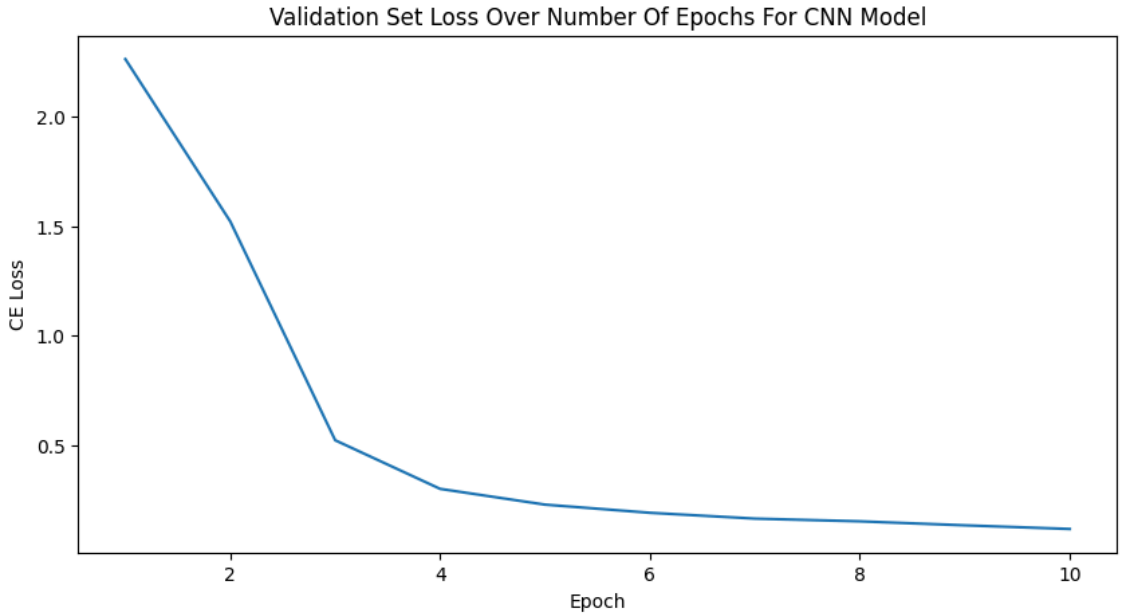
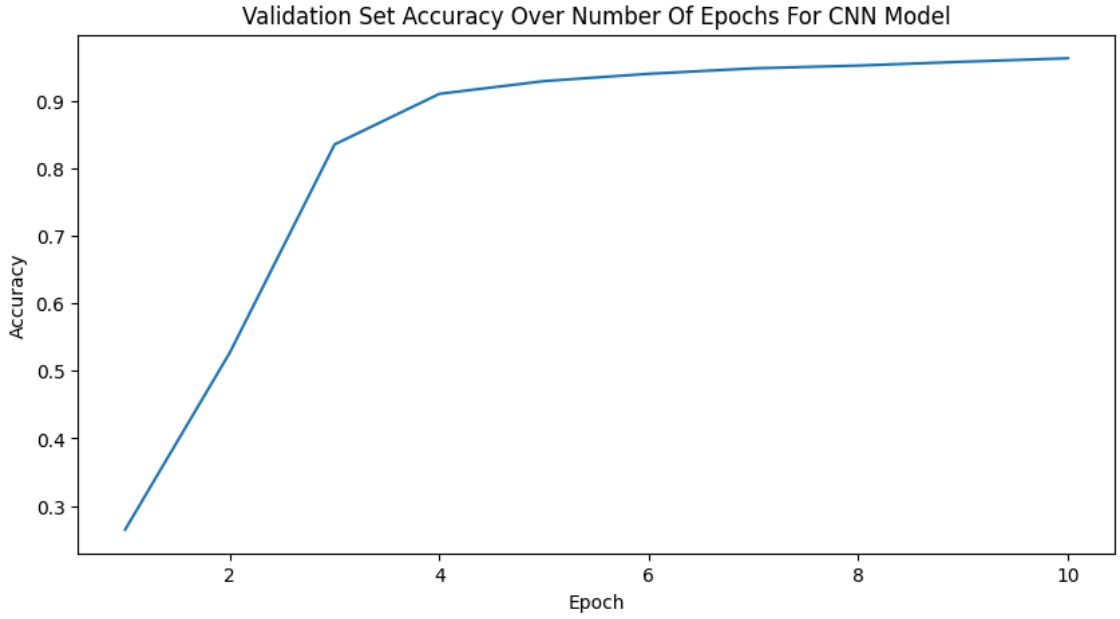
AML- Exercise 1

Ido Leder 206977852

2- Supervised Classification:

The plots for the loss and accuracy of the CNN model on the test set:

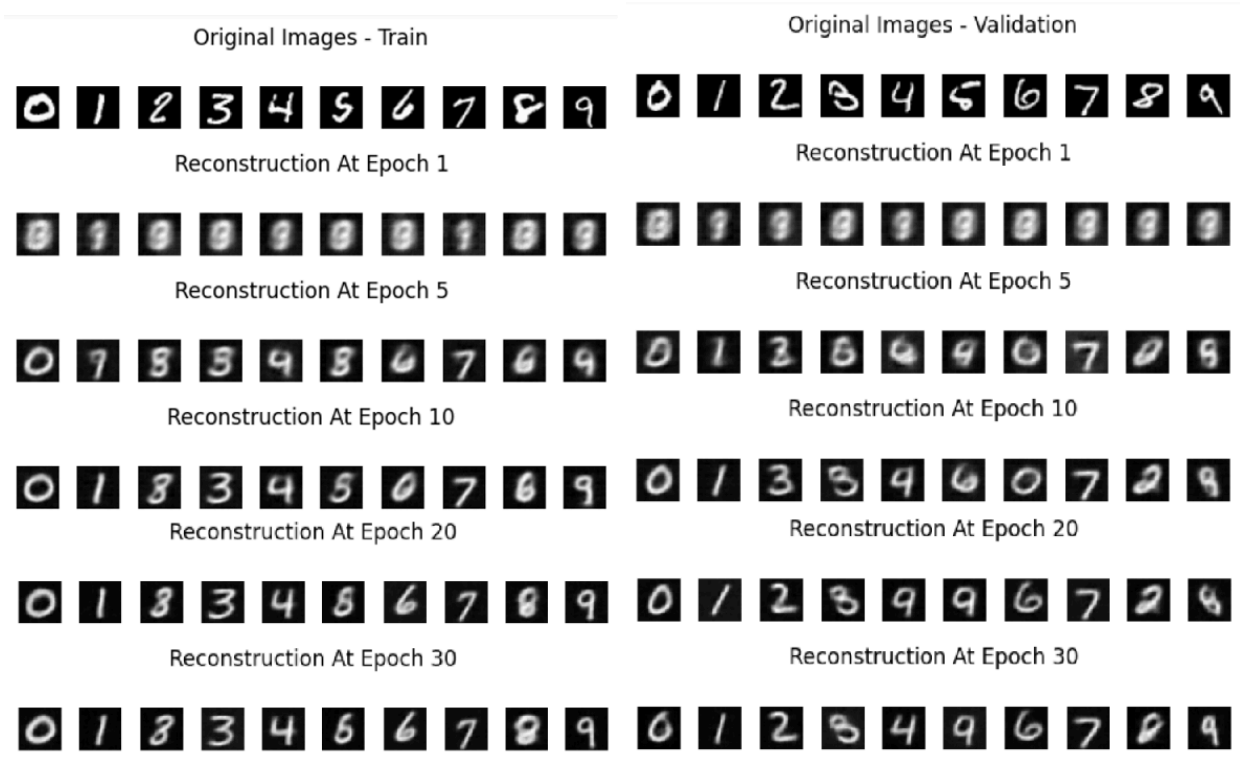




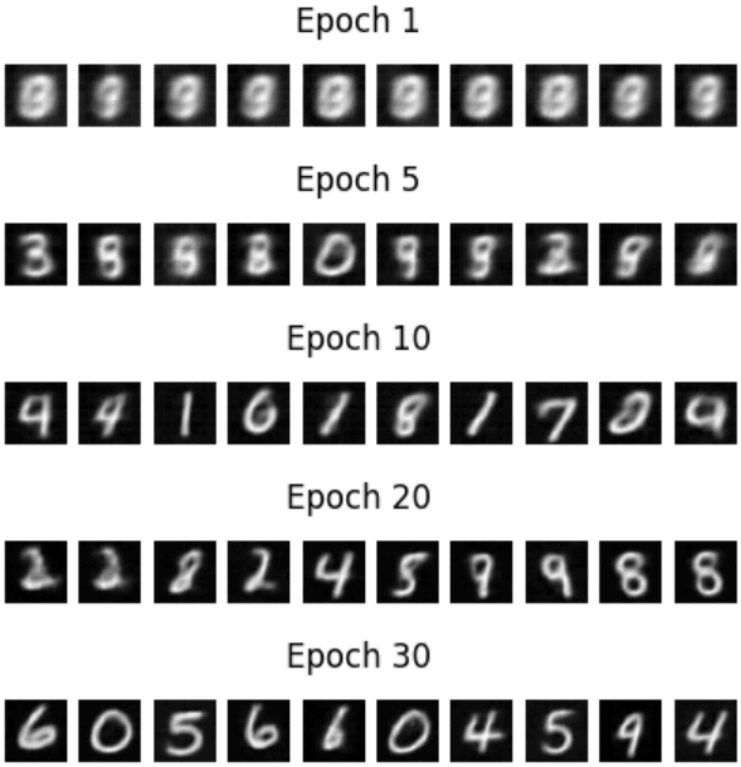
3- VAE:

Q1:

It doesn’t seem like the auto-encoder overfit the train data, since its reconstructions on the test set get better as the epochs progress and end up being pretty similar in quality to the reconstructions on the train set.

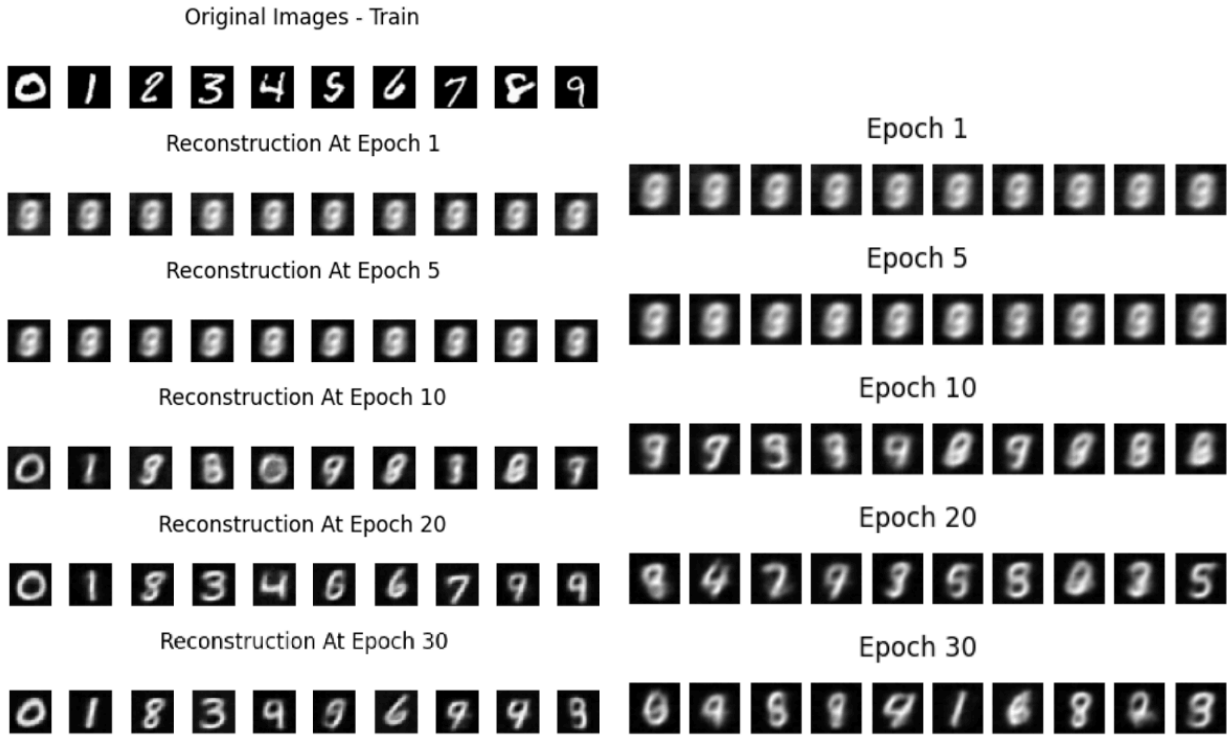


Q2: These are the samples by epoch:



Q3:

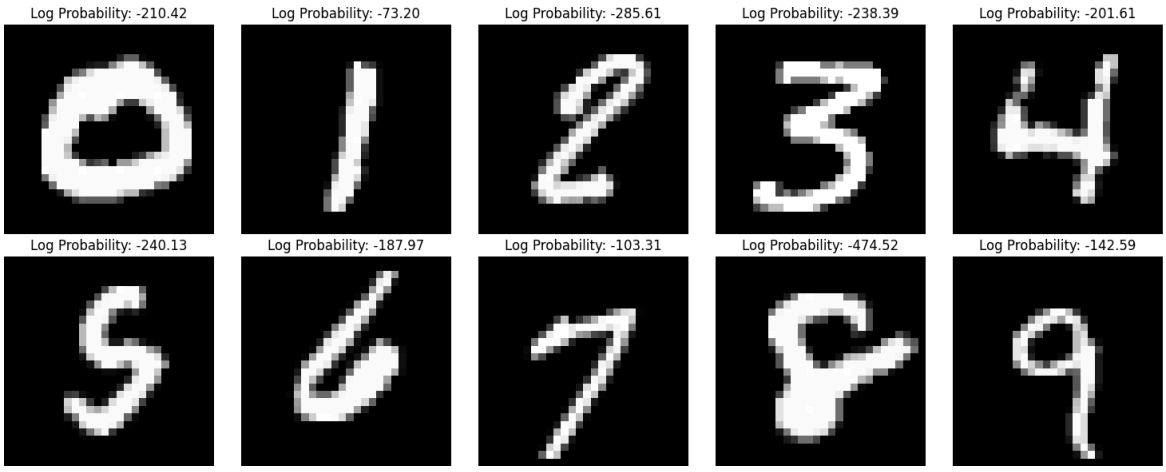
Reconstructions of train in the left image, and samples by epoch in the right



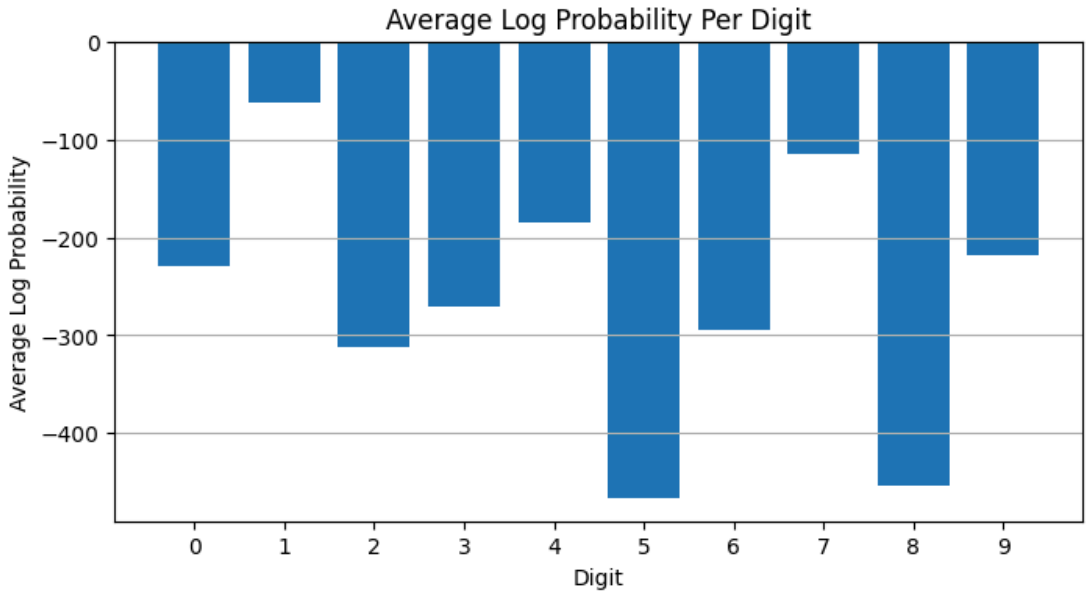
- As we can clearly see, the quality of the latent optimizer model reconstructions is worse than those of the amortized model. The latent model kept a separate q for each image in the train set but the decoder is still shared between them. The decoder is trying to generalize all of the images in the set to decrease reconstruction loss, but instead of receiving latent codes that are learnt together by the encoder (like in the amortized case), it receives latent codes that are learnt independently and might not resemble the true distribution. This leads to decreased reconstruction quality of some of the images, and to worse q vectors by the latent optimizer model.
- The samples with the latent optimization model are a lot less clear than the ones from the amortized model. Even after 30 epochs the samples are not all that close to resembling clear digits, like the ones we saw with the amortized model. We can derive that our initialization was not sufficient to establish a prior that approximates the real distribution well.

Q4:

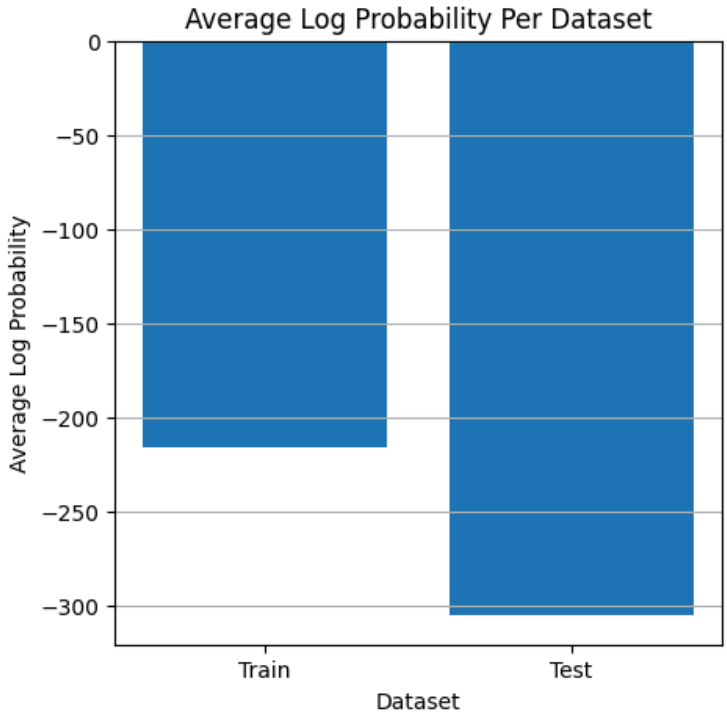
(a) Single image from each digit, sampled from the train set:



(b) The digit that got the highest log probability is 1, likely due to the fact that its shape is pretty unique(straight line when most others have curves) compared to the other digits, and it has less variation in the way people draw it. Therefore it was easier for our model to distinguish it from the other digits.



(c) The average log probability on the samples from the train dataset was higher than the average log probability on the samples from the test dataset. This aligns with the fact that the model used those samples to train, leading to it assigning them higher log probabilities.



Throughout the exercise I have used ChatGPT for help with plotting syntax, torch syntax, debugging and explanations of how the models work.