Homework 5 Answer

Zhao Wang u0905676

April 24, 2019

1 Paper Problems

1.
$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x} + \frac{\partial z}{\partial y_3}\frac{\partial y_3}{\partial x}$$

$$= 3\cdot \frac{2e^{-x^2-yz}x}{\left(1+e^{-x^2-yz}\right)^2} + (-e^{-x})\cdot \frac{e^{-yz-x^2}z}{\left(1+e^{-x^2-yz}\right)^2} + \cos(x)\cdot \frac{e^{-yz-x^2}y}{\left(1+e^{-x^2-yz}\right)^2}$$

$$= \frac{6e^{-x^2-yz}x - e^{-x^2-x-yz}z + e^{-x^2-yz}y\cos(x)}{\left(e^{-x^2-yz}+1\right)^2}$$

2.
$$z_1^1 = \sigma(w_{01}^1 + w_{11}^1 x_1 + w_{21}^1 x_2) = \sigma(-1 + -2*1 + -3*1) = 0.00247$$

$$z_2^1 = \sigma(w_{02}^1 + w_{12}^1 x_1 + w_{22}^1 x_2) = 0.99753$$

$$z_1^2 = \sigma(w_{01}^2 + w_{11}^2 z_1^1 + w_{21}^2 z_2^1) = 0.01803$$

$$z_2^2 = \sigma(w_{02}^2 + w_{12}^2 z_1^1 + w_{22}^2 z_2^1) = 0.98197$$

$$y = w_{01}^3 + w_{11}^3 z_1^2 + w_{21}^3 z_2^2 = -2.43690$$

3. We get $y = -2.43690$ from question 2, then we have

$$\frac{\partial L}{\partial w_{01}^3} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial w_{01}^3} = (y - y^*) = -3.43690$$

$$\frac{\partial L}{\partial w_{11}^3} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial w_{11}^3} = (y - y^*)(z_1^2) = -0.06198$$

$$\frac{\partial L}{\partial w^3_{21}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial w^3_{21}} = (y - y^*)(z^2_2) = -3.37494$$

$$\frac{\partial L}{\partial w^2_{01}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_1}\frac{\partial z^2_1}{\partial w^2_{01}} = (y - y^*)(w^3_{11})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{01}} = -0.12170$$

$$\frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_1}\frac{\partial z^2_1}{\partial w^2_{11}} = (y - y^*)(w^3_{11})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{11}} = -0.00031$$

$$\frac{\partial L}{\partial w^2_{21}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_1}\frac{\partial z^2_1}{\partial w^2_{21}} = (y - y^*)(w^3_{11})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{21}} = -0.12141$$

$$\frac{\partial L}{\partial w^2_{02}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_2}\frac{\partial z^2_2}{\partial w^2_{02}} = (y - y^*)(w^3_{21})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{02}} = 0.09128$$

$$\frac{\partial L}{\partial w^2_{12}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_2}\frac{\partial z^2_2}{\partial w^2_{12}} = (y - y^*)(w^3_{21})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{12}} = 0.00023$$

$$\frac{\partial L}{\partial w^2_{22}} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z^2_2}\frac{\partial z^2_2}{\partial w^2_{22}} = (y - y^*)(w^3_{21})\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^2_{22}} = 0.09105$$

then,

$$\frac{\partial L}{\partial w^1_{01}} = \frac{\partial L}{\partial y}(\frac{\partial y}{\partial z^2_1}\frac{\partial z^2_1}{\partial w^1_{01}}+\frac{\partial y}{\partial z^2_2}\frac{\partial z^2_2}{\partial w^1_{01}}) = (y-y^*)((w^3_{11})(\frac{\partial z^2_1}{\partial z^1_1}\frac{\partial z^1_1}{\partial w^1_{01}})+(w^3_{21})(\frac{\partial z^2_2}{\partial z^1_1}\frac{\partial z^1_1}{\partial w^1_{01}})) = 0.0010506$$

$$\frac{\partial L}{\partial w^1_{11}} = (y - y^*)((w^3_{11})((w^2_{11})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{11}})) + (w^3_{21})((w^2_{12})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{11}}))) = 0.0010506$$

$$\frac{\partial L}{\partial w^1_{21}} = (y - y^*)((w^3_{11})((w^2_{11})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{21}})) + (w^3_{21})((w^2_{12})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{21}}))) = 0.0010506$$

$$\frac{\partial L}{\partial w^1_{02}} = (y - y^*)((w^3_{11})((w^2_{21})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{02}})) + (w^3_{21})((w^2_{22})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{02}}))) = 0.0015759$$

$$\frac{\partial L}{\partial w^1_{12}} = (y - y^*)((w^3_{11})((w^2_{21})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{12}})) + (w^3_{21})((w^2_{22})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{12}}))) = 0.0015759$$

$$\frac{\partial L}{\partial w^1_{22}} = (y - y^*)((w^3_{11})((w^2_{21})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{22}})) + (w^3_{21})((w^2_{22})(\frac{\partial \sigma}{\partial s}\frac{\partial s}{\partial w^1_{22}}))) = 0.0015759$$

4. (1) Objective function $= \min_{\mathbf{w}} \sum_i^m \log(1 + \exp(-y_i \mathbf{w}^\top \mathbf{x}_i)) + \frac{1}{2}\mathbf{w}^\top \mathbf{w}$

   Gradient of the objective function: $\begin{bmatrix} \sum_i^m -\frac{y_i \mathbf{x}_{i0} e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}{1+e^{-y_i \mathbf{w}^\top \mathbf{x}_i}} + \mathbf{w}_0 \\ \vdots \\ \sum_i^m -\frac{y_i \mathbf{x}_{in} e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}{1+e^{-y_i \mathbf{w}^\top \mathbf{x}_i}} + \mathbf{w}_n \end{bmatrix}$

   (2) The fisrt three steps show below:
   Step 1:
   $\mathbf{w} = [-0.005, 0.010, -0.003, -0.010]$
   $\nabla J(\mathbf{w}) = [-0.5, 1.0, -0.3, -1.0]$

   Step 2:
   $\mathbf{w} = [0.00494, 0.02987, 0.01687, -0.01994]$
   $\nabla J(\mathbf{w}) = [-1.98576, -3.97151, -3.97151, 1.98576]$

   Step 3:
   $\mathbf{w} = [0.00104, 0.02934, 0.02334, -0.02252]$
   $\nabla J(\mathbf{w}) = [-1.55479, -0.20730, 2.59131, -1.03652]$

2 Practice

1. See Github repo

2. (a) We have $\gamma_0 = 0.0000001, d = 0.000001$

var	Training Error	Testing Error
0.01	0.312	0.296
0.1	0.292	0.288
0.5	0.297	0.286
1	0.25	0.25
3	0.317	0.294
5	0.298	0.286
10	0.299	0.292
100	0.241	0.258

   (b) We have $\gamma_0 = 0.0000001, d = 0.000001$

var	Training Error	Testing Error
0.01	0.287	0.276
0.1	0.225	0.242
0.5	0.288	0.282
1	0.24	0.258
3	0.291	0.278
5	0.241	0.254
10	0.250	0.252
100	0.305	0.308

(c) We use the error result from the ML estimation to minus the MAP estimation and get difference:

Train_Diff	Testing_Diff
-0.009	-0.02
-0.063	-0.046
0.002	-0.004
-0.01	0.008
-0.003	-0.016
-0.045	-0.032
-0.042	-0.04
0.047	0.05

In summary, the ML estimation performs better than the MAP estimation. Different from the hyperparameter C acts on the learning rate, v adds variance to the gradient directly so v can scale up and down.

3.

(a) Our result gets:

```
Layer 0 gradient:
[[-3.43689523]
 [-0.061967  ]
 [-3.37492823]]

Layer 1 gradient:
[[-0.12169947  0.09127461]
 [-0.00030092  0.00022569]
 [-0.12139856  0.09104892]]

Layer 2 gradient:
[[ 0.00105061  0.00157591]
 [ 0.00105061  0.00157591]
 [ 0.00105061  0.00157591]]
```

(b) d = 0.001

w	Training	Testing
5	0.688	0.800
10	1.032	1.200
25	0.344	0.400
50	0.344	0.400
100	44.610	44.200

(c) After initialize all the weights with 0, we get:

w	Training	Testing
5	44.610	44.200
10	44.610	44.200
25	44.610	44.200
50	44.610	44.200
100	44.610	44.200

In summary, both training and testing become the same value and very large.

(d) The logistic regression and SVM both performs better than the neural network.

(e) I tried this part with two different editions of python (2 & 3.6) because I cannot test with python 2 on cade lab. Tensorflow is not allowed to install on cade machine so I write a 3.6 edition for data generating and test.

"tanh" & "Xavier" initialization

w/d	Training	Testing
5/3	0.958	0.952
5/5	0.791	0.796
5/9	0.995	1.000
10/3	0.987	0.982
10/5	0.998	0.996
10/9	1.000	1.000
25/3	0.982	0.984
25/5	0.993	0.990
25/9	1.000	0.998
50/3	0.984	0.980
50/5	1.000	1.000
50/9	1.000	1.000
100/3	0.997	0.996
100/5	1.000	1.000
100/9	1.000	1.000

"RELU" & "he" initialization

w/d	Training	Testing
5/3	0.958	0.952
5/5	0.791	0.796
5/9	0.995	1.000
10/3	0.987	0.982
10/5	0.998	0.996
10/9	1.000	1.000
25/3	0.982	0.984
25/5	0.993	0.990
25/9	1.000	0.998
50/3	0.984	0.980
50/5	1.000	1.000
50/9	1.000	1.000
100/3	0.997	0.996
100/5	1.000	1.000
100/9	1.000	1.000

4. See Github repo