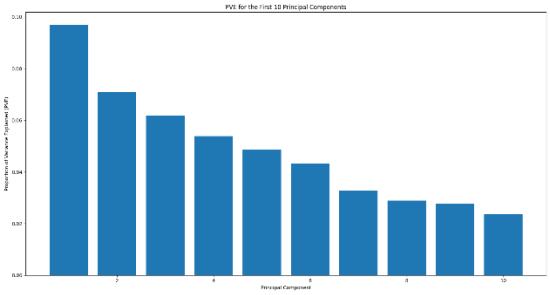
Question 1.1



Proportion of Variance Explained (PVE) for the first 10 principal components:

[0.09704664 0.07095924 0.06169089 0.05389419 0.04868797 0.04312231

0.0327193 0.02883895 0.02762029 0.02357001]

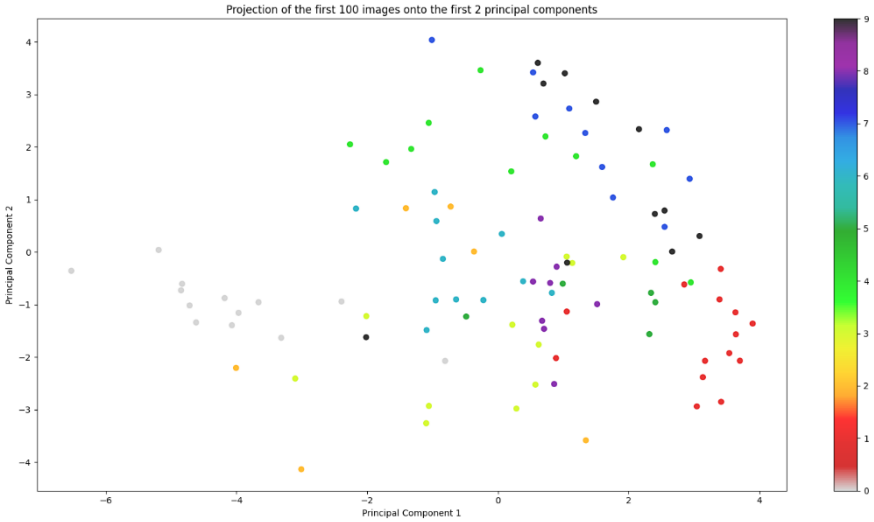
Question 1.2

Minimumu number of principle components to explain 70% of the data: 26

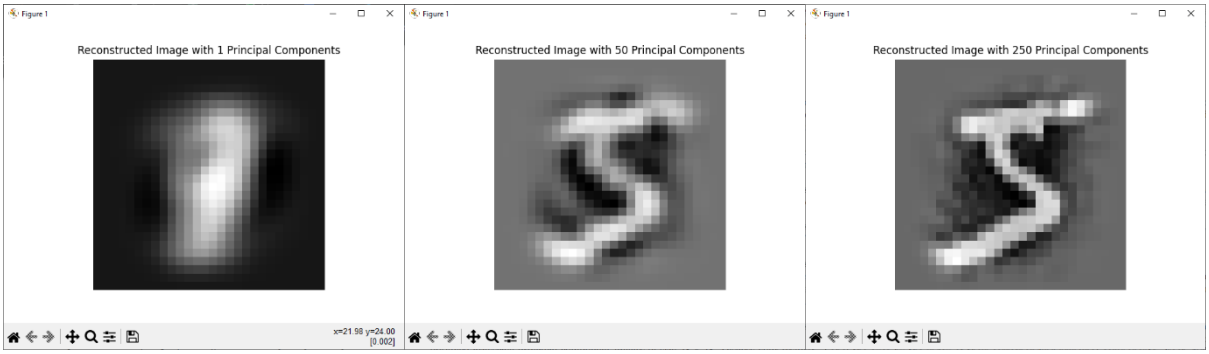
Question 1.3

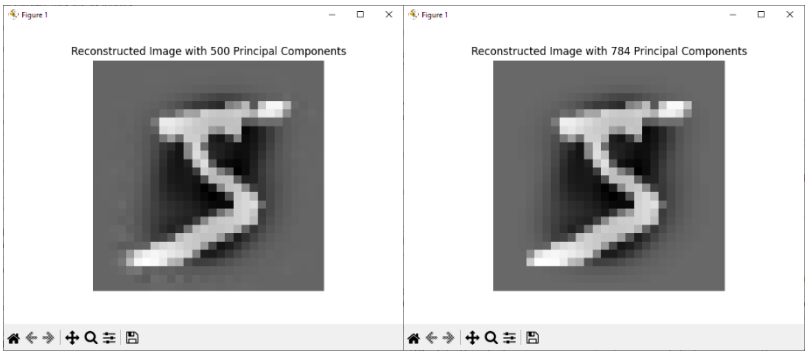
Looking at the 10 largest PC's we can see that each PC brings out a different feature from the images. For example, PC 1 looks like an oval that could be a part of 0 whereas PC 5 brings out strokes in the middle of the image.

Question 1.4



Question 1.5





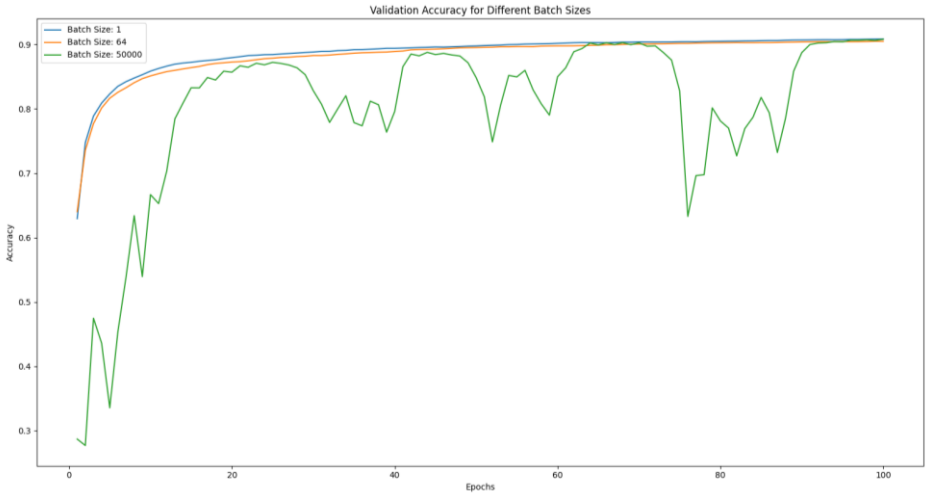
As the number of principal components increases, the image becomes more an more similar to the original image.

Question 2.1

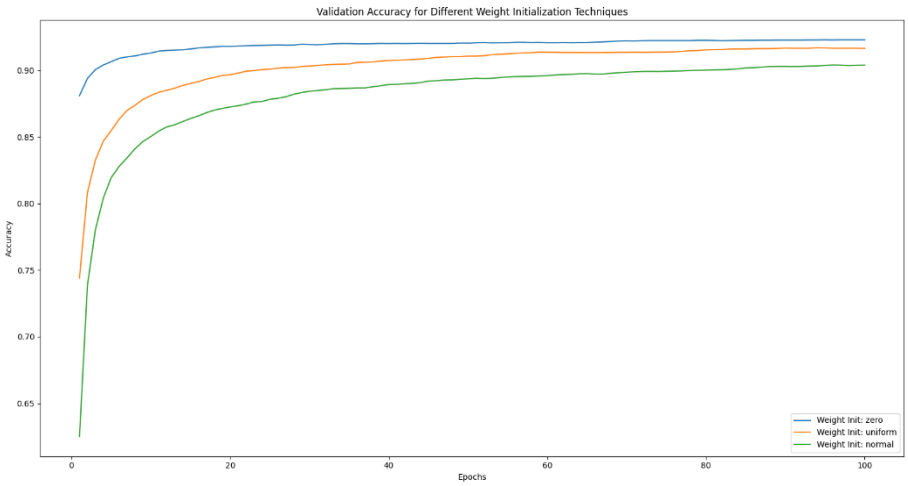
```
Test Accuracy: 90.39
Confusion Matrix:
 [[953   0   2   3   1  15  10   6   5   0]
 [  0 942  11   7   0   2   2   3  28   1]
 [  7  15 848  34  10   8  14  12  41   5]
 [  2   1  28 854   0  51   2  10  33  13]
 [  2   2   6   3 885   2  11   9  15  60]
 [ 15   2   8  50  16 801  16   9  67  10]
 [ 16   2  16   1  15  25 912   2   6   0]
 [  1   7  18  16  10   1   0 879   7  56]
 [  7  12  17  41  15  59  12  10 802  19]
 [  7   7   1  13  52   9   0  40  22 843]]
```

Question 2.2

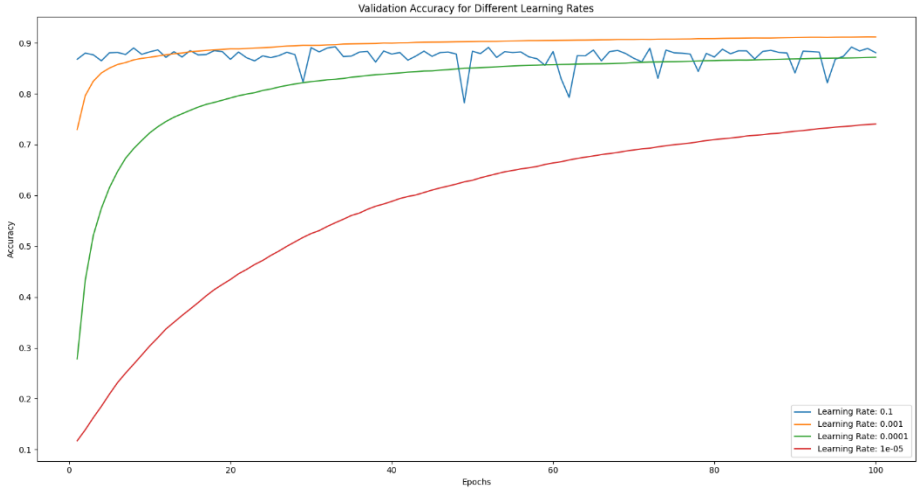
Batch Size Test Results



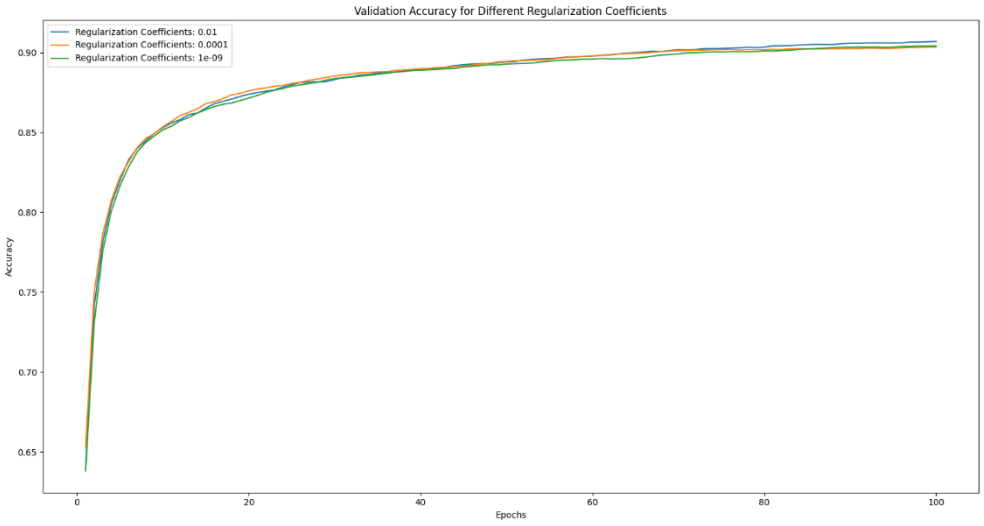
Weight Initialization Technique Test Results



Learning Rate Test Results



Regularization Coefficient Test Results



Question 2.3

batch_size = 64

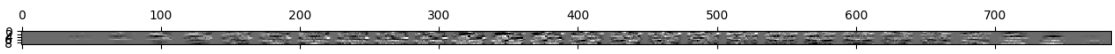
learning_rate = 1e-3

l2_reg_coefficient = 1e-2

weight initialization = zero initialization

```
Test Accuracy: 92.12
Confusion Matrix:
 [[958   0   2   3   0  14   8   3   5   1]
 [  0 947   8   7   0   3   3   2  25   1]
 [ 10  15 840  32   9   7  17  11  48   7]
 [  5   1  23 848   1  56   3  12  33  13]
 [  2   2   6   4 861   3  14   7  18  78]
 [ 16   3   8  50  14 799  17   8  68  13]
 [ 16   2  13   2  13  27 911   2   9   0]
 [  1   8  18  12   8   2   0 871   9  66]
 [  9  14  16  41  12  65  13  11 791  23]
 [  8   6   2  12  45  11   0  40  21 850]]
```

Question 2.4



Question 2.5

```
Class 0 - Precision: 0.9346, Recall: 0.9638, F1 Score: 0.9490, F2 Score: 0.9578
Class 1 - Precision: 0.9489, Recall: 0.9508, F1 Score: 0.9498, F2 Score: 0.9504
Class 2 - Precision: 0.8974, Recall: 0.8434, F1 Score: 0.8696, F2 Score: 0.8537
Class 3 - Precision: 0.8388, Recall: 0.8523, F1 Score: 0.8455, F2 Score: 0.8495
Class 4 - Precision: 0.8941, Recall: 0.8653, F1 Score: 0.8795, F2 Score: 0.8709
Class 5 - Precision: 0.8095, Recall: 0.8022, F1 Score: 0.8058, F2 Score: 0.8037
Class 6 - Precision: 0.9239, Recall: 0.9156, F1 Score: 0.9197, F2 Score: 0.9172
Class 7 - Precision: 0.9007, Recall: 0.8754, F1 Score: 0.8879, F2 Score: 0.8803
Class 8 - Precision: 0.7702, Recall: 0.7950, F1 Score: 0.7824, F2 Score: 0.7899
Class 9 - Precision: 0.8080, Recall: 0.8543, F1 Score: 0.8305, F2 Score: 0.8446
```

The model works with an accuracy around 0.85, with some classes performin better than others. Class 0 and 1 seem to be performing way better than other classes.