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). 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`.
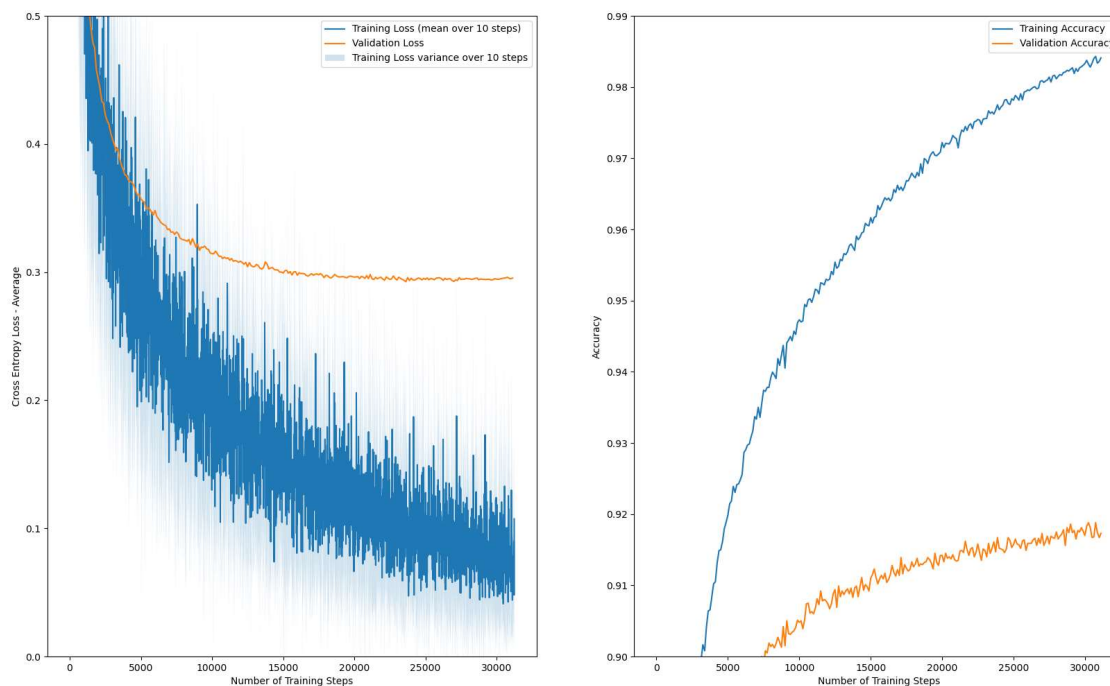
Task 1

task 1a)

Fill in task 1a image of hand-written notes which are easy to read, or latex equations here

task 1a)

Fill in task 1a image of hand-written notes which are easy to read, or latex equations here
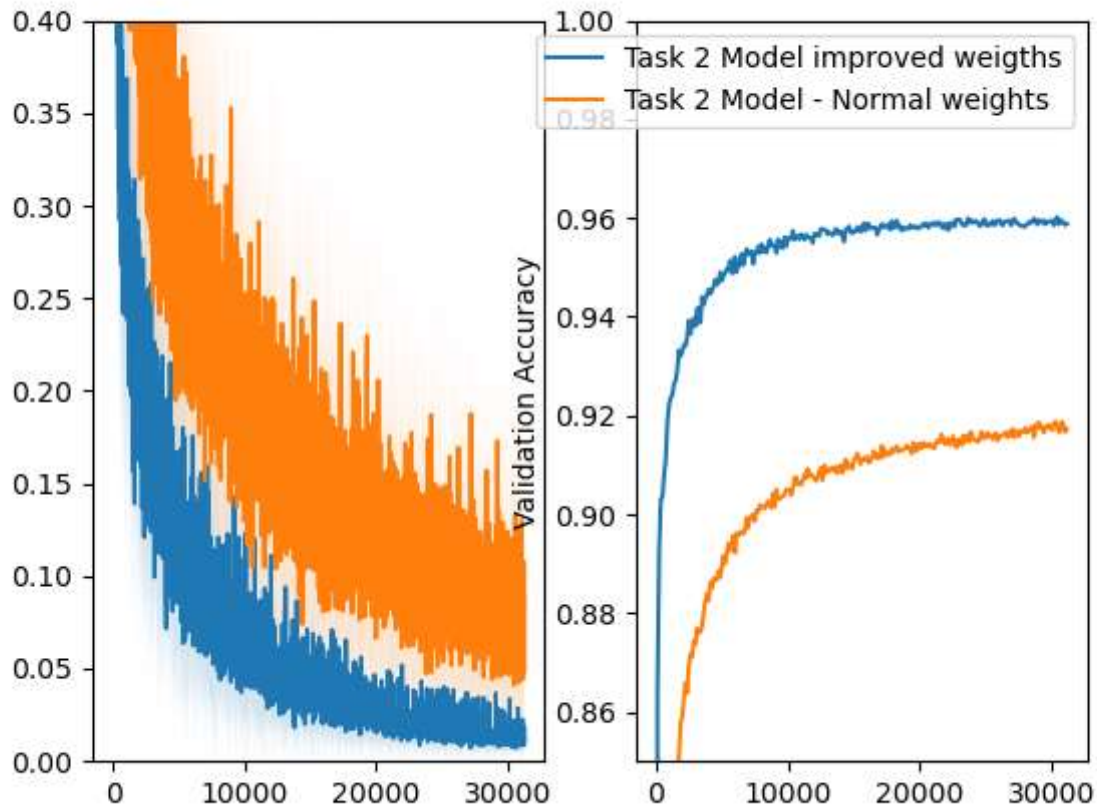
# Task 2

Task 2c)

Task 2d)

Parameters = num weights + biases = (50240 + 640) = 50880 Inspected from debugger

Task 3

Taask 3a) First improved weights. This results in inmproved convergance rate, and increased accuracy.

Task 3b) Both uses improved weights from task a. We can clearly se that both accuracy and loss is further improved by using the improved sigmoid funciton! Increased valdidation accuracy also suggest that we have less overfitting/memorizing of the learning data. Convergance rate is also further improved! Derivation of the derivative is shown in the handwritten notes!
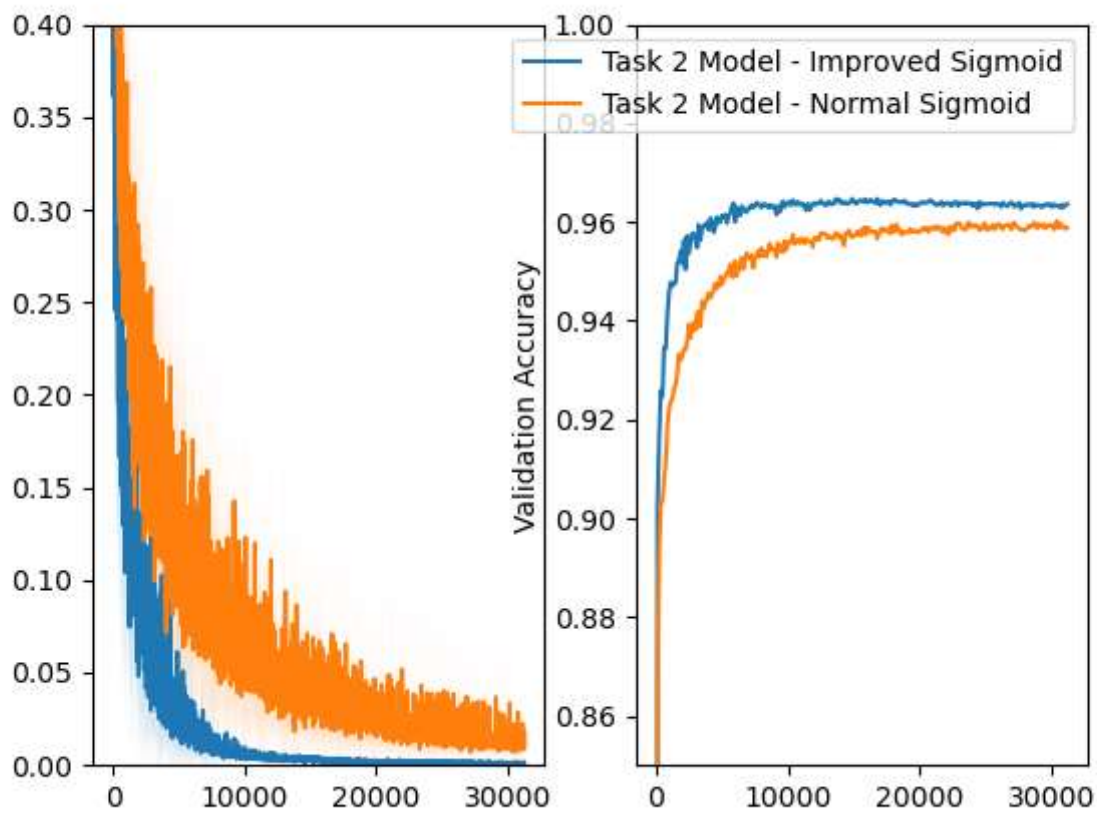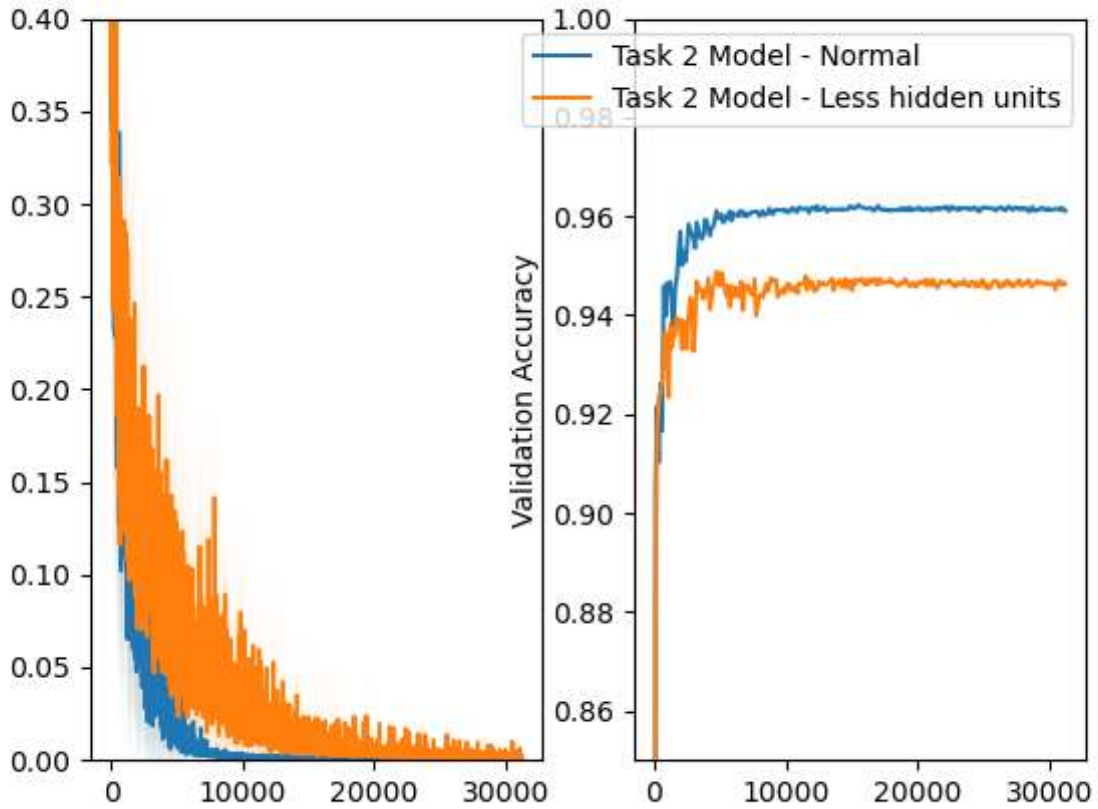
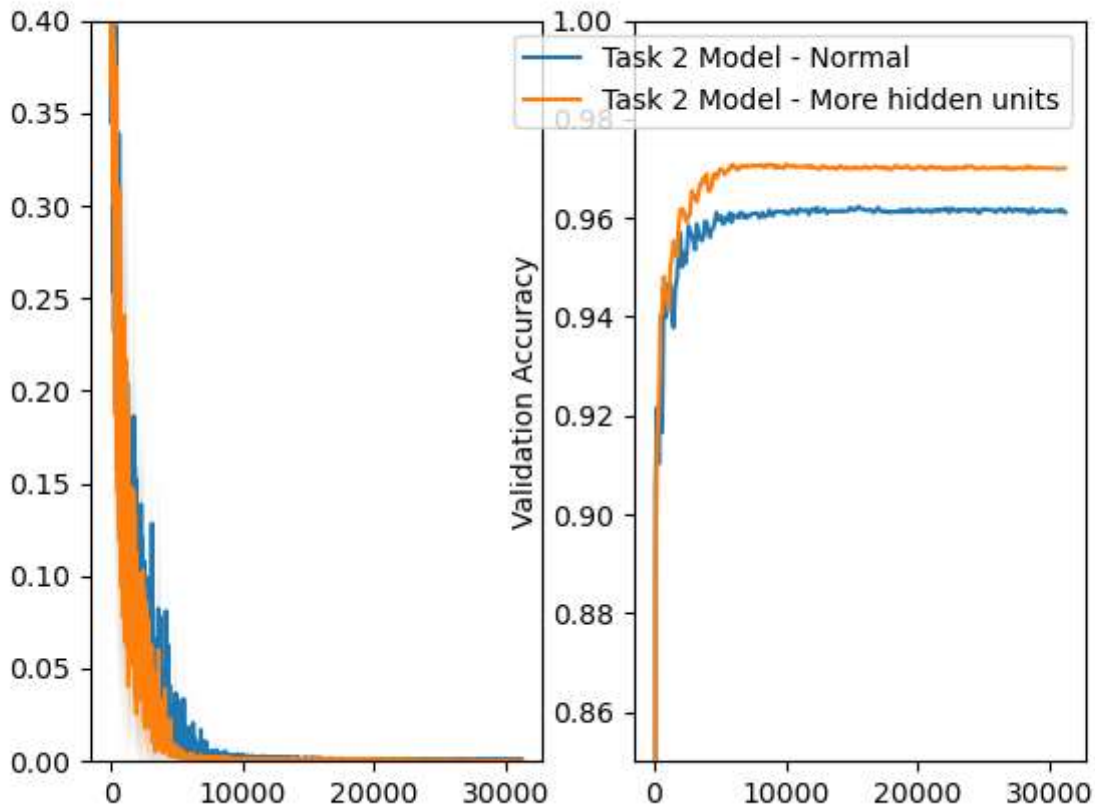Task 3c) Both useses improvements from a and b.
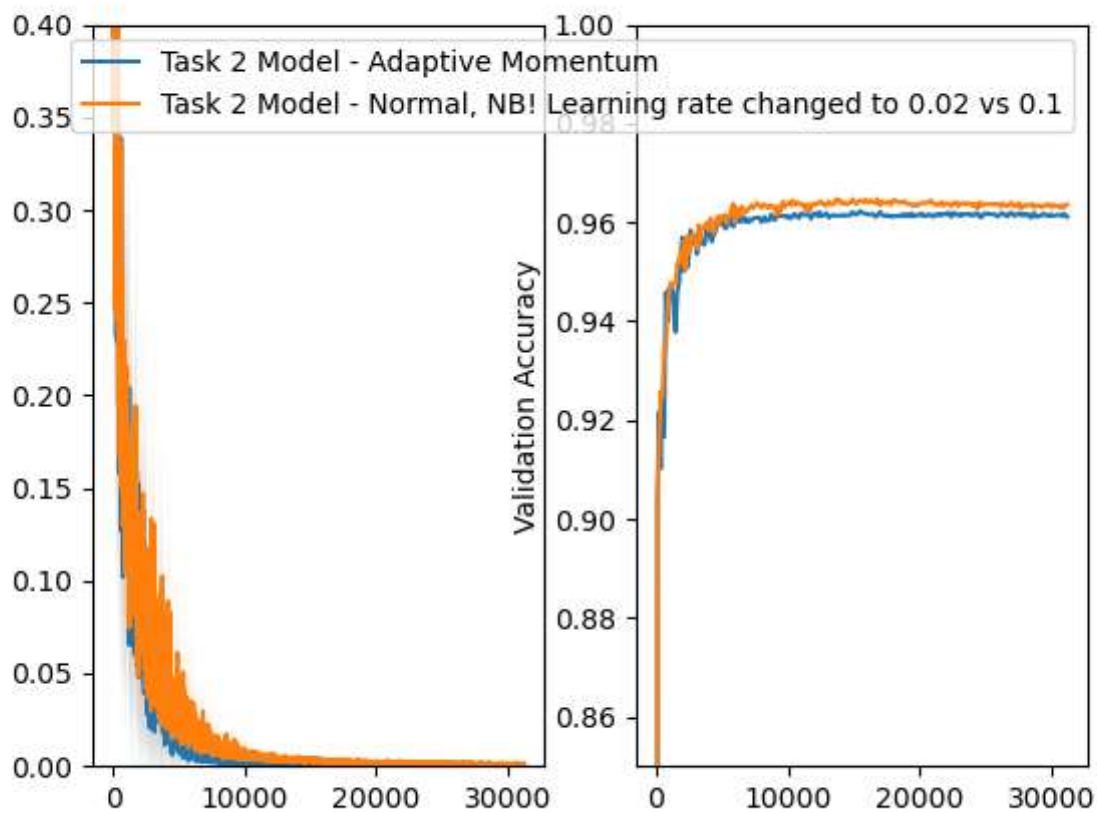
Task 4

Task 4a)

The acurracy gets worse even tho the loss stil convergece to zero

Task 4b)

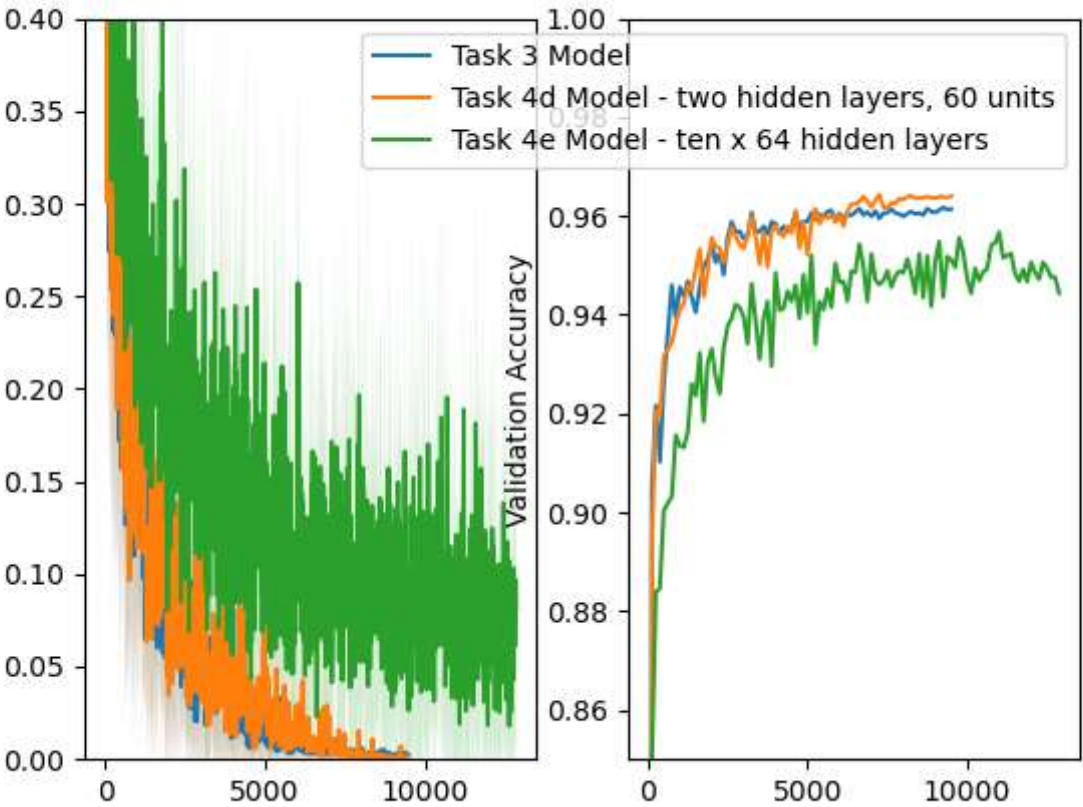The acurracy gets worse even tho the loss stil convergece to zero. The running time also increse.

Task 4d)

This plot compares one hidden layers with 64 units to one with two hidden layers with 60 units. These shouls have aproximatly the same number of parameters.Parameters = num weights + biases = $785 \cdot 60 + 60 \cdot 60 + 60*10$ = 51300 witch is close to the previous number of parameters! The one with two hidden layers is a little better! The difference is small and may be due to a little more parameters then the otehr!
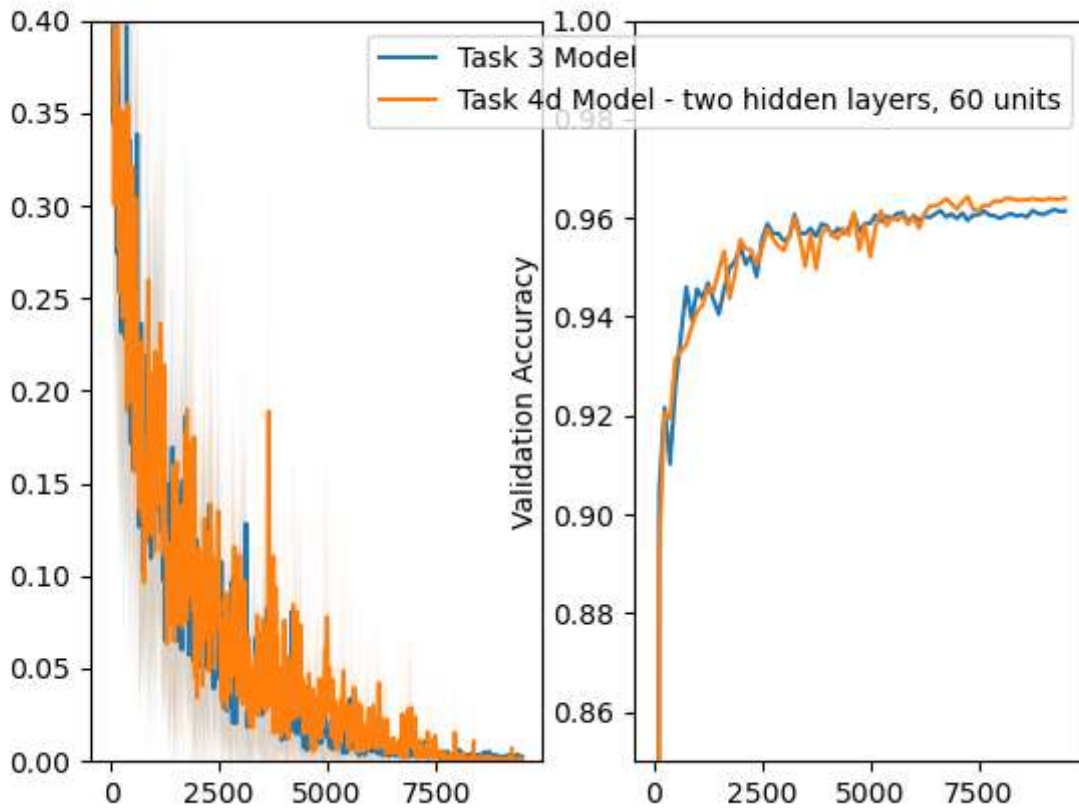
Task 4e)

Model from task 3 vs 4d vs 4e. Model from 4e has alot more paramters and the running time was really large. More parameters alows the model to learn training data better, however it may increase the chance for overfitting. It can be seen that model 3 with just one hidden layer, and the one with two outperformes ten! And i requires alot less computing! Early stopping was implemented to cut in computation time. Run time without was 342.85s Runtime with was 132.17s.

In [ ]: