

Assignment 1 Report

This is an outline for your report to ease the amount of work required to create your report. Jupyter notebook supports markdown, and I recommend you to check out this [cheat sheet](#). If you are not familiar with markdown.

Before delivery, **remember to convert this file to PDF**. You can do it in two ways:

1. Print the webpage (ctrl+P or cmd+P)
2. Export with latex. This is somewhat more difficult, but you'll get somewhat of a "prettier" PDF. Go to File -> Download as -> PDF via LaTeX. You might have to install nbconvert and pandoc through conda; `conda install nbconvert pandoc`.

Task 1

task 1a)

$$C^n(\omega) = -(y^n \cdot \ln(\hat{y}^n) + (1 - y^n) \cdot \ln(1 - \hat{y}^n))$$

$$\hat{y} = f(x) = \frac{1}{1 + \exp(-z)}, \rightarrow z = \omega^T \cdot x = \sum_i^I \omega_i \cdot x_i$$

Assuming the y^n vector is constant:

$$\frac{\partial C^n}{\partial \omega_i} = -\frac{\partial}{\partial \omega_i} (y^n \cdot \ln(f(x^n))) - \frac{\partial}{\partial \omega_i} ((1 - y^n) \cdot \ln(1 - f(x^n)))$$

Performing the partial derivation on the two separate terms in the previous equation we get:

Starting with the first term:

$$First = -y^n \cdot \frac{\partial}{\partial \omega_i} (\ln(f(x^n))) = -y^n \cdot \frac{1}{f(x^n)} \cdot \frac{\partial f(x^n)}{\partial \omega_i} = -\frac{y^n}{f(x^n)} \cdot \frac{x_i^n \cdot \exp(-z)}{(1 + \exp(-z))^2} = -y^n \hat{y}^n$$

Continuing with the second term:

$$Second = -(1 - y^n) \frac{1}{1 - f(x^n)} \cdot \left(-\frac{x_i^n \cdot \exp(-z)}{(1 + \exp(-z))^2} \right)$$

Expanding on what $1 - f(x^n)$ is:

$$1 - f(x^n) = \frac{1 + \exp(-z)}{1 + \exp(-z)} - \frac{1}{1 + \exp(-z)} = \frac{\exp(-z)}{1 + \exp(-z)}$$

Hence;

$$\frac{1}{1 - f(x^n)} = \frac{1 + \exp(-z)}{\exp(-z)} = \frac{1}{\exp(-z)} \cdot \frac{1}{f(x^n)}$$

Going back to the *second* term:

$$-(1 - y^n) \frac{1}{\exp(-z)} \cdot \frac{1}{f(x^n)} \cdot \left(-\frac{x_i^n \cdot \exp(-z)}{(1 + \exp(-z))^2} \right) = -(1 - y^n) \frac{1}{\exp(-z)} \cdot \frac{1}{f(x^n)} \cdot (-x_i^n \cdot \exp(-z))$$

Hence;

$$Second = (1 - y^n) \cdot x_i^n \hat{y}^n$$

This gives us:

$$\frac{\partial C^n}{\partial \omega_i} = x_i^n (-y^n \hat{y}^n \cdot \exp(-z) + (1 - y^n) \cdot \hat{y}^n) = x_i^n (-y^n \hat{y}^n \cdot \exp(-z) + \hat{y}^n - y^n \hat{y}^n) = x_i^n (\hat{y}^n -$$

Seeing that $(1 - \exp(-z)) = \frac{1}{f(x^n)}$, we finally get:

$$\frac{\partial C^n}{\partial \omega_i} = x_i^n (\hat{y}^n - y^n) = -(y^n - \hat{y}^n) x_i^n \rightarrow QED$$



task 1b)

To find the partial derivative $\frac{\partial C^n}{\partial \omega_{k,j}}$ using the chain rule.

$$\frac{\partial C^n}{\partial \omega_{k,j}} = \frac{\partial C^n}{\partial \hat{y}_n^k} \frac{\partial \hat{y}_n^k}{\partial z_k} \frac{\partial z_k}{\partial \omega_{k,j}}$$

The first partial derivative is straightforward:

$$\frac{\partial C^n}{\partial \hat{y}_n^k} = - \sum_{k=1}^K y_k^n \frac{1}{\hat{y}_n^k}$$

For the last partial derivative, we note that for all $\omega_{k,i}$ except $\omega_{k,j}$, the partial derivative becomes zero. Thus, we get:

$$\frac{\partial z_k}{\partial \omega_{k,j}} = x_j^n$$

For the second partial derivative, we split the derivation into two parts. In the first case, we assume $k = k'$:

$$\frac{\partial \hat{y}_n^k}{\partial z_k} = \frac{\exp(z_k) \sum_{k'} \exp(z_{k'}) - \exp(z_k) \exp(z_k)}{(\sum_{k'} \exp(z_{k'}))^2} = \frac{\exp(z_k)}{\sum_{k'} \exp(z_{k'})} \left(\frac{\sum_{k'} \exp(z_{k'})}{\sum_{k'} \exp(z_{k'})} - \frac{\exp(z_k)}{\sum_{k'} \exp(z_{k'})} \right)$$

In the second case, $k \neq k'$:

$$\frac{\partial \hat{y}_k^n}{\partial z_k} = \frac{0 - \exp(z_k) \exp(z_{k'})}{(\sum_{k'} \exp(z_{k'}))^2} = -\hat{y}_k^n \hat{y}_{k'}^n$$

Combining this, we get

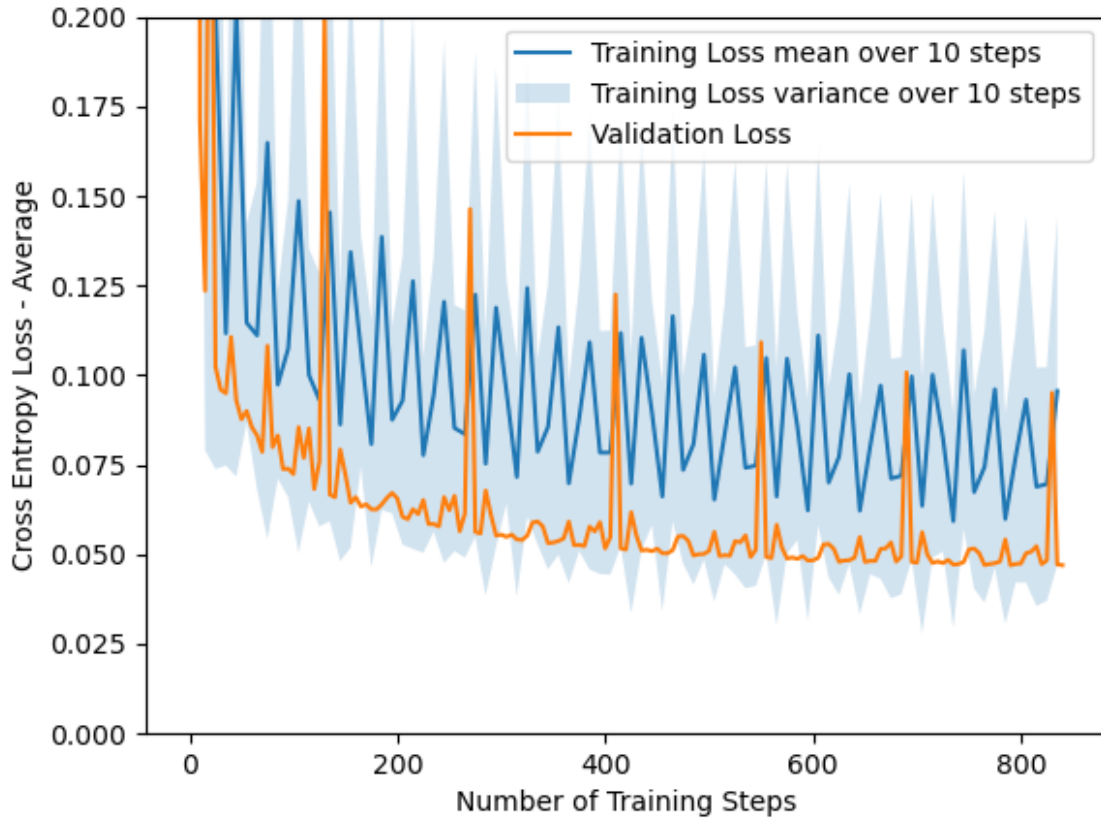
$$\frac{\partial C^n}{\partial \omega_{k,j}} = \frac{\partial z_k}{\partial \omega_{k,j}} \left(-\sum_{k \neq k'} y_k^n \frac{1}{\hat{y}_k^n} \frac{\partial \hat{y}_k^n}{\partial z_k} - y_{k'} \frac{1}{\hat{y}_{k'}} \frac{\partial \hat{y}_{k'}^n}{\partial z_{k'}} \right)$$

$$\frac{\partial C^n}{\partial \omega_{k,j}} = x_j^n \left(-\sum_{k \neq k'} y_k^n \frac{1}{\hat{y}_k^n} (-\hat{y}_k^n \hat{y}_{k'}^n) - y_{k'} \frac{1}{\hat{y}_{k'}} \hat{y}_k^n (1 - \hat{y}_k^n) \right)$$

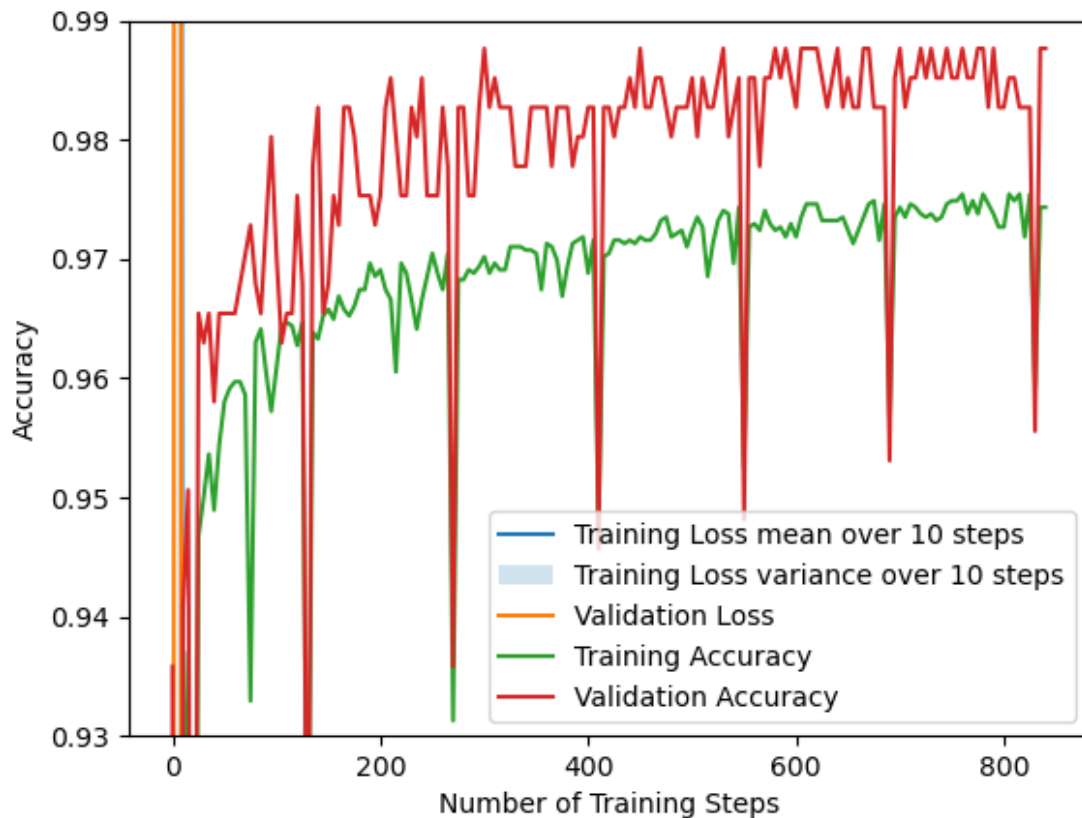
$$\frac{\partial C^n}{\partial \omega_{k,j}} = x_j^n \left(\sum_{k \neq k'} y_k^n \hat{y}_{k'}^n + y_k^n \hat{y}_{k'}^n - y_{k'} \right) = x_j^n \left(\sum_k y_k^n \hat{y}_{k'}^n - y_{k'} \right) = x_j^n (\hat{y}_{k'}^n \sum_k y_k^n - y_{k'}) = -x_j^n (y_{k'} - \hat{y}_{k'}^n)$$

Task 2

Task 2b)



Task 2c)



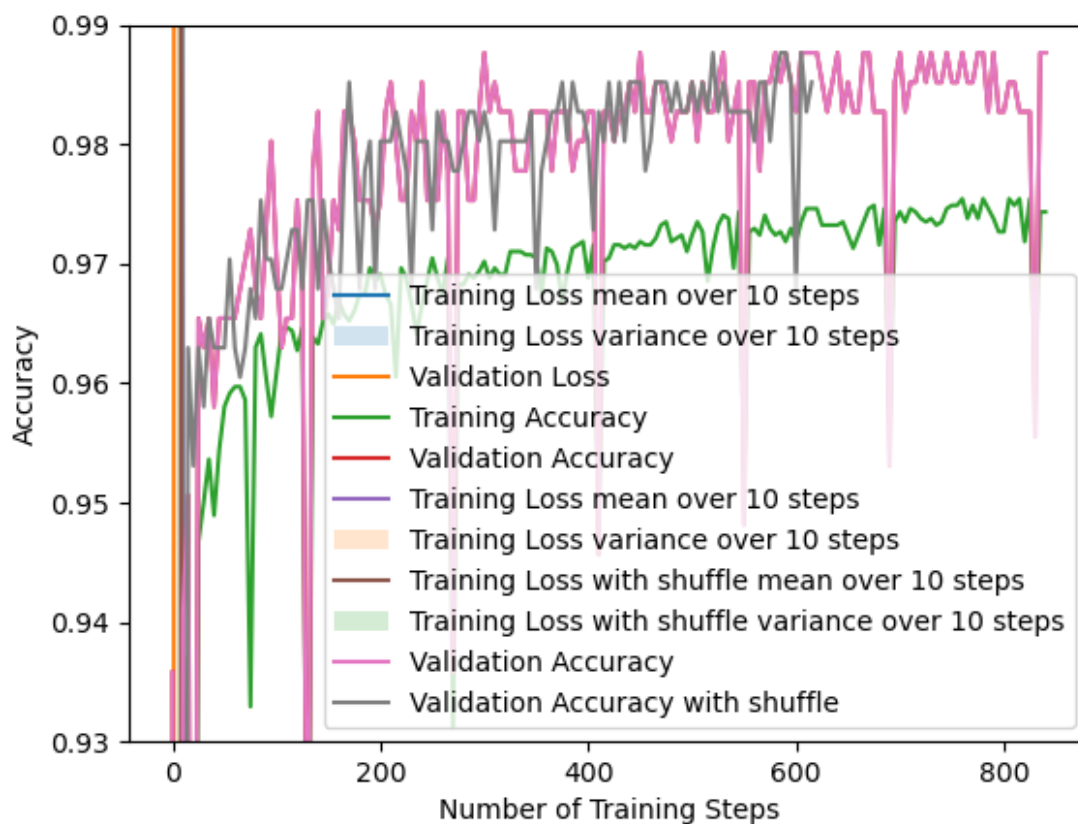
Task 2d)

Early stopping generally kicks in after 30 epochs. With shuffling this number varies between 15 and 35, but is usually approximately 25.

Task 2e)

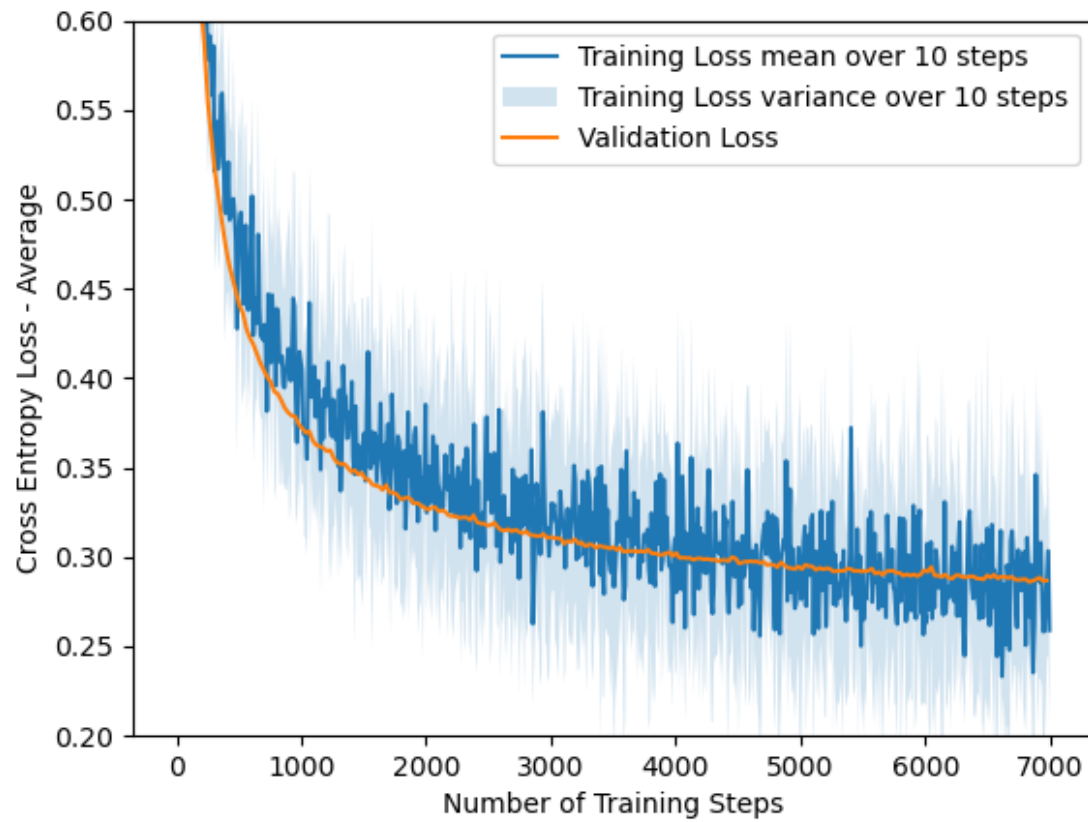
When using logistic regression with gradient descent, there is a risk of being stuck in local minima where the gradient is near zero. Without shuffling, the dataset is static and these minima will not be properly avoided. Using shuffling however, provides a way to rearrange the dataset between

epochs, so a local minima will no longer be in the same place as before.

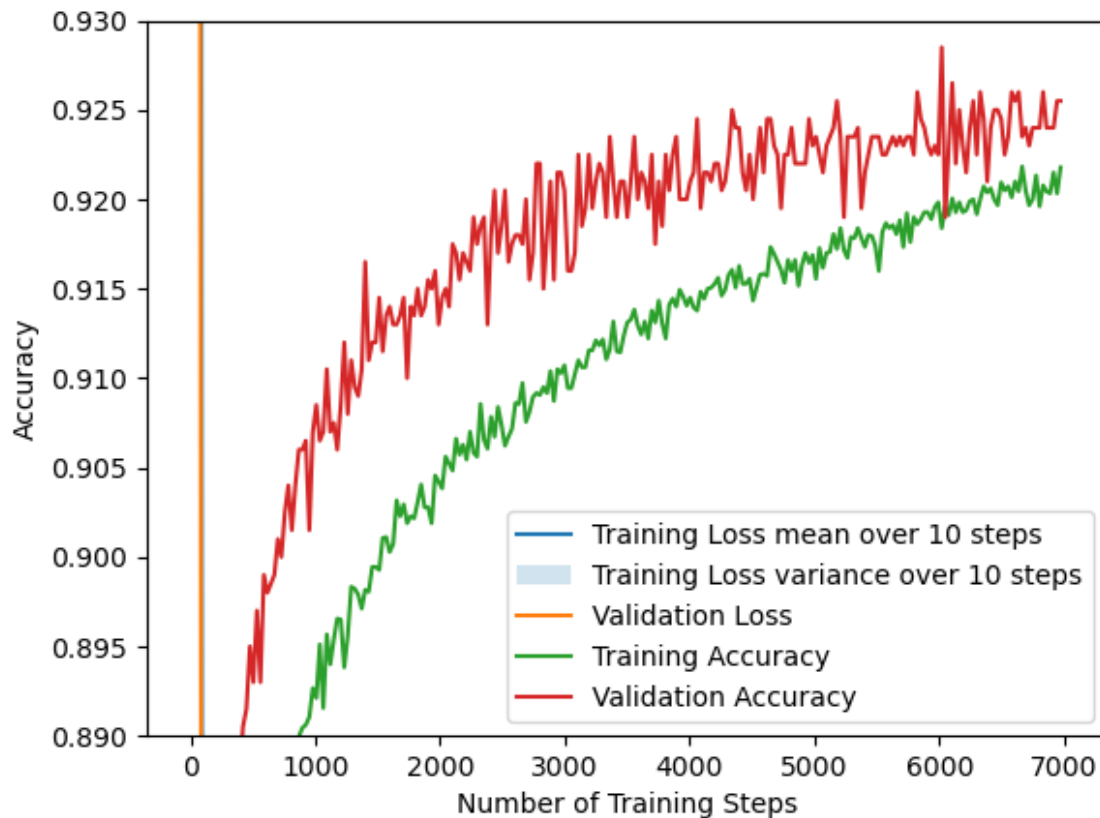


Task 3

Task 3b)



Task 3c)



Task 3d)

In terms of train loss and validation loss, we observe that near the end, the train loss seems to have lower mean than the validation loss. This might be an indicator of overfitting, as the model then better describes the training data than the validation data. This means the model is not general enough, and therefore cannot account for biases in the validation data set.

Task 4

Task 4a)

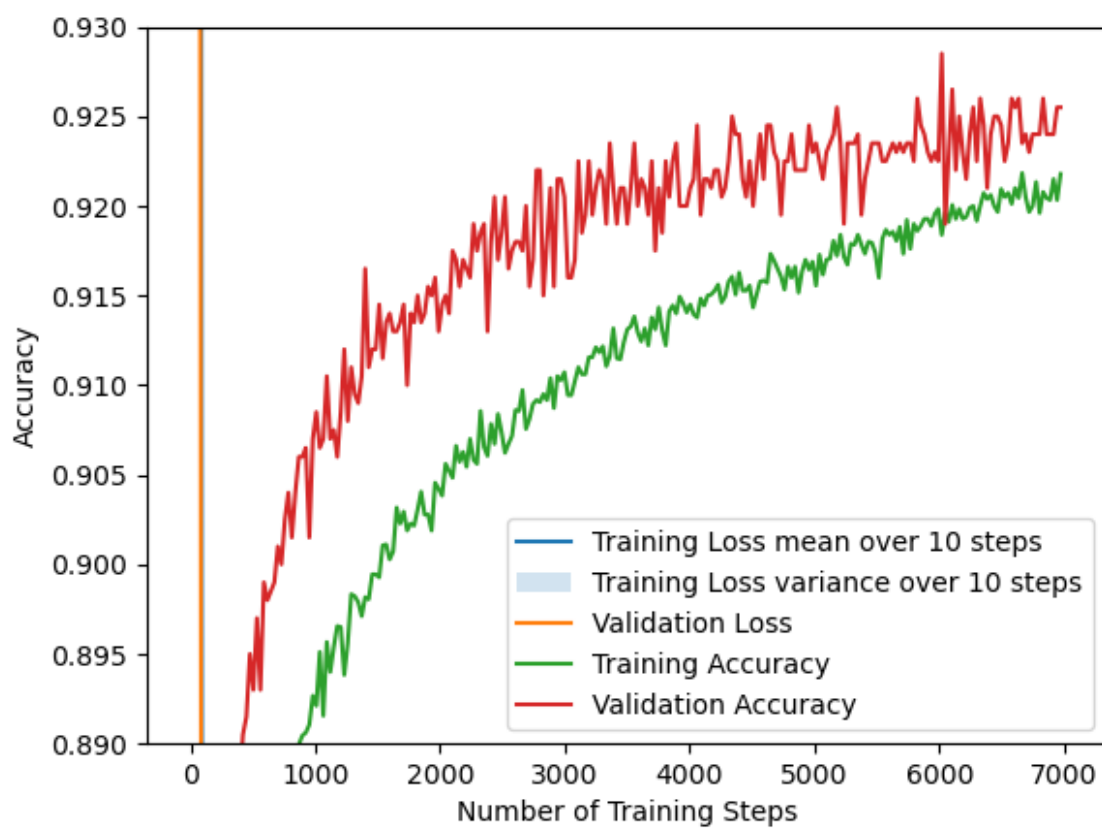
$$\frac{\partial J}{\partial \omega} = \frac{\partial C}{\partial \omega} + \frac{\partial R}{\partial \omega} = \frac{-x(y - \hat{y})}{N} + \frac{\lambda}{2} \frac{\partial}{\partial \omega} \omega(\omega)^T = \frac{-x(y - \hat{y})}{N} + \lambda \omega$$

Task 4b)

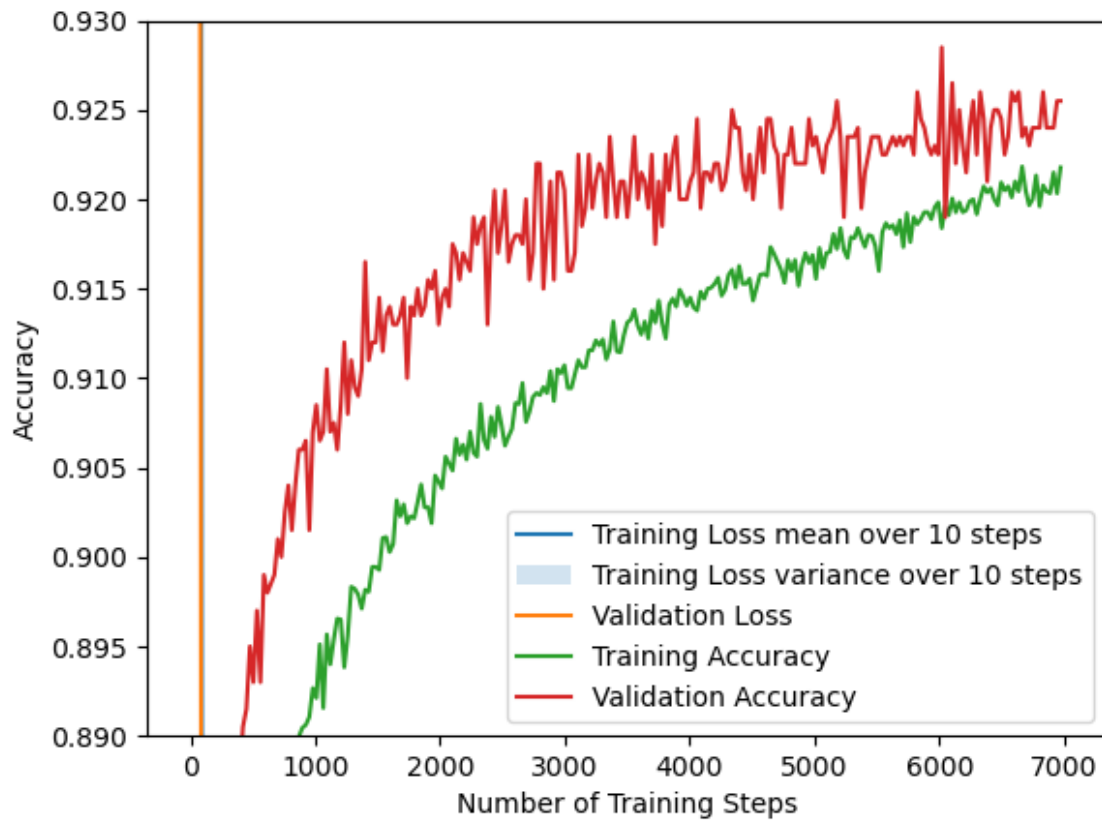
FILL IN ANSWER

Task 4c)

FILL IN ANSWER



Task 4d)



Task 4e)

FILL IN ANSWER