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 (https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet)](https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet). 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 somehwat 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)

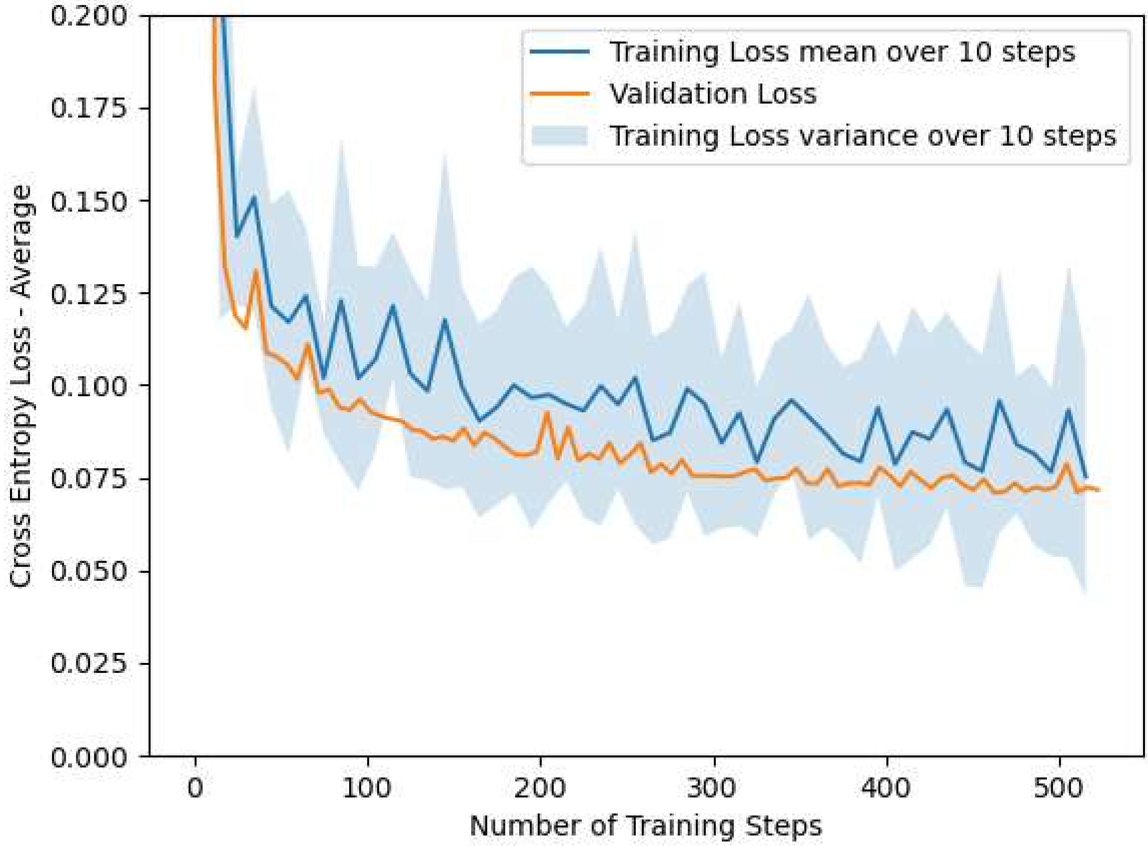
Appended

task 1b)

APpended

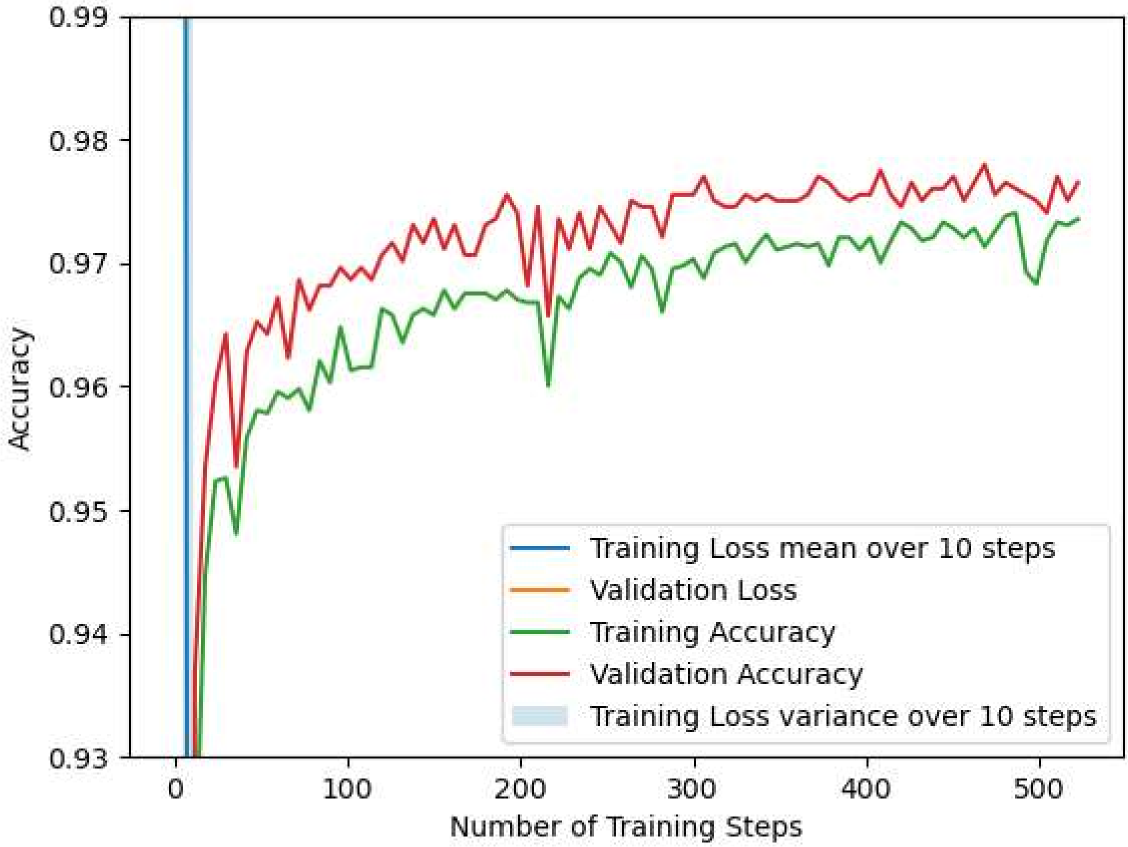
Task 2

Task 2b)



Training Loss mean over 10 steps
Validation Loss
Training Loss variance over 10 steps
Cross Entropy Loss - Average
Number of Training Steps
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
100
200
300
400
500

Task 2c)



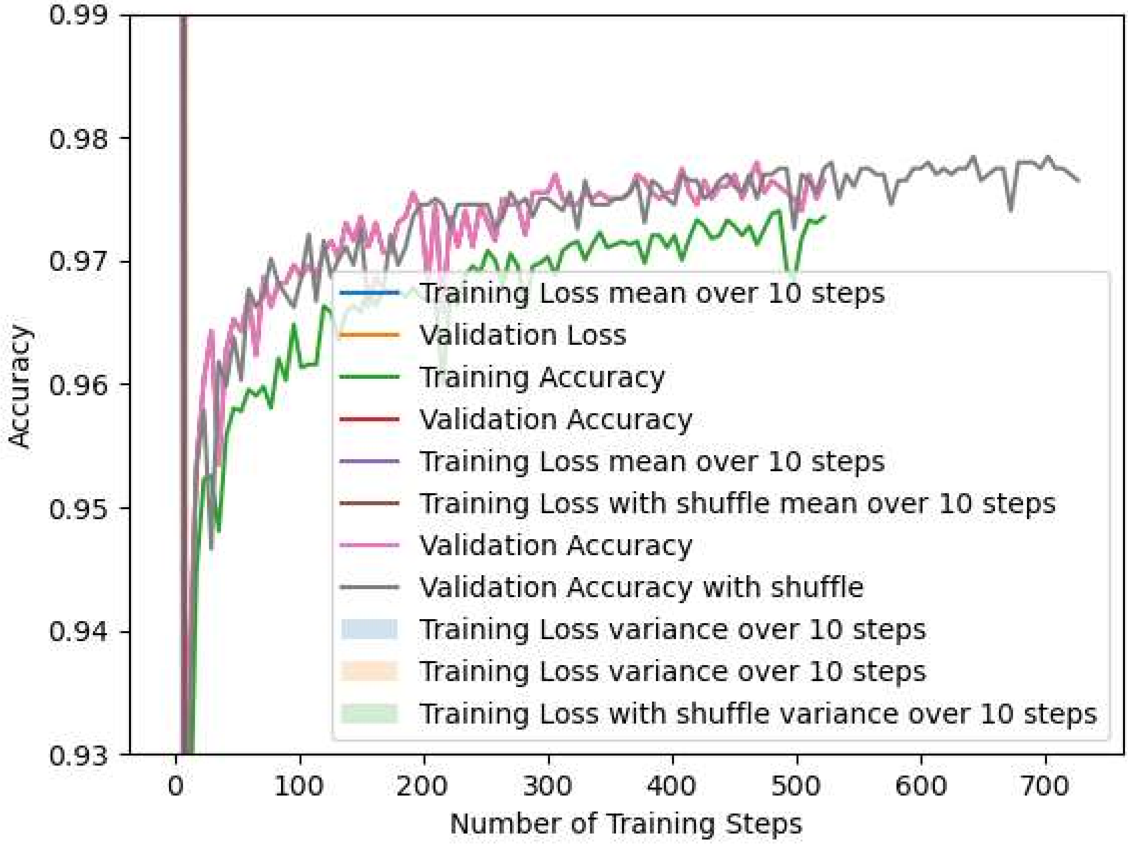
Training Loss mean over 10 steps
Validation Loss
Training Accuracy
Validation Accuracy
Training Loss variance over 10 steps
Accuracy
Number of Training Steps
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0
100
200
300
400
500

Task 2d)

Stops after 33 epochs

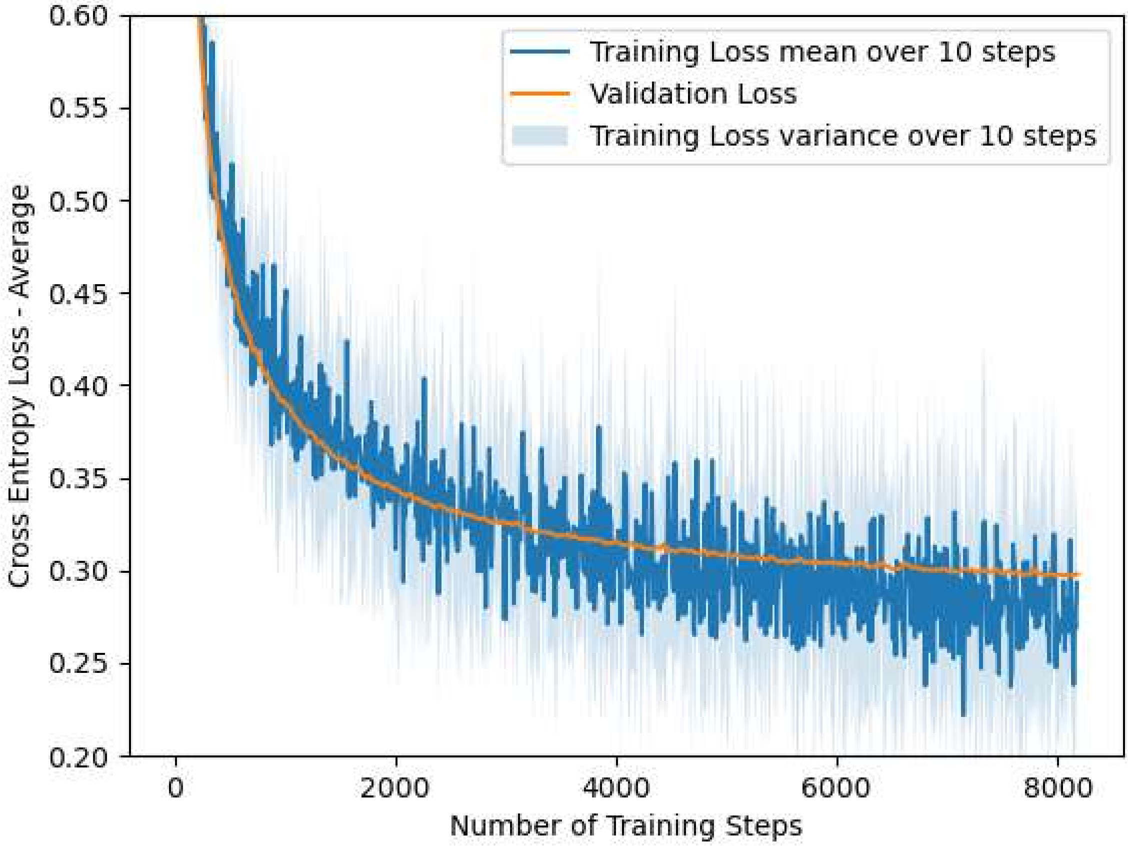
Task 2e)

Stopped after 16 epochs. More smood because it changes the order of trainig images every epoch

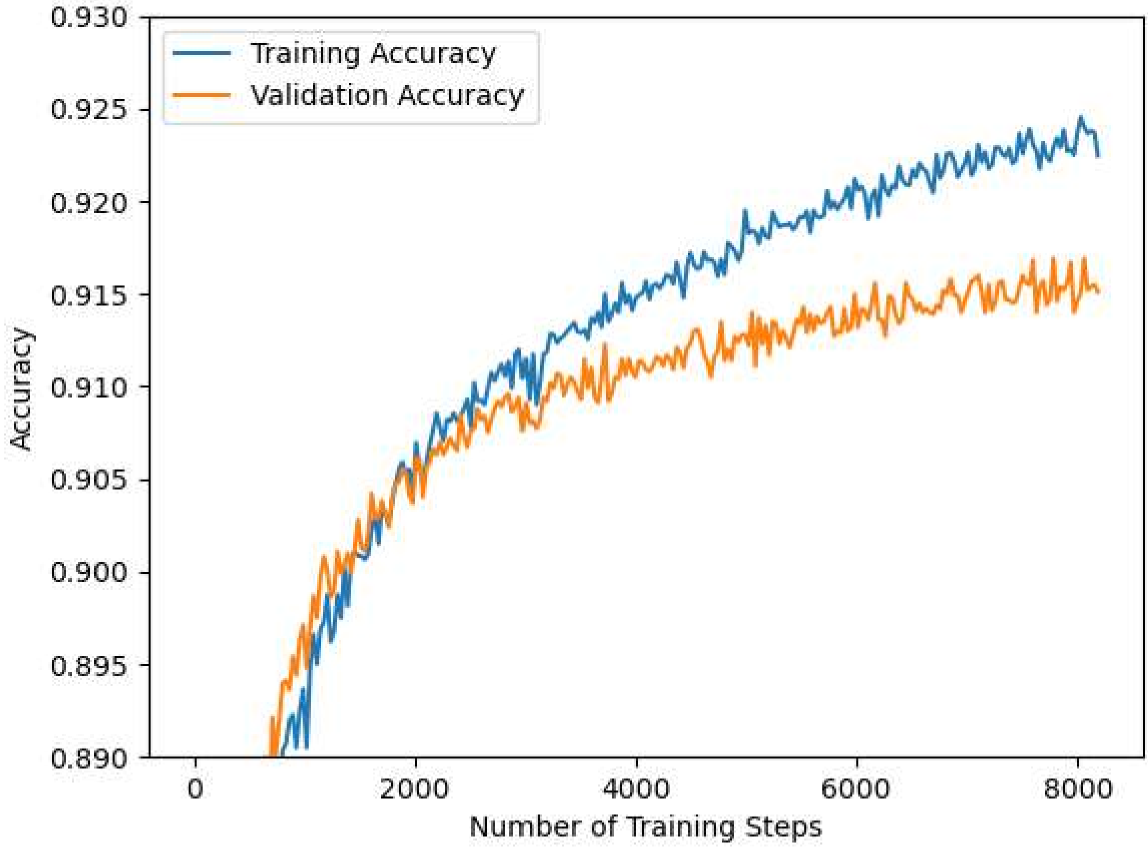


Task 3

Task 3b)


Training Loss mean over 10 steps
Validation Loss
Training Loss variance over 10 steps
Cross Entropy Loss - Average
Number of Training Steps

Task 3c)


Training Accuracy
Validation Accuracy
Accuracy
Number of Training Steps

Task 3d)

Yes! I belive there are overfitting because the training accuracy is increasing even tho the validation accuracy is flatting out! overfitting starts at around 3k steps!

Task 4

Task 4a)

Appended

Task 4b)

Cant do this. Training do not work as intended and loss is greater then 1?!

Task 4c)

FILL IN ANSWER

Task 4d)

Because the linear equation for exact back propegation is altered and not the ideeal correct expretion

Task 4e)

FILL IN ANSWER

Task 1

a) $C(w) = \frac{1}{N} \sum_{n=1}^{N} C^n$

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

$$\frac{\partial f(x^n)}{\partial w_i} = x_i^n f(x^n)(1 - f(x^n))$$

$$\frac{\partial C^n(w)}{\partial w_i} = x_i^n \, C(w^n)(1 - C(w^n))$$

b) $$\frac{\partial C^n(w)}{\partial w_{kj}} = \frac{\partial C}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w}$$

$$\left.\frac{\partial C}{\partial a}\right|_{a=\hat{y}=f(z)} = \frac{\partial\left(-\sum_{k=1}^{K} y_k^n \ln(\hat{y}_k^n)\right)}{\partial \hat{y}_k^n}$$

$$i=j, \quad \frac{\partial \hat{y}_i}{\partial x_i} = \frac{e^{x_j}\sum_k e^{x_k} - e^{x_i}e^{x_i}}{(\sum_k e^{x_k})^2} = \hat{y}_i(1-y_i)$$

$$i\neq j, \quad \frac{\partial y_i}{\partial x_j} = \frac{0 - e^{x_i}e^{x_j}}{(\sum_k e^{x_k})^2} = -\hat{y}_i\hat{y}_j$$

$$\frac{\partial C}{\partial y_i} = -\sum_i y_i \frac{1}{\hat{y}_i}$$

$$\frac{\partial C}{\partial x_j} = -\sum_{i\neq j} y_i \frac{1}{y_i}\frac{\partial y_i}{\partial x_j} - y_j\frac{1}{\hat{y}_j}\frac{\partial y_j}{\partial x_j}$$

$$= -\sum_{i\neq j} y_i \frac{1}{\hat{y}_i}(-\hat{y}_i(-\hat{y}_i y_j)) - y_i\frac{1}{\hat{y}_j}\hat{y}_j(1-\hat{y}_j)$$

$$= y_j - y_i$$

$$\Rightarrow \frac{\partial C}{\partial a} = -x_j(y_i - \hat{y}_i)$$

4a)

$$R(w) = \|w\|^2 = \frac{1}{2}\sum_{i,j} w_{i,j}^2$$

$$J = C(w) + \lambda R(w)$$

$$\frac{\partial J}{\partial w} = \frac{\partial C}{\partial w} + \frac{\partial\left(\lambda \frac{1}{2}\sum_{i,j} w_{i,j}^2\right)}{\partial w} = \frac{\partial C}{\partial w} + \lambda \cdot w$$

already found!

X_train : Images for training

Y_train : lables for training

Image1 = Image1.reshape(28, 28)

plt.imshow(Image1)

$\hat{y}$: prediction

y : ground truth

Gradient descent:

$$W_{t+1} = W_t - \alpha \frac{\partial \hat{L}(w)}{\partial W}$$

loss plot

Y-axis : avg loss

X-axis : Number of training steps

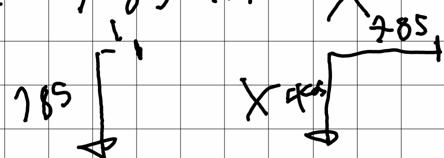
X : 4005×785

rader

kolumner

W : 785×1

$WX = 785 \times 1 \quad X \quad 4005 \times 785$



XW går

$W^T X = 1 \times 785 \quad \times \quad 4005 \times 785$