

## **Table of Contents:**

Question 1a): Figure 1, Figure 2 in OUTPUT.pdf

Question 1b): Figure 3, Figure 4 in OUTPUT.pdf and Page 2 in OTHER.pdf

Question 2a-g): Pages 7-9 in OTHER.pdf

Question 3a): Lines 83-91 in source.py

Question 3b): Lines 83-91 in source.py

Question 3c): in OUTPUT.pdf [sorry I put it in between Question 1b)]

Question 3d): Lines 119-140 in source.py

Question 3e): Pages 2-4 in OTHER.pdf

Question 3f): Figure 5, Figure 6 in OUTPUT.pdf

Question 4a): Lines 171-174 in source.py

Question 4b): Page 4 in OTHER.pdf

Question 4c): Lines 182-186 in source.py

Question 4d): Figure 7, Figure 8, Figure 9, Figure 10 in OUTPUT.pdf

Question 4e): Pages 4-5 in OTHER.pdf

Question 4f): Page 10 in OTHER.pdf

Question 4g): Figure 11, Figure 12 in OUTPUT.pdf

Question 4h): Pages 5-6 in OTHER.pdf

**Question #1b)**

```
Testing error for training session 1 = 0.0346
Testing error for training session 2 = 0.0373
Testing error for training session 3 = 0.1287
Testing error for training session 4 = 0.0359
Testing error for training session 5 = 0.035
Testing error for training session 6 = 0.11
Testing error for training session 7 = 0.0371
Testing error for training session 8 = 0.13
Testing error for training session 9 = 0.037
Smallest test error = 0.0346
```

**Question #3e)**

```
Learning_rate_init = 0.18 ; Momentum = 0.65 ; Alpha = 0.165

Testing error for training session 1 = 3.52
Training error for training session 1 = 1.48
Testing error for training session 2 = 3.33
Training error for training session 2 = 1.3
Testing error for training session 3 = 3.3
Training error for training session 3 = 1.15
Testing error for training session 4 = 3.15
Training error for training session 4 = 1.03
Testing error for training session 5 = 2.99
Training error for training session 5 = 0.88
Testing error for training session 6 = 3.07
Training error for training session 6 = 0.86
Testing error for training session 7 = 2.93
Training error for training session 7 = 0.81
Testing error for training session 8 = 3.02
Training error for training session 8 = 0.75
Testing error for training session 9 = 2.95
Training error for training session 9 = 0.69
Testing error for training session 10 = 3.01
Training error for training session 10 = 0.7
Testing error for training session 11 = 2.93
Training error for training session 11 = 0.67
Testing error for training session 12 = 2.91
Training error for training session 12 = 0.61
Testing error for training session 13 = 3.06
Training error for training session 13 = 0.63
```

Testing error for training session 14 = 3.14  
Training error for training session 14 = 0.57  
Testing error for training session 15 = 3.06  
Training error for training session 15 = 0.6  
Testing error for training session 16 = 2.95  
Training error for training session 16 = 0.61  
Testing error for training session 17 = 2.98  
Training error for training session 17 = 0.55  
Testing error for training session 18 = 2.86  
Training error for training session 18 = 0.5  
Testing error for training session 19 = 2.79  
Training error for training session 19 = 0.54  
Testing error for training session 20 = 2.86  
Training error for training session 20 = 0.53  
Testing error for training session 21 = 2.72  
Training error for training session 21 = 0.52  
Testing error for training session 22 = 2.79  
Training error for training session 22 = 0.51  
Testing error for training session 23 = 2.76  
Training error for training session 23 = 0.48  
Testing error for training session 24 = 2.87  
Training error for training session 24 = 0.55  
Testing error for training session 25 = 2.83  
Training error for training session 25 = 0.48  
Testing error for training session 26 = 2.82  
Training error for training session 26 = 0.42  
Testing error for training session 27 = 2.8  
Training error for training session 27 = 0.51  
Testing error for training session 28 = 2.72  
Training error for training session 28 = 0.49  
Testing error for training session 29 = 2.83  
Training error for training session 29 = 0.45  
Testing error for training session 30 = 2.58  
Training error for training session 30 = 0.48  
Testing error for training session 31 = 2.74  
Training error for training session 31 = 0.46  
Testing error for training session 32 = 2.6  
Training error for training session 32 = 0.44  
Testing error for training session 33 = 2.87  
Training error for training session 33 = 0.43  
Testing error for training session 34 = 2.74  
Training error for training session 34 = 0.43  
Testing error for training session 35 = 2.54  
Training error for training session 35 = 0.42  
Testing error for training session 36 = 2.71

```
Training error for training session 36 = 0.51
Testing error for training session 37 = 2.76
Training error for training session 37 = 0.5
Testing error for training session 38 = 2.68
Training error for training session 38 = 0.42
Testing error for training session 39 = 2.66
Training error for training session 39 = 0.47
Testing error for training session 40 = 2.72
Training error for training session 40 = 0.44
Testing error for training session 41 = 2.72
Training error for training session 41 = 0.44
Testing error for training session 42 = 2.71
Training error for training session 42 = 0.5
Testing error for training session 43 = 2.57
Training error for training session 43 = 0.42
Testing error for training session 44 = 2.65
Training error for training session 44 = 0.47
Testing error for training session 45 = 2.65
Training error for training session 45 = 0.45
Testing error for training session 46 = 2.8
Training error for training session 46 = 0.46
Testing error for training session 47 = 2.72
Training error for training session 47 = 0.48
Testing error for training session 48 = 2.8
Training error for training session 48 = 0.5
Testing error for training session 49 = 2.61
Training error for training session 49 = 0.46
Testing error for training session 50 = 2.89
Training error for training session 50 = 0.39
Minimum test error = 2.54
[Finished in 88.7s]
```

**Question #4b)** 2.74405361693e-28

**Question #4e) \*last 10 lines\***

```
Iteration number = 955
Testing error for training session 955 = 4.58
Training error for training session 955 = 1.79
Mean training loss for training session 955 = 0.48
Iteration number = 960
Testing error for training session 960 = 4.58
Training error for training session 960 = 1.78
Mean training loss for training session 960 = 0.48
Iteration number = 965
```

```
Testing error for training session 965 = 4.58
Training error for training session 965 = 1.77
Mean training loss for training session 965 = 0.48
Iteration number = 970
Testing error for training session 970 = 4.57
Training error for training session 970 = 1.77
Mean training loss for training session 970 = 0.48
Iteration number = 975
Testing error for training session 975 = 4.57
Training error for training session 975 = 1.76
Mean training loss for training session 975 = 0.47
Iteration number = 980
Testing error for training session 980 = 4.57
Training error for training session 980 = 1.76
Mean training loss for training session 980 = 0.47
Iteration number = 985
Testing error for training session 985 = 4.56
Training error for training session 985 = 1.75
Mean training loss for training session 985 = 0.47
Iteration number = 990
Testing error for training session 990 = 4.56
Training error for training session 990 = 1.74
Mean training loss for training session 990 = 0.47
Iteration number = 995
Testing error for training session 995 = 4.56
Training error for training session 995 = 1.74
Mean training loss for training session 995 = 0.47
Iteration number = 1000
Testing error for training session 1000 = 4.55
Training error for training session 1000 = 1.73
Mean training loss for training session 1000 = 0.47
```

#### Question #4g)

Minimum test error during 50 epochs of training = **4.07**  
(I tried to bring it lower than that but couldn't do so. I just need more time to tune the hyperparameters) □

#### Question #4h) batchSize = 3000; momentum = 0.99

```
Iteration number = 0
Testing error for training session 0 = 20.4
Training error for training session 0 = 20.53
Mean training loss for training session 0 = 4.76
```

Iteration number = 5  
Testing error for training session 5 = 8.11  
Training error for training session 5 = 6.27  
Mean training loss for training session 5 = 1.61  
Iteration number = 10  
Testing error for training session 10 = 6.48  
Training error for training session 10 = 4.19  
Mean training loss for training session 10 = 1.01  
Iteration number = 15  
Testing error for training session 15 = 5.46  
Training error for training session 15 = 2.71  
Mean training loss for training session 15 = 0.58  
Iteration number = 20  
Testing error for training session 20 = 4.76  
Training error for training session 20 = 1.82  
Mean training loss for training session 20 = 0.38  
Iteration number = 25  
Testing error for training session 25 = 4.31  
Training error for training session 25 = 1.34  
Mean training loss for training session 25 = 0.31  
Iteration number = 30  
Testing error for training session 30 = 4.1  
Training error for training session 30 = 1.19  
Mean training loss for training session 30 = 0.33  
Iteration number = 35  
Testing error for training session 35 = 4.07  
Training error for training session 35 = 1.27  
Mean training loss for training session 35 = 0.39  
Iteration number = 40  
Testing error for training session 40 = 4.14  
Training error for training session 40 = 1.42  
Mean training loss for training session 40 = 0.45  
Iteration number = 45  
Testing error for training session 45 = 4.21  
Training error for training session 45 = 1.51  
Mean training loss for training session 45 = 0.48  
Iteration number = 50  
Testing error for training session 50 = 4.22  
Training error for training session 50 = 1.53  
Mean training loss for training session 50 = 0.48

# CSC411 Assignment 3

Akhil Gupta  
(000359011)

H2. a) prove that  $\frac{\partial \sigma_k}{\partial z_j} = -\sigma_k o_j$ ,  $k \neq j$

$$\begin{aligned}
 \frac{\partial \sigma_k}{\partial z_j} &= \frac{\partial e^{z_k}}{\sum_i e^{z_i}} = \frac{0 - e^{z_k} \cdot e^{z_j}}{\left(\sum_i e^{z_i}\right)^2} = -\frac{e^{z_k} \cdot e^{z_j}}{\left(\sum_i e^{z_i}\right)^2} \\
 &\quad \overbrace{\phantom{\frac{\partial \sigma_k}{\partial z_j}}^{\partial z_j}} \\
 &= \frac{-e^{z_k}}{\sum_i e^{z_i}} \cdot \frac{e^{z_j}}{\sum_i e^{z_i}} \\
 &= \underline{-\sigma_k o_j}
 \end{aligned}$$

b) prove that  $\frac{\partial \sigma_k}{\partial z_k} = (1 - \sigma_k)\sigma_k$

$$\begin{aligned}
 \frac{\partial \sigma_k}{\partial z_k} &= \frac{\partial e^{z_k}}{\sum_i e^{z_i}} = \frac{e^{z_k} \cdot \sum_i e^{z_i} - e^{z_k} \cdot e^{z_k}}{\left(\sum_i e^{z_i}\right)^2} \\
 &\quad \overbrace{\phantom{\frac{\partial \sigma_k}{\partial z_k}}^{\partial z_k}} \\
 &= \frac{e^{z_k} \left( \sum_i e^{z_i} - e^{z_k} \right)}{\left(\sum_i e^{z_i}\right)^2} \\
 &= \frac{e^{z_k}}{\sum_i e^{z_i}} \cdot \left( \frac{\sum_i e^{z_i} - e^{z_k}}{\sum_i e^{z_i}} \right) \\
 &= \underline{\sigma_k (1 - \sigma_k)}
 \end{aligned}$$

c) derive formulas for  $\frac{\partial z_i}{\partial w_{kj}} \cdot h_j$

$$\frac{\partial z_i}{\partial w_{kj}} = \frac{\partial \sum_{j=1}^J w_{kj} h_j + w_{k0}}{\partial w_{kj}} = h_j \quad (\text{hidden layer})$$

d) prove that for  $j \geq 1$ ,

$$\frac{\partial L(w, w_0)}{\partial w_{kj}} = \sum_n [o_k^{(n)} - t_k^{(n)}] h_j^{(n)}$$

$$\begin{aligned} \Rightarrow \frac{\partial L(w, w_0)}{\partial w_{kj}} &= \frac{\partial E}{\partial w_{kj}} = \sum_{n=1}^N \frac{\partial E}{\partial o_k^{(n)}} \cdot \frac{\partial o_k^{(n)}}{\partial z_k^{(n)}} \cdot \frac{\partial z_k^{(n)}}{\partial w_{kj}} \\ &= \sum_{n=1}^N \left( -\frac{t_k^{(n)}}{o_k^{(n)}} \right) \cdot \left( o_k^{(n)}(1 - o_k^{(n)}) \right) \cdot \left( h_j^{(n)} \right) \\ &= \sum_{n=1}^N \left( -t_k^{(n)} \right) \cdot \left( 1 - o_k^{(n)} \right) \cdot \left( h_j^{(n)} \right) \\ &= \sum_n [o_k^{(n)} - t_k^{(n)}] \cdot h_j^{(n)} \end{aligned}$$

e) prove that  $\frac{\partial L(w, w_0)}{\partial w_{k0}} = \sum_n [o_k^{(n)} - t_k^{(n)}]$

if we let  $h_0 = 1$ , then  $z_k = \sum_{j=1}^J w_{kj} + w_{k0}$ , and the proof in part d) goes

through for  $j=0$ , thus,

$$\frac{\partial L(w, w_0)}{\partial w_{k0}} = \sum_n [o_k^{(n)} - t_k^{(n)}] \cdot h_0^{(n)} = \sum_n [o_k^{(n)} - t_k^{(n)}]$$

f) prove that  $\frac{\partial L(w, w_0)}{\partial w} = (O - T)^T H$

$k_j^{th}$  entry of  $w = \frac{\partial L(w, w_0)}{\partial w_{kj}}$

$$\begin{aligned} \left[ \frac{\partial L(w, w_0)}{\partial w} \right]_{kj} &= \frac{\partial L(w, w_0)}{\partial w_{kj}} \\ &= \sum_n [o_k^{(n)} - t_k^{(n)}] \cdot h_j^{(n)} \\ &= \sum_n [O - T]_{nk} \cdot H_{nj} \\ &= \sum_n [O - T]^T_{kn} \cdot H_{nj} \\ &= [(O - T)^T \cdot H]_{kj} \end{aligned}$$

therefore,  $\frac{\partial L(w, w_0)}{\partial w} = (O - T)^T \cdot H$  since all components are equal.

g) derive formula for  $\frac{\partial L(w, w_0)}{\partial w_0}$

$$\begin{aligned} \frac{\partial L(w, w_0)}{\partial w_0} &= \sum_n [o_k^{(n)} - t_k^{(n)}] \\ &= \sum_n [O - T]_{nk} \\ &= \underbrace{\sum_n [O - T]}_N / N \quad (\text{row-wise addition of } O - T) \end{aligned}$$

$N \rightarrow \infty$

Question #4f)

Ritik Gupta  
1000357071

$$C = \frac{L}{N} + \frac{\alpha}{2} \sum_{kj=1} w_{kj}^2$$

$$\frac{\partial C}{\partial w} = \frac{\partial L/N}{\partial w} + \frac{\alpha}{2} (2w_{kj})$$

$$\frac{\partial C}{\partial w} = \frac{\partial L/N}{\partial w} + \alpha_w ; \quad \frac{\partial C}{\partial w_0} = \frac{\partial L/N}{\partial w_0}$$

$$\frac{\partial C}{\partial w} = dw_2 + \alpha w_2 ; \quad \frac{\partial C}{\partial w_0} = db_2$$

---