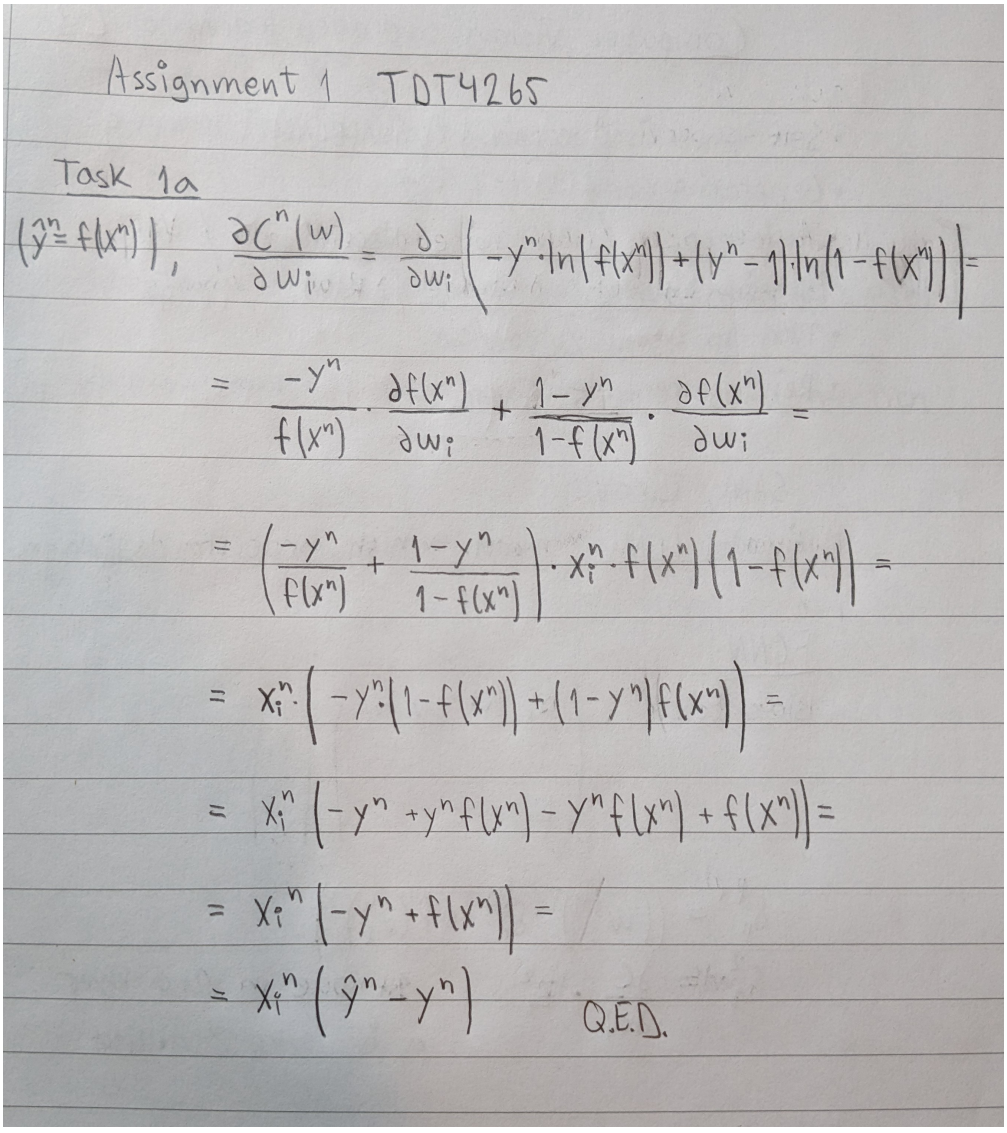
Assignment 1 Report Olof Ljunggren

TDT4265 Computer vision and deep learning

Task 1

Task 1a)



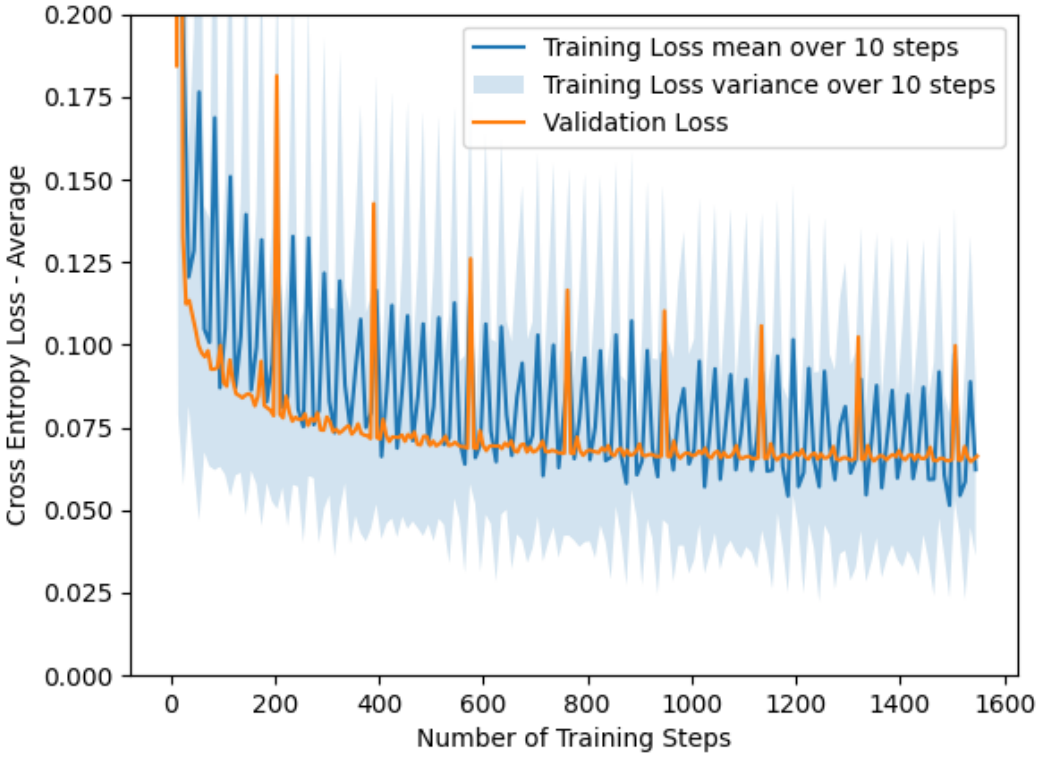
Task 2

Task 2a)

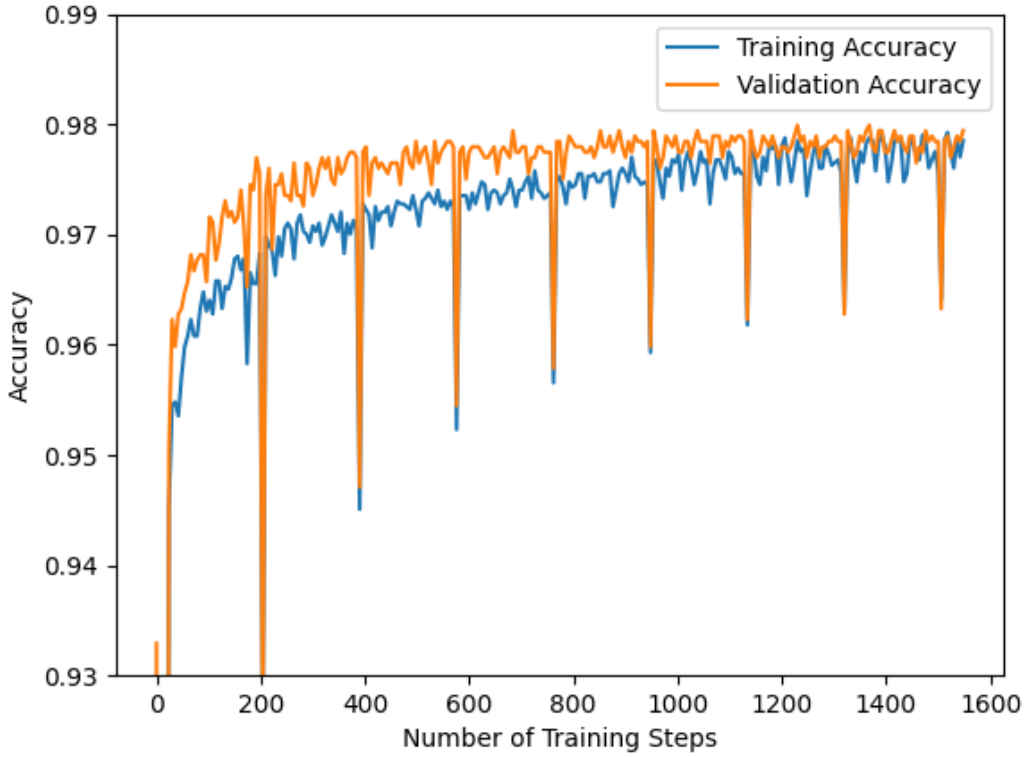
- Implementation of normalization, forward propagation, backward propagation and cross entropy loss was done here.

Task 2b)

- Figure with training and validation loss:



Task 2c)

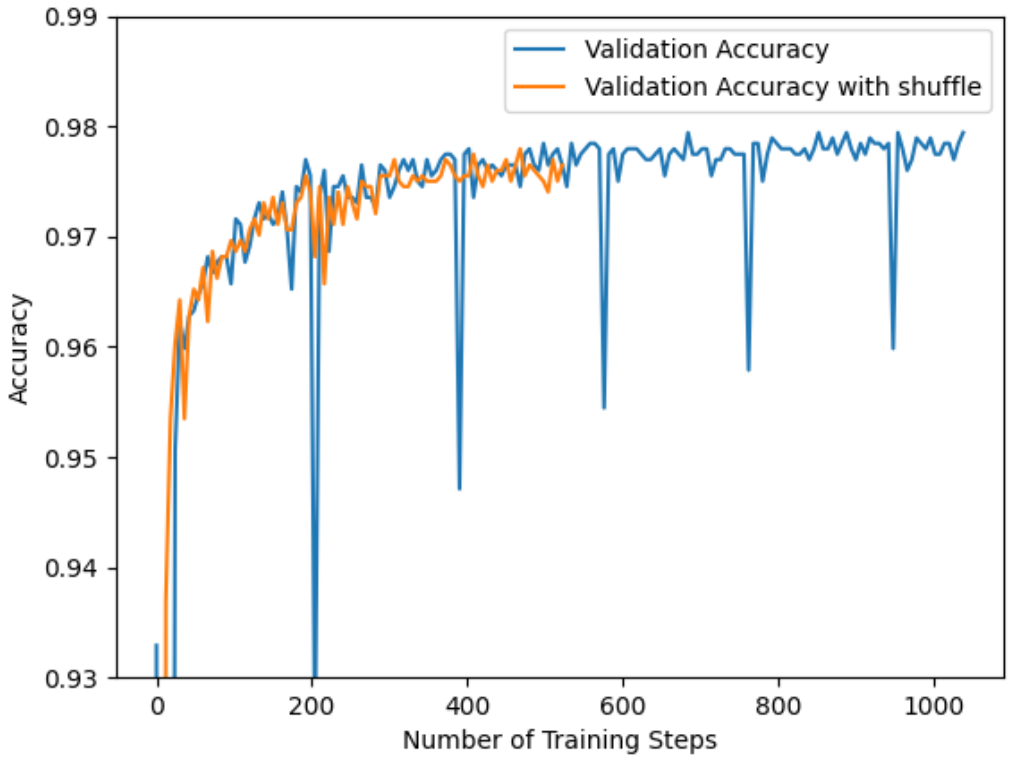


Task 2d)

- The early stop kicked in at epoch 33. (Started counting at 0). This was done by checking if the lowest of the 10 recent values was the oldest. In that case the loss has not improved for last 10 values.

Task 2e)

- Figure with shuffled vs unshuffled dataset. The spikes appears in the unshuffled set because it kind of lose generalization. There will probably be spikes when we get some data points from number 2 and some from set with number 3. Since the data is not shuffled the structure of the initial data matters.



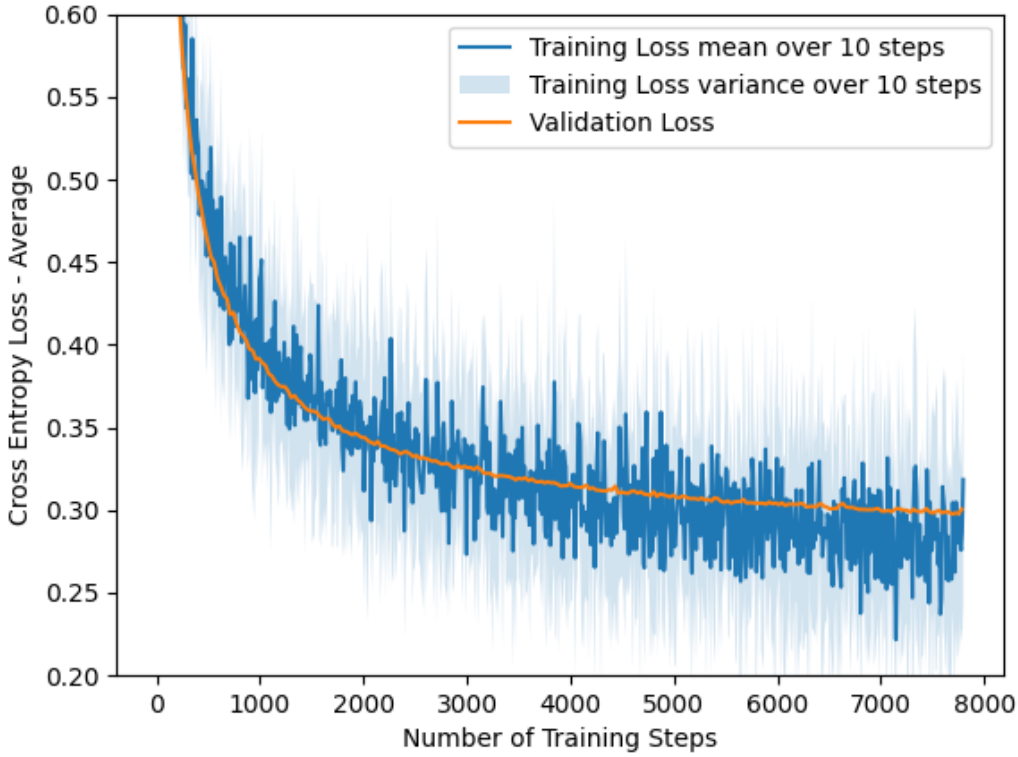
Task 3

Task 2a)

- Implementation of one-hot encoding, forward propagation, backward propagation and multiple label cross entropy loss was done here.

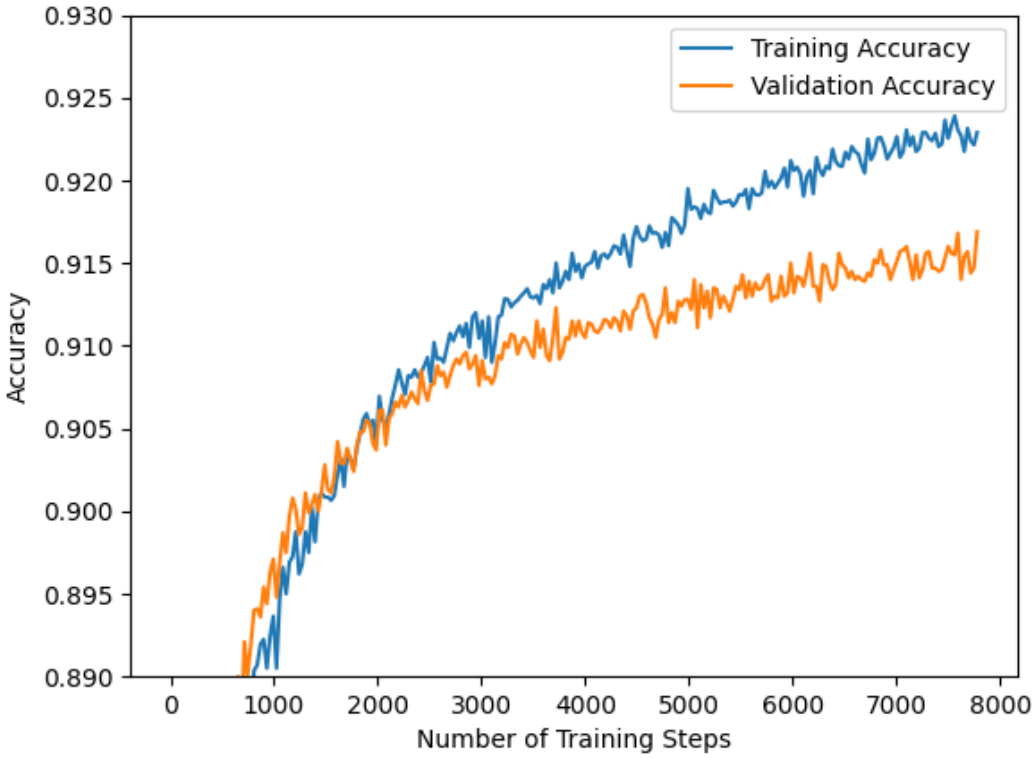
Task 3b)

- Figure with training and validation loss for the softmax regression problem:



Task 3c)

- Figure with training and validation accuracy for the multilabel regression problem.



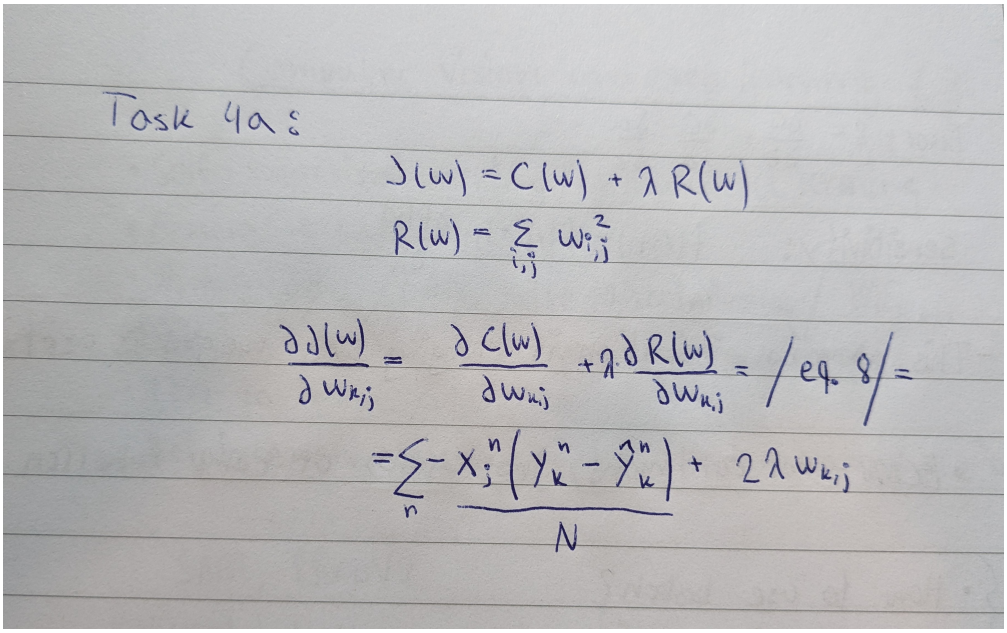
Task 3d)

We see some signs of overfitting. This can be seen by looking at the diffrence between the both accuracies. Since the validation data is not improving as much as the training data we are probably overfitting the data to the training set.

Task 4

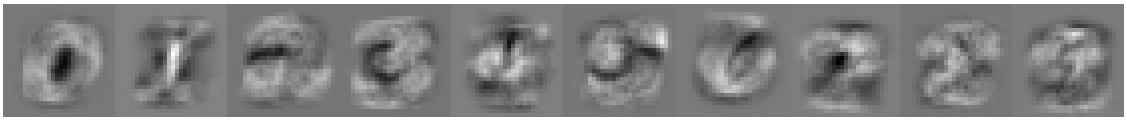
Task 4a)

- Since derivation is associative the cost from the multi-class cross-entropy cost will stay the same but we will have an extra term. This new term will punish large weights.

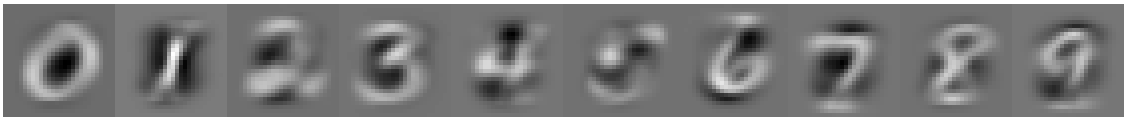


Task 4b)

- Figure with lambda = 0:



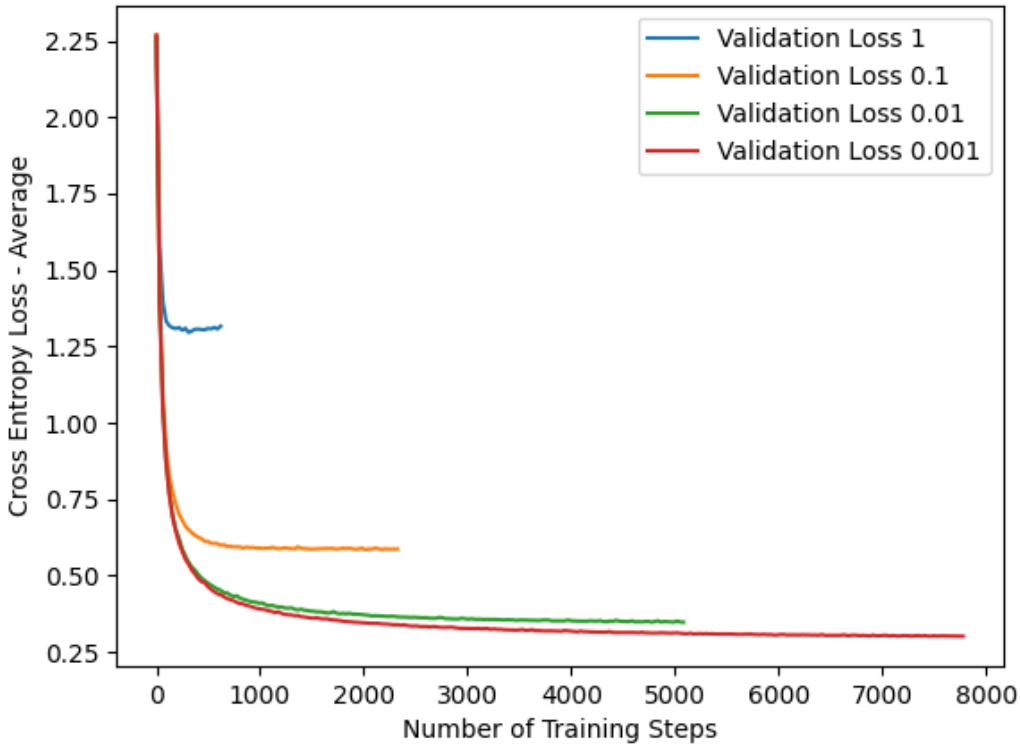
- Figure with lambda = 1:



- It is less noisy weigths in the second case since the lambda = 1 penalize high weigths which corresponds to high frequenciers and noise.

Task 4c)

- Figure with validation loss for diffrent values for lambda. (lambda: 1.0, 0.1, 0.01, 0.001)

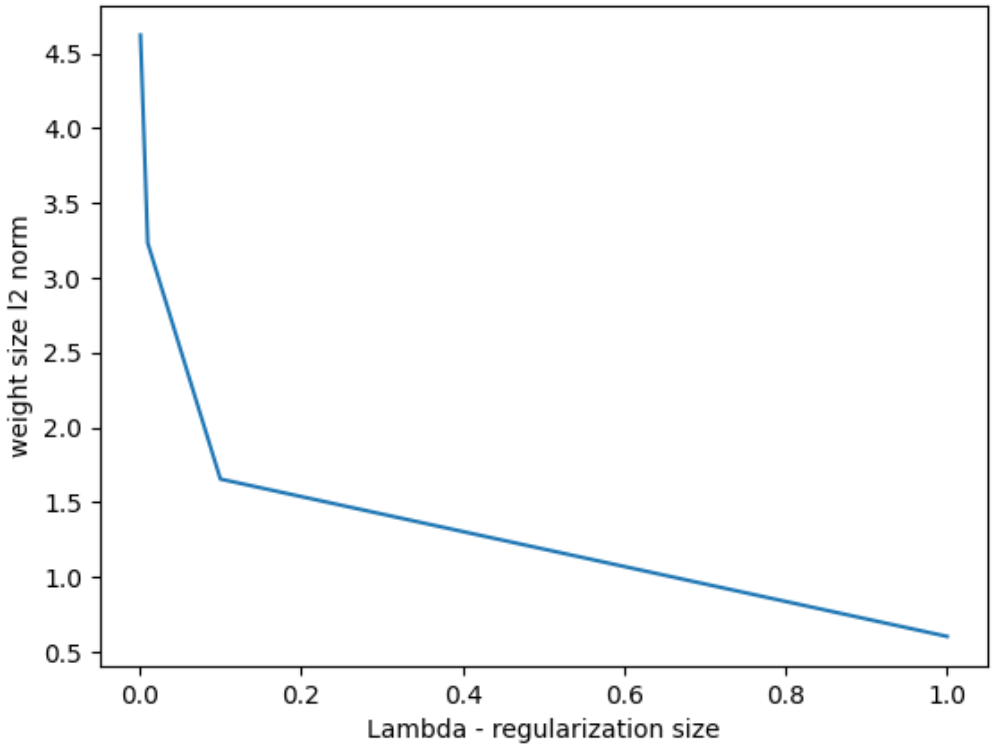


Task 4d)

- The precision degrades with increasing lambda and it penalizes complex solutions. This gives generalization but of course simplifies models. I believe this simplification is the main reason for this.

Task 4e)

- Figure with penalizing factor lambda plotted against the Frobenius norm size of the trained weigths. As expected the value of the generalization constant lambda directly affects the size of the weigths.


weight size l2 norm
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Lambda - regularization size