HW 8 Report

Original Images: 1, 7, 6, 5, 9

Effect of Training Epochs:

LR = 0.001

With layers: `[784, 392, 196, 98, 32, 98, 196, 392, 784]`

We tried early stopping with patience = 20 and Epochs = 400. Early stopping did not trigger Final training loss = 0.0080

Test Loss = 0.0314 hence high overfitting

Time taken was > 1 hour

For patience = 10 we get Early stopping at 178 epochs. Similar results obtained at patience = 7. Hence, we use Max Epochs = 300 now on

Effect of Learning Rate:

(with layers same as earlier with latent space = 32)

1) LR = 0.001:

Time taken = 39 minutes

Train loss = 0.012 and Test Loss = 0.029

2) Due to lack of BatchNorm and regularization, we find overflow at learning rate > 0.007. So we just show results for Learning Rate = 0.007:

Time = 20 mins

Train Loss = 0.047 and Test Loss = 0.057

Effect of Latent Space:

1) Size = 32: Mentioned earlier in case of Learning Rate 0.001

2) Size = 16:

Time = 50 mins

Train Loss = 0.013 and Test Loss = 0.043

3) Size = 4: Even for LR = 0.001 model cannot train. So we take LR 0.0005 now on:

```
[784, 392, 196, 98, 49, 24, 16, 4, 16, 24, 49, 98, 196, 392, 784]
```
Time = 1hr

Train loss= 0.073

Test loss = 0.079

Latent Space size 1:

```
[784, 392, 196, 98, 49, 24, 16, 4, 16, 24, 49, 98, 196, 392, 784]
```
Train loss = 0.110 and Test loss = 0.112 with time = 1 hr

PCA with latent Space = 32:

Discussion:

1) Very irregular training curve for high learning rates and low latent space size
2) Higher latent space and low learning rates with high epochs give best results
3) PCA severely underperforms as compared to nonlinear networks of the same size.
4) High Learning Rate is unable to train the network and gets stuck in a local minima