CS 189 HW7 Write-Up

Qingyang Zhao

TOTAL POINTS

60 / 80

QUESTION 1

1 Getting Started 10 / 10

✓ + **10 pts** **Correct**

+ **0 pts** Wrong/Missing

QUESTION 2

Backpropogation Algorithm for Neural Networks 60 pts

2.1 a 0 / 0

✓ + **0 pts** **--**

2.2 b 10 / 10

✓ + **10 pts** **Correct**

+ **8 pts** Almost Correct

+ **5 pts** Medium Progress

+ **2 pts** Trivial Progress

+ **0 pts** Wrong/Missing

2.3 c 0 / 0

✓ + **0 pts** **--**

2.4 d 0 / 0

✓ + **0 pts** **--**

2.5 e 10 / 10

✓ + **10 pts** **Correct**

+ **8 pts** Minor calculation error

+ **5 pts** Correct chain rule

+ **0 pts** Incorrect

+ **6 pts** Correct code, no detailed solution with steps written out.

2.6 f 0 / 10

+ **10 pts** Correct

+ **8 pts** Mostly Correct (misnamed a variable, forgot a negative sign, off by a constant)

+ **7 pts** minor conceptual mistake

+ **5 pts** update of either bias completely wrong or weight completely wrong or

+ **3 pts** only completed one part (either bias or weight) of code and that part is mostly correct

+ **2 pts** No code, but correct concept via equation

✓ + **0 pts** **Incorrect or blank**

2.7 g 10 / 10

✓ + **3.4 pts** **tanh**

✓ + **3.3 pts** **linear**

✓ + **3.3 pts** **relu**

+ **0 pts** Incorrect

2.8 h 10 / 10

✓ + **10 pts** **Correct/Reasonable Numbers**

+ **8 pts** Minor Errors

+ **5 pts** Halfway There

+ **2 pts** Correct Direction

+ **0 pts** Incorrect/Blank

2.9 i 0 / 10

+ **10 pts** Correct

+ **8 pts** Minor Error

+ **5 pts** Partial

✓ + **0 pts** **Incorrect/Missing**

2.10 j 0 / 0

+ **10 pts** Correct/Close

+ **8 pts** Some Error

+ **5 pts** Partial

✓ + **0 pts** **Incorrect/Missing**

2.11 k 0 / 0

+ **10 pts** Graphs and Justification

+ **4 pts** Claim without justification

✓ + **0 pts** **No Answer**

QUESTION 3

3 Your Own Question 10 / 10

✓ + **10 pts** **Question and Solution**

+ **0 pts** Incorrect or blank

HW7-1

a) Names — Email Address

Qi Wan — jhun0324@berkeley.edu

Description of Team: Best Group Ever

How did I work?

Comments:

b) I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

Qingyang Zhao

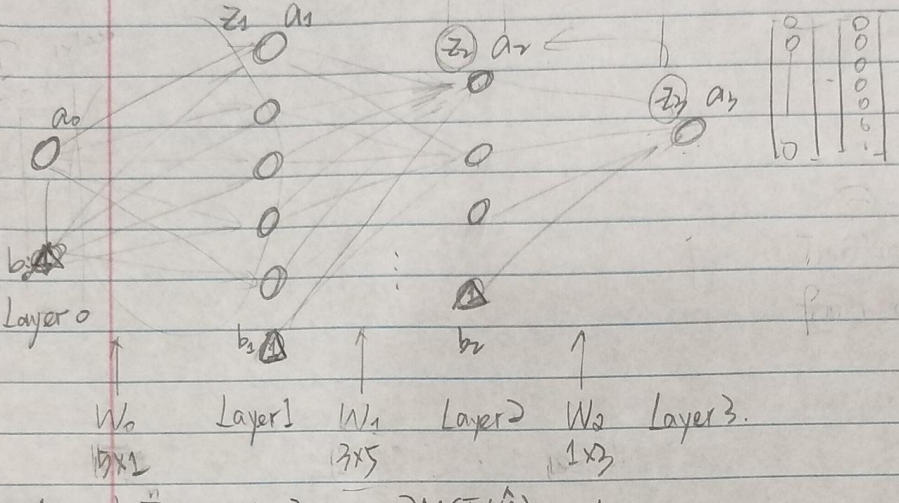
1 Getting Started 10 / 10

✓ + 10 pts Correct

+ 0 pts Wrong/Missing

CS289A HW07 $l_i(\vec{a}) = f(W_i \vec{a} + \vec{b}_i)$ $i = 0, 1, \dots, \ell$.

a)



b) $MSE(\hat{y}) = \frac{1}{2}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2^2$ $\quad \frac{\partial MSE(\hat{y})}{\partial \hat{y}} = -\frac{1}{2} \cdot 2(y - \hat{y})$

c) $\sigma_{linear}(z) = z$ $\quad \sigma'_{linear}(z) = 1$

$\sigma_{ReLU} = \begin{cases} 0 & z \geq 0 \\ z & \text{otherwise} \end{cases}$ $\quad \sigma'_{ReLU} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$

$\sigma_{tanh} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $\quad \sigma'_{tanh} = 1 - \sigma_{tanh}^2$

d) code

e) Ultimate Goal: (learning rate)

$$W_{ij}^{(\ell)} := W_{ij}^{(\ell)} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b)$$

$$b_i^{(\ell)} := b_i^{(\ell)} - \alpha \frac{\partial}{\partial b_i} J(W, b)$$

Overall Gradient for layer ℓ $\quad i \to$ node $\quad j \to$ next layer's node.

$$\frac{\partial J(W,b)}{\partial W_{ij}^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

$$\frac{\partial J(W,b)}{\partial b_i^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

for output layer n_ℓ.

$$\delta_i^{(n_\ell)} = \frac{\partial}{\partial z_i^{(n_\ell)}} J(W,b) = \boxed{-(y - a_i^{(n_\ell)}) \cdot f'(z_i^{(n_\ell)})}$$

for $\ell = n_\ell - 1$ $\quad \delta_i^{(\ell)} = \frac{\partial}{\partial z_i^{(\ell)}} J(W,b) = \sum_i \left(\frac{\partial J(W,b)}{\partial z_i^{(n_\ell)}} \cdot \frac{\partial z_i^{(n_\ell)}}{\partial a_i^{(n_\ell - 1)}} \right) \frac{\partial a_i^{(n_\ell - 1)}}{\partial z_i^{(n_\ell - 1)}} = \sum_i \left(\delta_i^{(n_\ell)} \cdot W_{ij}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)})$

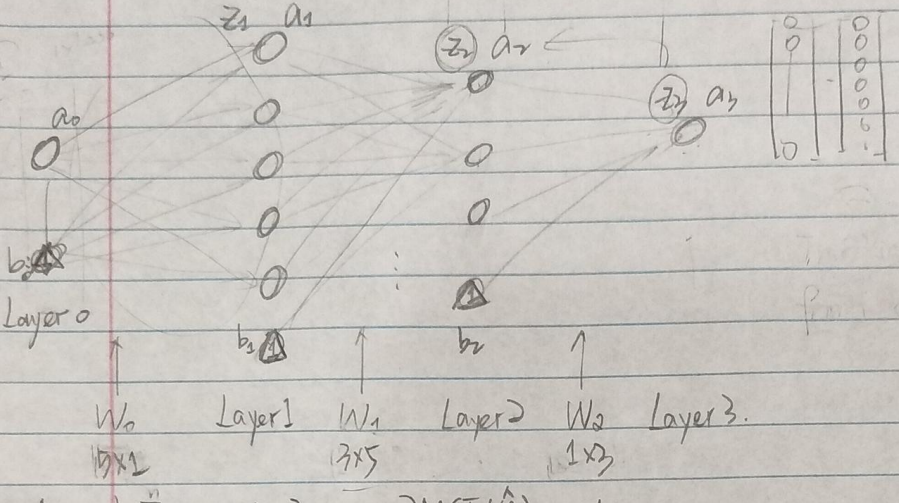
ℓ $\quad \delta_i^{(\ell)} = \sum_{i'} \left(\frac{\partial J(W,b)}{\partial z_{i'}^{(\ell+1)}} \cdot \frac{\partial z_{i'}^{(\ell+1)}}{\partial a_i^{(\ell)}} \right) \cdot \frac{\partial a_i^{(\ell)}}{\partial z_i^{(\ell)}} = \sum_{i'} \left(\delta_{i'}^{(\ell+1)} \cdot W_{ji'}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)}) \Rightarrow \boxed{\delta^{(\ell)} = (W^{(\ell)})^T \cdot \delta^{(\ell+1)} \bullet \sigma'(z^{(\ell)})}$

2.1 a 0 / 0

✓ + 0 pts --

CS289A HW07 $l_i(\vec{a}) = f(W_i \vec{a} + \vec{b}_i)$ $i = 0, 1, \dots, \ell$.

a)



b) $MSE(\hat{y}) = \frac{1}{2}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2^2$ $\quad \frac{\partial MSE(\hat{y})}{\partial \hat{y}} = -\frac{1}{2} \cdot 2(y - \hat{y})$

c) $\sigma_{linear}(z) = z$ $\quad \sigma'_{linear}(z) = 1$

$\sigma_{ReLU} = \begin{cases} 0 & z \geq 0 \\ z & \text{otherwise} \end{cases}$ $\quad \sigma'_{ReLU} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$

$\sigma_{tanh} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $\quad \sigma'_{tanh} = 1 - \sigma_{tanh}^2$

d) code

e) Ultimate Goal: (learning rate)

$$W_{ij}^{(\ell)} := W_{ij}^{(\ell)} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b)$$

$$b_i^{(\ell)} := b_i^{(\ell)} - \alpha \frac{\partial}{\partial b_i} J(W, b)$$

Overall Gradient for layer ℓ $\quad i \to$ node $\quad j \to$ next layer's node.

$$\frac{\partial J(W,b)}{\partial W_{ij}^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

$$\frac{\partial J(W,b)}{\partial b_i^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

for output layer n_ℓ.

$$\delta_i^{(n_\ell)} = \frac{\partial}{\partial z_i^{(n_\ell)}} J(W,b) = \boxed{-(y - a_i^{(n_\ell)}) \cdot f'(z_i^{(n_\ell)})}$$

for $\ell = n_\ell - 1$ $\quad \delta_i^{(\ell)} = \frac{\partial}{\partial z_i^{(\ell)}} J(W,b) = \sum_i \left(\frac{\partial J(W,b)}{\partial z_i^{(n_\ell)}} \cdot \frac{\partial z_i^{(n_\ell)}}{\partial a_i^{(n_\ell - 1)}} \right) \frac{\partial a_i^{(n_\ell - 1)}}{\partial z_i^{(n_\ell - 1)}} = \sum_i \left(\delta_i^{(n_\ell)} \cdot W_{ij}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)})$

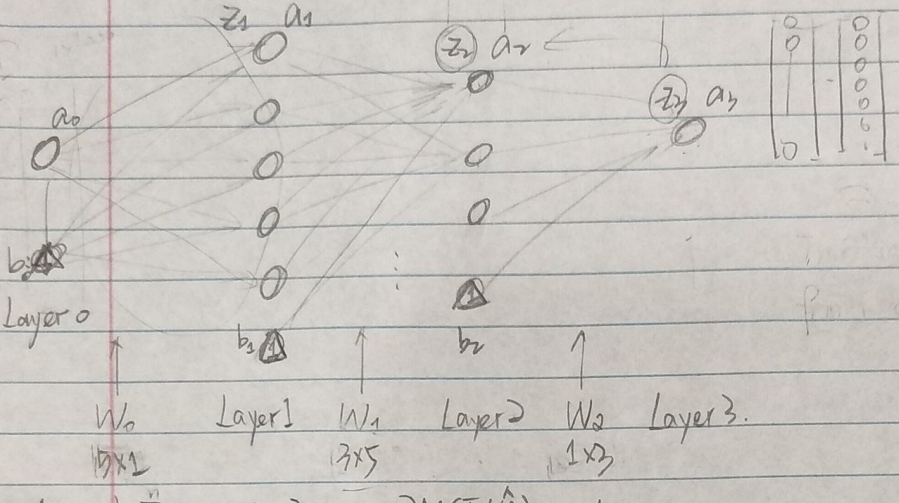
ℓ $\quad \delta_i^{(\ell)} = \sum_{i'} \left(\frac{\partial J(W,b)}{\partial z_{i'}^{(\ell+1)}} \cdot \frac{\partial z_{i'}^{(\ell+1)}}{\partial a_i^{(\ell)}} \right) \cdot \frac{\partial a_i^{(\ell)}}{\partial z_i^{(\ell)}} = \sum_{i'} \left(\delta_{i'}^{(\ell+1)} \cdot W_{ji'}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)}) \Rightarrow \boxed{\delta^{(\ell)} = (W^{(\ell)})^T \cdot \delta^{(\ell+1)} \bullet \sigma'(z^{(\ell)})}$

2.2 b **10 / 10**

- ✓ **+ 10 pts Correct**
- **+ 8 pts** Almost Correct
- **+ 5 pts** Medium Progress
- **+ 2 pts** Trivial Progress
- **+ 0 pts** Wrong/Missing

CS289A HW07 $l_i(\vec{a}) = f(W_i \vec{a} + \vec{b}_i)$ $i = 0, 1, \dots, \ell$.

a)



b) $MSE(\hat{y}) = \frac{1}{2}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2^2$ $\quad \frac{\partial MSE(\hat{y})}{\partial \hat{y}} = -\frac{1}{2} \cdot 2(y - \hat{y})$

c) $\sigma_{linear}(z) = z$ $\quad \sigma'_{linear}(z) = 1$

$\sigma_{ReLU} = \begin{cases} 0 & z \geq 0 \\ z & \text{otherwise} \end{cases}$ $\quad \sigma'_{ReLU} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$

$\sigma_{tanh} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $\quad \sigma'_{tanh} = 1 - \sigma_{tanh}^2$

d) code

e) Ultimate Goal: (learning rate)

$$W_{ij}^{(\ell)} := W_{ij}^{(\ell)} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b)$$

$$b_i^{(\ell)} := b_i^{(\ell)} - \alpha \frac{\partial}{\partial b_i} J(W, b)$$

Overall Gradient for layer ℓ $\quad i \to$ node $\quad j \to$ next layer's node.

$$\frac{\partial J(W,b)}{\partial W_{ij}^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

$$\frac{\partial J(W,b)}{\partial b_i^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

for output layer n_ℓ.

$$\delta_i^{(n_\ell)} = \frac{\partial}{\partial z_i^{(n_\ell)}} J(W,b) = \boxed{-(y - a_i^{(n_\ell)}) \cdot f'(z_i^{(n_\ell)})}$$

for $\ell = n_\ell - 1$ $\quad \delta_i^{(\ell)} = \frac{\partial}{\partial z_i^{(\ell)}} J(W,b) = \sum_i \left(\frac{\partial J(W,b)}{\partial z_i^{(n_\ell)}} \cdot \frac{\partial z_i^{(n_\ell)}}{\partial a_i^{(n_\ell - 1)}} \right) \frac{\partial a_i^{(n_\ell - 1)}}{\partial z_i^{(n_\ell - 1)}} = \sum_i \left(\delta_i^{(n_\ell)} \cdot W_{ij}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)})$

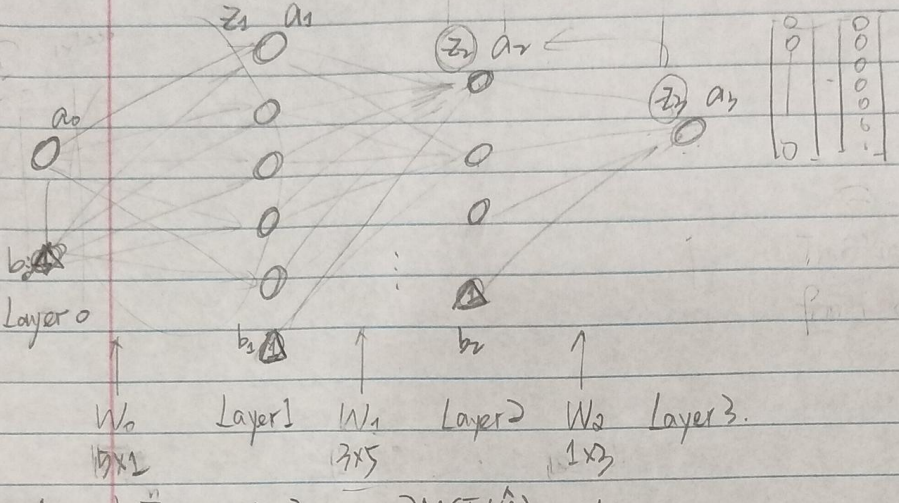
ℓ $\quad \delta_i^{(\ell)} = \sum_{i'} \left(\frac{\partial J(W,b)}{\partial z_{i'}^{(\ell+1)}} \cdot \frac{\partial z_{i'}^{(\ell+1)}}{\partial a_i^{(\ell)}} \right) \cdot \frac{\partial a_i^{(\ell)}}{\partial z_i^{(\ell)}} = \sum_{i'} \left(\delta_{i'}^{(\ell+1)} \cdot W_{ji'}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)}) \Rightarrow \boxed{\delta^{(\ell)} = (W^{(\ell)})^T \cdot \delta^{(\ell+1)} \bullet \sigma'(z^{(\ell)})}$

2.3 C 0 / 0

✓ + 0 pts --

CS289A HW07 $l_i(\vec{a}) = f(W_i \vec{a} + \vec{b}_i)$ $i = 0, 1, \dots, \ell$.

a)



b) $MSE(\hat{y}) = \frac{1}{2}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2^2$ $\quad \frac{\partial MSE(\hat{y})}{\partial \hat{y}} = -\frac{1}{2} \cdot 2(y - \hat{y})$

c) $\sigma_{linear}(z) = z$ $\quad \sigma'_{linear}(z) = 1$

$\sigma_{ReLU} = \begin{cases} 0 & z \geq 0 \\ z & \text{otherwise} \end{cases}$ $\quad \sigma'_{ReLU} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$

$\sigma_{tanh} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $\quad \sigma'_{tanh} = 1 - \sigma_{tanh}^2$

d) code

e) Ultimate Goal: (learning rate)

$$W_{ij}^{(\ell)} := W_{ij}^{(\ell)} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b)$$

$$b_i^{(\ell)} := b_i^{(\ell)} - \alpha \frac{\partial}{\partial b_i} J(W, b)$$

Overall Gradient for layer ℓ $\quad i \to$ node $\quad j \to$ next layer's node.

$$\frac{\partial J(W,b)}{\partial W_{ij}^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

$$\frac{\partial J(W,b)}{\partial b_i^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

for output layer n_ℓ.

$$\delta_i^{(n_\ell)} = \frac{\partial}{\partial z_i^{(n_\ell)}} J(W,b) = \boxed{-(y - a_i^{(n_\ell)}) \cdot f'(z_i^{(n_\ell)})}$$

for $\ell = n_\ell - 1$ $\quad \delta_i^{(\ell)} = \frac{\partial}{\partial z_i^{(\ell)}} J(W,b) = \sum_i \left(\frac{\partial J(W,b)}{\partial z_i^{(n_\ell)}} \cdot \frac{\partial z_i^{(n_\ell)}}{\partial a_i^{(n_\ell - 1)}} \right) \frac{\partial a_i^{(n_\ell - 1)}}{\partial z_i^{(n_\ell - 1)}} = \sum_i \left(\delta_i^{(n_\ell)} \cdot W_{ij}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)})$

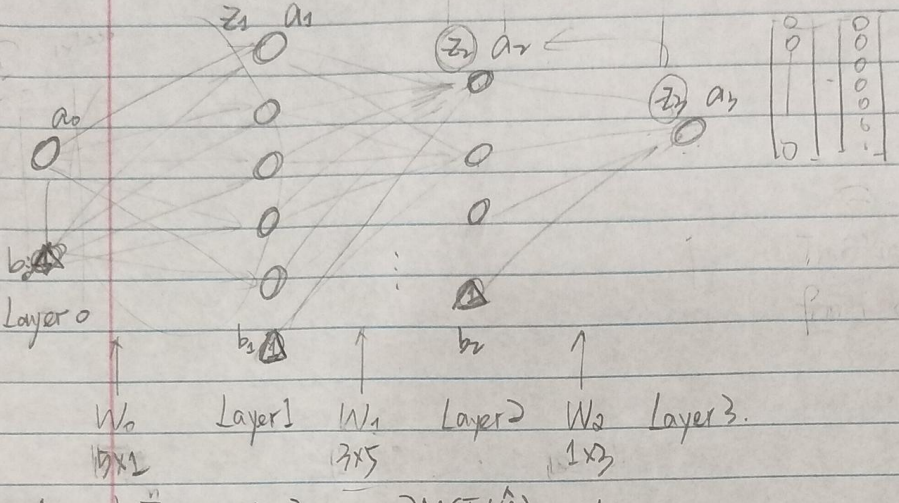
ℓ $\quad \delta_i^{(\ell)} = \sum_{i'} \left(\frac{\partial J(W,b)}{\partial z_{i'}^{(\ell+1)}} \cdot \frac{\partial z_{i'}^{(\ell+1)}}{\partial a_i^{(\ell)}} \right) \cdot \frac{\partial a_i^{(\ell)}}{\partial z_i^{(\ell)}} = \sum_{i'} \left(\delta_{i'}^{(\ell+1)} \cdot W_{ji'}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)}) \Rightarrow \boxed{\delta^{(\ell)} = (W^{(\ell)})^T \cdot \delta^{(\ell+1)} \bullet \sigma'(z^{(\ell)})}$

2.4 d 0 / 0

✓ + 0 pts --

CS289A HW07 $l_i(\vec{a}) = f(W_i \vec{a} + \vec{b}_i)$ $i = 0, 1, \dots, \ell$.

a)



b) $MSE(\hat{y}) = \frac{1}{2}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|_2^2$ $\quad \frac{\partial MSE(\hat{y})}{\partial \hat{y}} = -\frac{1}{2} \cdot 2(y - \hat{y})$

c) $\sigma_{linear}(z) = z$ $\quad \sigma'_{linear}(z) = 1$

$\sigma_{ReLU} = \begin{cases} 0 & z \geq 0 \\ z & \text{otherwise} \end{cases}$ $\quad \sigma'_{ReLU} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$

$\sigma_{tanh} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ $\quad \sigma'_{tanh} = 1 - \sigma_{tanh}^2$

d) code

e) Ultimate Goal: (learning rate)

$$W_{ij}^{(\ell)} := W_{ij}^{(\ell)} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b)$$

$$b_i^{(\ell)} := b_i^{(\ell)} - \alpha \frac{\partial}{\partial b_i} J(W, b)$$

Overall Gradient for layer ℓ $\quad i \to$ node $\quad j \to$ next layer's node.

$$\frac{\partial J(W,b)}{\partial W_{ij}^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

$$\frac{\partial J(W,b)}{\partial b_i^{(\ell)}} = \frac{1}{m}\sum_{k=1}^{n} \frac{\partial}{\partial W_{ij}^{(\ell)}} J(W, b; x^{(k)} y^{(k)})$$

for output layer n_ℓ.

$$\delta_i^{(n_\ell)} = \frac{\partial}{\partial z_i^{(n_\ell)}} J(W,b) = \boxed{-(y - a_i^{(n_\ell)}) \cdot f'(z_i^{(n_\ell)})}$$

for $\ell = n_\ell - 1$ $\quad \delta_i^{(\ell)} = \frac{\partial}{\partial z_i^{(\ell)}} J(W,b) = \sum_i \left(\frac{\partial J(W,b)}{\partial z_i^{(n_\ell)}} \cdot \frac{\partial z_i^{(n_\ell)}}{\partial a_i^{(n_\ell - 1)}} \right) \frac{\partial a_i^{(n_\ell - 1)}}{\partial z_i^{(n_\ell - 1)}} = \sum_i \left(\delta_i^{(n_\ell)} \cdot W_{ij}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)})$

ℓ $\quad \delta_i^{(\ell)} = \sum_{i'} \left(\frac{\partial J(W,b)}{\partial z_{i'}^{(\ell+1)}} \cdot \frac{\partial z_{i'}^{(\ell+1)}}{\partial a_i^{(\ell)}} \right) \cdot \frac{\partial a_i^{(\ell)}}{\partial z_i^{(\ell)}} = \sum_{i'} \left(\delta_{i'}^{(\ell+1)} \cdot W_{ji'}^{(\ell)} \right) \sigma_i'^{(\ell)}(z_i^{(\ell)}) \Rightarrow \boxed{\delta^{(\ell)} = (W^{(\ell)})^T \cdot \delta^{(\ell+1)} \bullet \sigma'(z^{(\ell)})}$

2.5 e 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor calculation error

+ 5 pts Correct chain rule

+ 0 pts Incorrect

+ 6 pts Correct code, no detailed solution with steps written out.

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

Linear

node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

2.6 f **0 / 10**

+ **10 pts** Correct
+ **8 pts** Mostly Correct (misnamed a variable, forgot a negative sign, off by a constant)
+ **7 pts** minor conceptual mistake
+ **5 pts** update of either bias completely wrong or weight completely wrong or
+ **3 pts** only completed one part (either bias or weight) of code and that part is mostly correct
+ **2 pts** No code, but correct concept via equation

✓ + **0 pts** **Incorrect or blank**

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

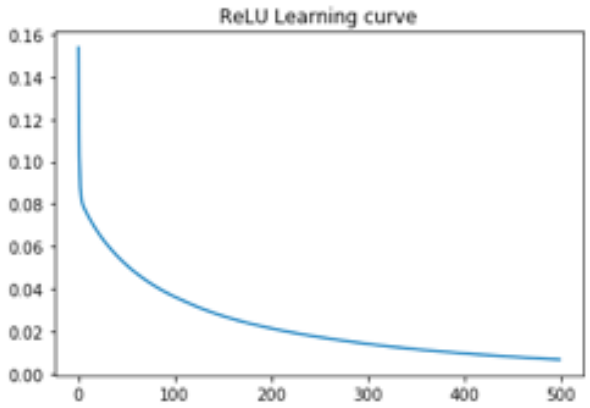
Linear

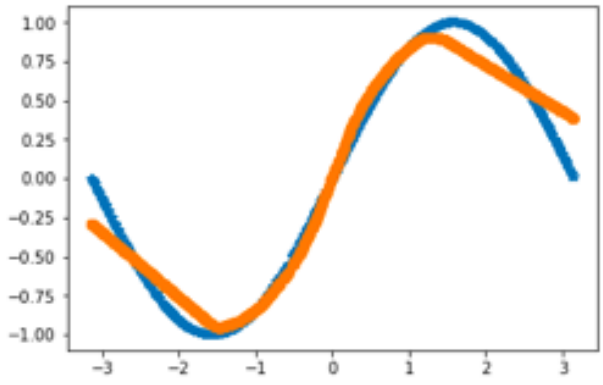
node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

ReLU:

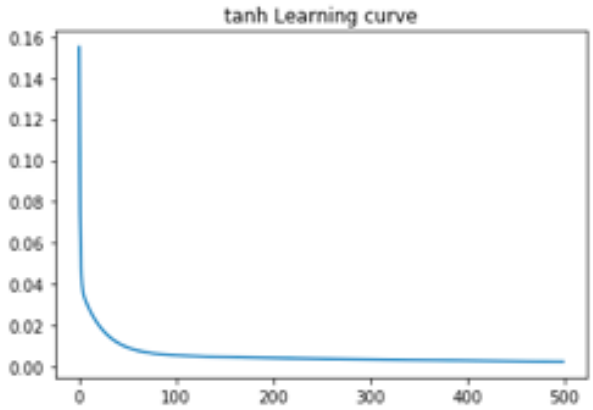
```
ReLU MSE: 0.00684167531631
```

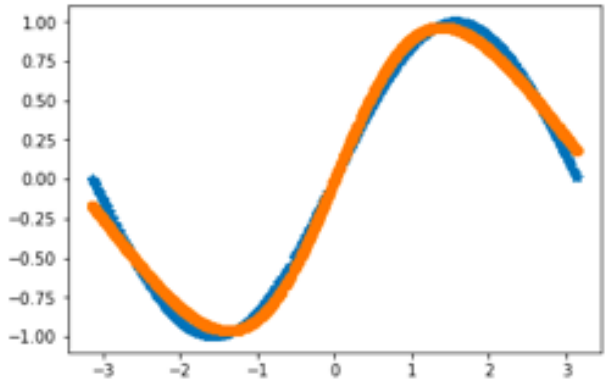




tanh:

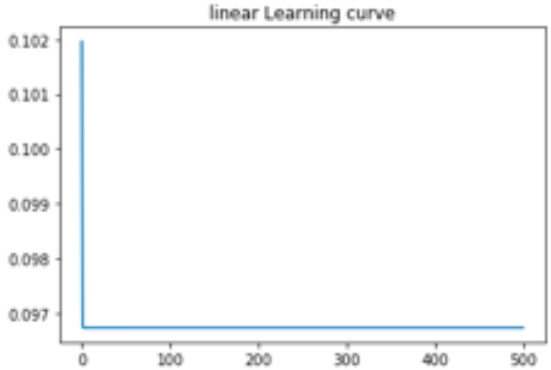
```
tanh MSE: 0.00224040515126
```

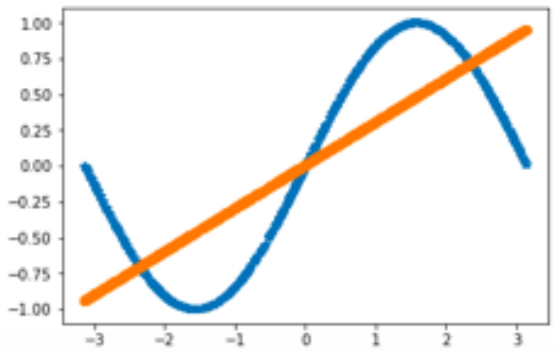




Linear:

```
linear MSE: 0.0967346963549
```





2.7 g **10 / 10**

✓ + **3.4 pts** tanh

✓ + **3.3 pts** linear

✓ + **3.3 pts** relu

+ **0 pts** Incorrect

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

Linear

node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

h) Error for depth 1:

```
NodeSize:5ReLU MSE: 0.192063956192
NodeSize:5tanh MSE: 0.0583296871098
NodeSize:5linear MSE: 0.0967370497222
NodeSize:10ReLU MSE: 0.201100031878
NodeSize:10tanh MSE: 0.0589004346013
NodeSize:10linear MSE: 0.0967350951925
NodeSize:25ReLU MSE: 0.217364704227
NodeSize:25tanh MSE: 0.0586952807986
NodeSize:25linear MSE: 0.0967366024839
NodeSize:50ReLU MSE: 0.204714233822
NodeSize:50tanh MSE: 0.0586419014661
NodeSize:50linear MSE: 0.0967359899126
```

Error for depth 2:

```
NodeSize:5ReLU MSE: 0.188586420577
NodeSize:5tanh MSE: 0.032066262804
NodeSize:5linear MSE: 0.0967354026467
NodeSize:10ReLU MSE: 0.213824199911
NodeSize:10tanh MSE: 0.0266254366438
NodeSize:10linear MSE: 0.0967347738012
NodeSize:25ReLU MSE: 0.187544468461
NodeSize:25tanh MSE: 0.0283601523481
NodeSize:25linear MSE: 0.0967347025341
NodeSize:50ReLU MSE: 0.180689589239
NodeSize:50tanh MSE: 0.0223763587072
NodeSize:50linear MSE: 0.0967347084438
```

Error for depth 3:

```
NodeSize:5ReLU MSE: 0.0233597881838
NodeSize:5tanh MSE: 0.0214648256952
NodeSize:5linear MSE: 0.096734718977
NodeSize:10ReLU MSE: 0.0162789613305
NodeSize:10tanh MSE: 0.01970545141
NodeSize:10linear MSE: 0.0967347269163
NodeSize:25ReLU MSE: 0.00262489530034
NodeSize:25tanh MSE: 0.00620770725242
NodeSize:25linear MSE: 0.0967347054405
NodeSize:50ReLU MSE: 0.00437158064479
NodeSize:50tanh MSE: 0.01021321621
NodeSize:50linear MSE: 0.0967347003152
```

2.8 h **10 / 10**

✓ **+ 10 pts** **Correct/Reasonable Numbers**

+ 8 pts Minor Errors

+ 5 pts Halfway There

+ 2 pts Correct Direction

+ 0 pts Incorrect/Blank

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

Linear

node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

2.9 i 0 / 10

+ 10 pts Correct

+ 8 pts Minor Error

+ 5 pts Partial

✓ + 0 pts Incorrect/Missing

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

Linear

node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

2.10 j **0 / 0**

+ **10 pts** Correct/Close
+ **8 pts** Some Error
+ **5 pts** Partial
✓ + **0 pts** **Incorrect/Missing**

f) code

g) ReLU : 0.0068
tanh : 0.0022
linear Activation: 0.0967

h) ReLU

node \ depth	1	2	3
5	0.1920	0.1886	0.0233
10	0.2011	0.2138	0.0128
25	0.2173	0.1875	0.00262
50	0.2049	0.1806	0.00437

tanh

node \ depth	1	2	3
5	0.0588	0.032	0.0215
10	0.0587	0.026	0.0197
25	0.0586	0.028	0.0062
50	0.0584	0.022	0.0102

Linear

node \ depth	1	2	3
5	0.0967	0.0967	0.0967
10	0.0967	0.0967	0.0967
25	0.0967	0.0967	0.0967
50	0.0967	0.0967	0.0967

j) code

2.11 k 0 / 0

+ **10 pts** Graphs and Justification

+ **4 pts** Claim without justification

✓ + **0 pts** **No Answer**

HW07

My Own Question

Is it the same when do $\frac{\partial MSE}{\partial a^{(l)}}$ and $\frac{\partial MSE}{\partial z^{(l)}}$?

Yes, except for we should iterate using different gradient.

If we assume $\frac{\partial MSE}{\partial a^l} = \delta^l$

for output layer $\delta^{nl} = -(y - a^{nl})$

for layer l $\frac{\partial MSE}{\partial a^l} = \frac{\partial MSE}{\partial a_{l+1}} \cdot \frac{\partial a_{l+1}}{\partial z_{l+1}} \cdot \frac{\partial z_{l+1}}{\partial a_l}$

$= (W^l)^T \left(\frac{\partial MSE}{\partial a_{l+1}} \odot f'(z^{(l+1)}) \right)_{l+1}$

Elementwise

→ difference

3 Your Own Question 10 / 10

✓ + 10 pts **Question and Solution**

+ 0 pts Incorrect or blank