  0.18834851  1.43272264]
 [ 0.38202831 -0.17530219 -1.33348241]
 [ 0.07009355 -0.03216399 -0.24466386]
 [ 0.09472154 -0.04346509 -0.33062865]
 [-0.13638384  0.06258276  0.47605228]]

Expected Result: Value should approximate these:
Loss: 11.16610900153398
Gradient wrt Center Vectors (dJ/dV):
[[ 0.          0.          0.          ]
 [ 0.          0.          0.          ]
 [-1.26947339 -1.36873189  2.45158957]
 [ 0.          0.          0.          ]
 [ 0.          0.          0.          ]]
Gradient wrt Outside Vectors (dJ/dU):
[[-0.41045956  0.18834851  1.43272264]
 [ 0.38202831 -0.17530219 -1.33348241]
 [ 0.07009355 -0.03216399 -0.24466386]
```

```
PS D:\Research\Yu Lab\NLP学习班\作业\第二次作业\Week 2 Assignment> python sgd.py
Running sanity checks...
iter 100: 0.004578
iter 200: 0.004353
iter 300: 0.004136
iter 400: 0.003929
iter 500: 0.003733
iter 600: 0.003548
iter 700: 0.003369
iter 800: 0.003200
iter 900: 0.003040
iter 1000: 0.002888
test 1 result: 8.414836786079764e-10
iter 100: 0.000000
iter 200: 0.000000
iter 300: 0.000000
iter 400: 0.000000
iter 500: 0.000000
iter 600: 0.000000
iter 700: 0.000000
iter 800: 0.000000
iter 900: 0.000000
iter 1000: 0.000000
test 2 result: 0.0
iter 100: 0.041205
iter 200: 0.039181
iter 300: 0.037222
iter 400: 0.035361
iter 500: 0.033593
iter 600: 0.031913
iter 700: 0.030318
iter 800: 0.028802
iter 900: 0.027362
iter 1000: 0.025994
test 3 result: -2.524451035823933e-09
ALL TESTS PASSED
```