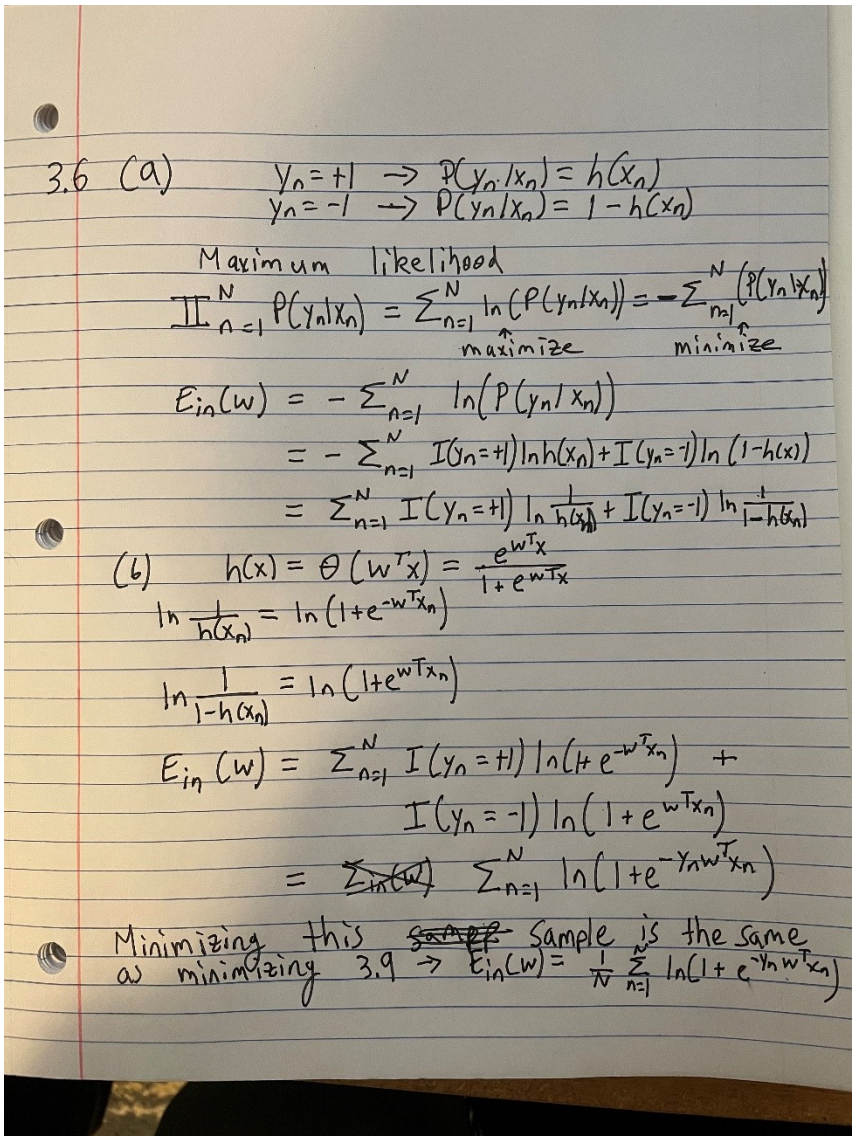
Team Evaluation

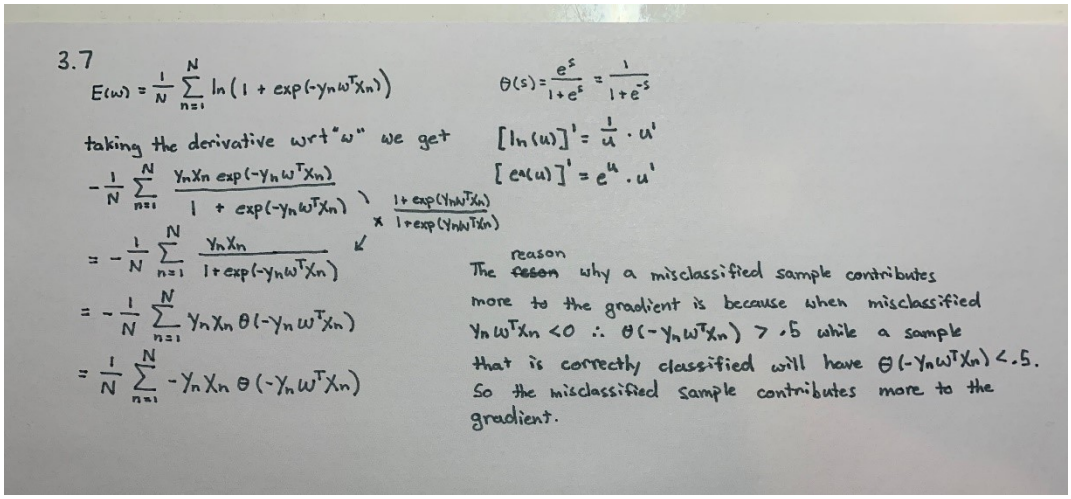
1. Brendan Rizzo
    a. Worked on Question 1 and came up with a great answer
2. Richard Huang
    a. Richard solved question 2 with little help by Eden.
3. Jiashang Cao
    a. I don't know where Jiashang went.
4. **Eden Seo**
    a. Helped Richard little bit on solving question 2.

1. Question 1 (Exercise 3.6 in LFD)
    a.



2. Question 2 (Exercise 3.7 in LFD)

a.



3.7

$$E(w) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1 + \exp(-y_n w^T x_n)\right)$$

$$\theta(s) = \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}}$$

taking the derivative wrt "w" we get

$$[\ln(u)]' = \frac{1}{u}\cdot u' \qquad [e^{(u)}]' = e^u \cdot u'$$

$$-\frac{1}{N}\sum_{n=1}^{N} \frac{y_n x_n \exp(-y_n w^T x_n)}{1+\exp(-y_n w^T x_n)} \times \frac{1+\exp(y_n w^T x_n)}{1+\exp(y_n w^T x_n)}$$

$$= -\frac{1}{N}\sum_{n=1}^{N} \frac{y_n x_n}{1+\exp(-y_n w^T x_n)}$$

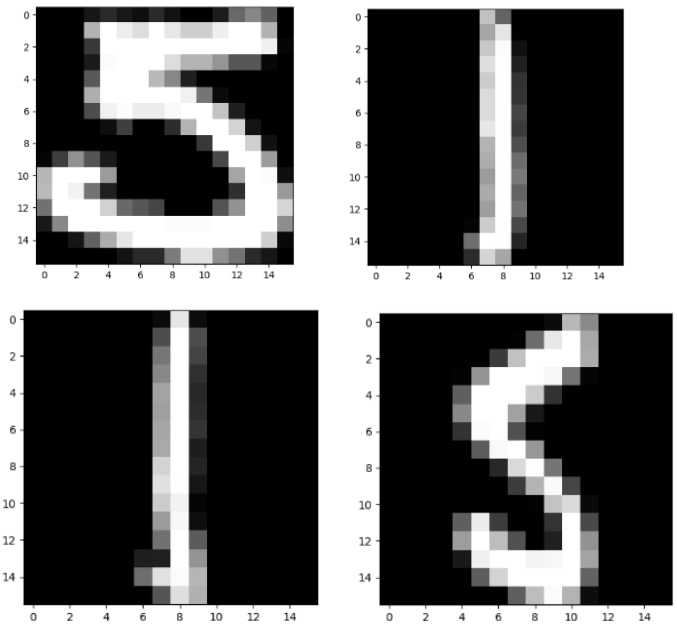
$$= -\frac{1}{N}\sum_{n=1}^{N} y_n x_n \theta(-y_n w^T x_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} -y_n x_n \theta(-y_n w^T x_n)$$

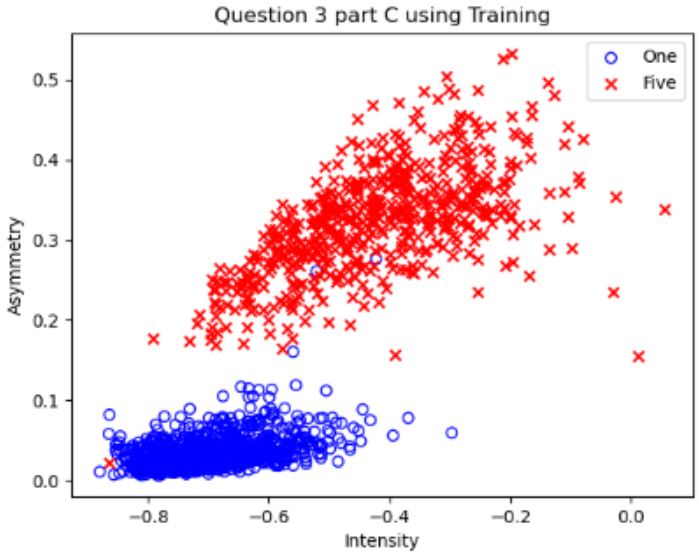
The ~~reason~~ reason why a misclassified sample contributes more to the gradient is because when misclassified $y_n w^T x_n < 0 \therefore \theta(-y_n w^T x_n) > .5$ while a sample that is correctly classified will have $\theta(-y_n w^T x_n) < .5$. So the misclassified sample contributes more to the gradient.

3. Question 3

a.



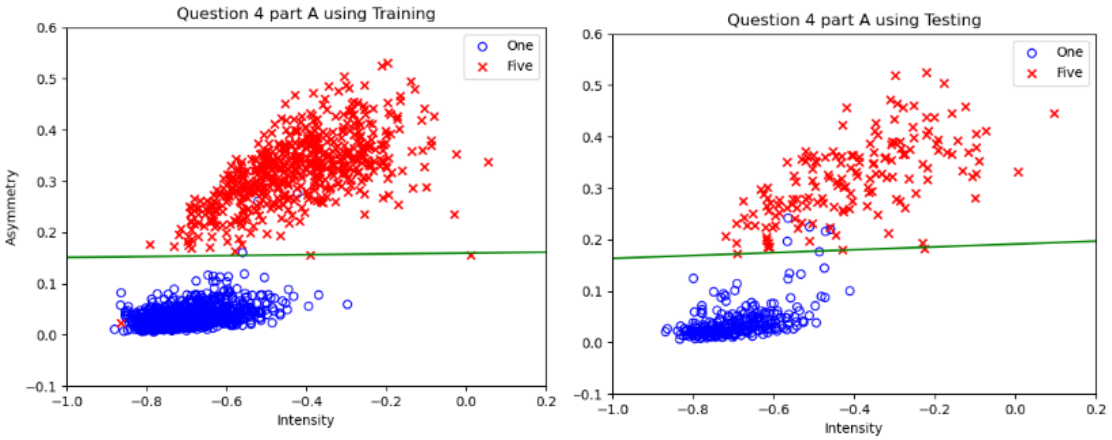
b. Intensity of One: [-0.7539140625, -0.77228125, … -0.44755859374999996]
Intensity of Five: [-0.11173828124999999, -0.56403515625, … -0.53423828125]
Symmetry of One: [0.029765625, 0.035273437500000004, … 0.04910546875]
Symmetry of Five: [0.42023828124999996, 0.21845703125, … 0.29608203125]




c.

4. Question 4

a.




b.

```
E_in:  0.0010790554053368
E_test:  0.006148133462409647
```

c.

```
E_train_3rd:  0.0019084895331687268
E_test_3rd:  0.009466150225192
```

d. We should not use 3rd order transform since it has higher E values than the original