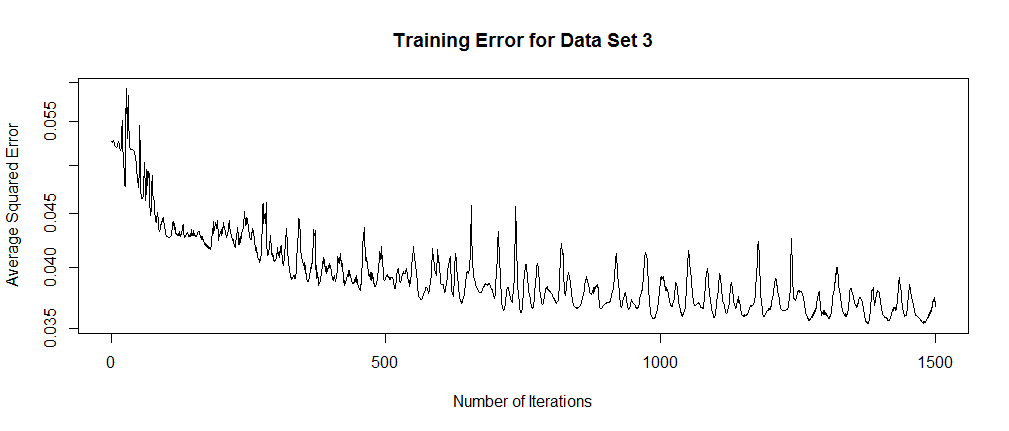
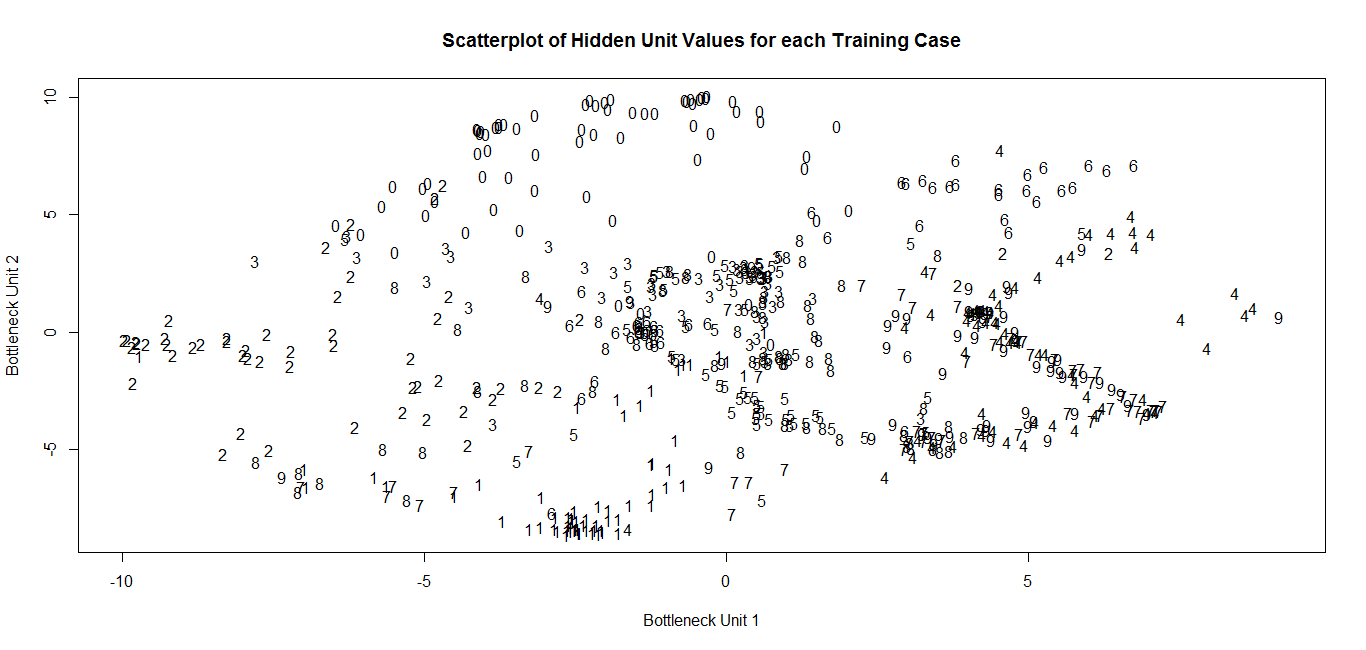
**MNIST Handwritten digit Data set – Dimensionality reduction using neural networks/ auto-encoders**

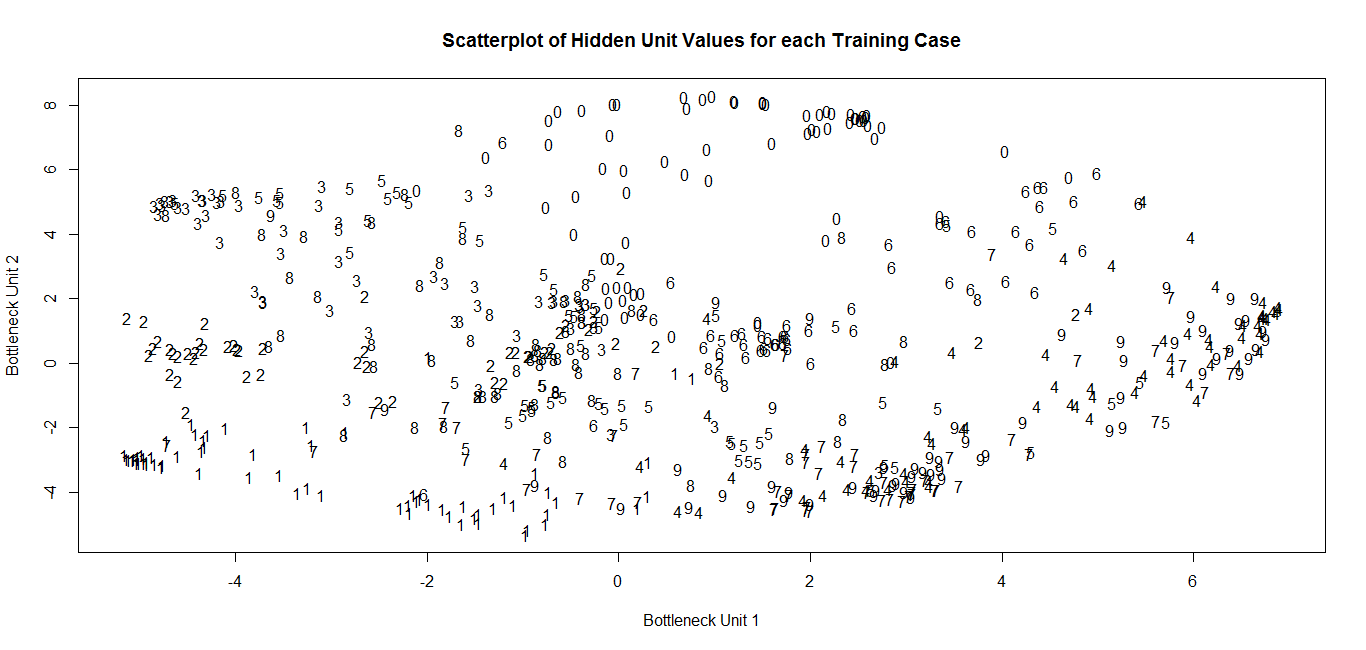
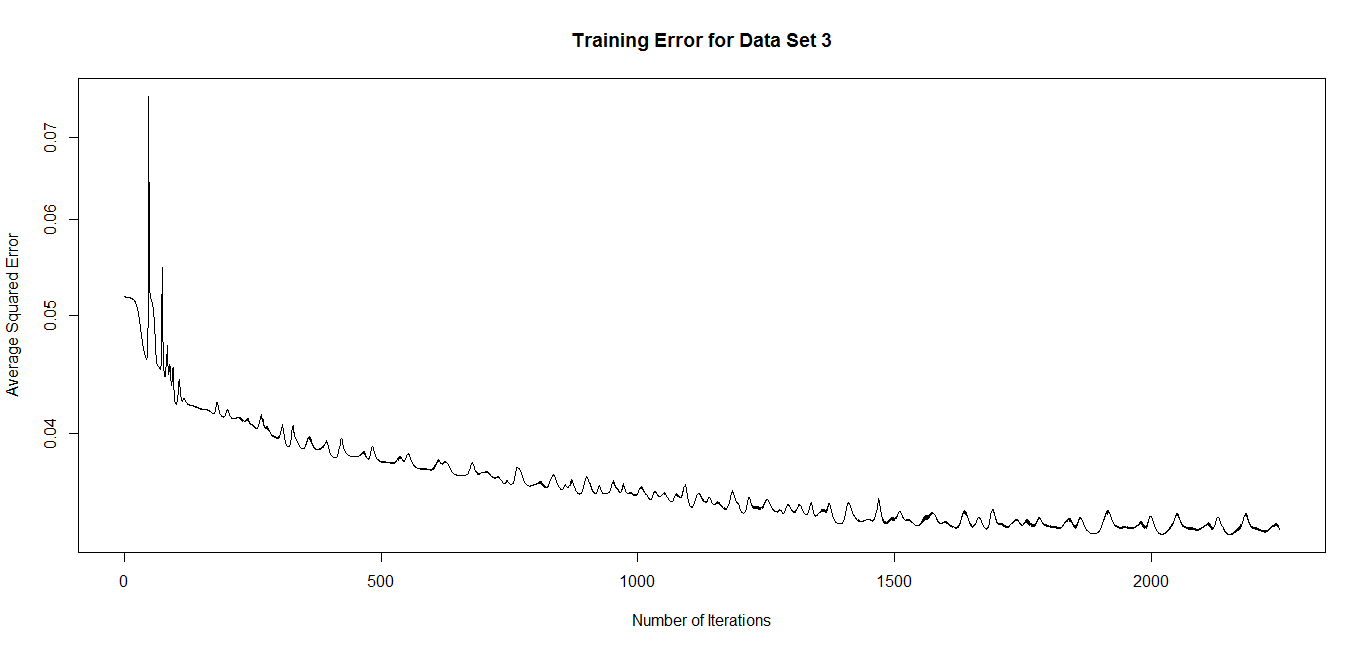
The purpose of this analysis is to explore the dimensionality reduction through trying out different iterations and learning rates on a dataset of handwritten digits (as represented by 196 values) through the use of a multilayer perceptron network.

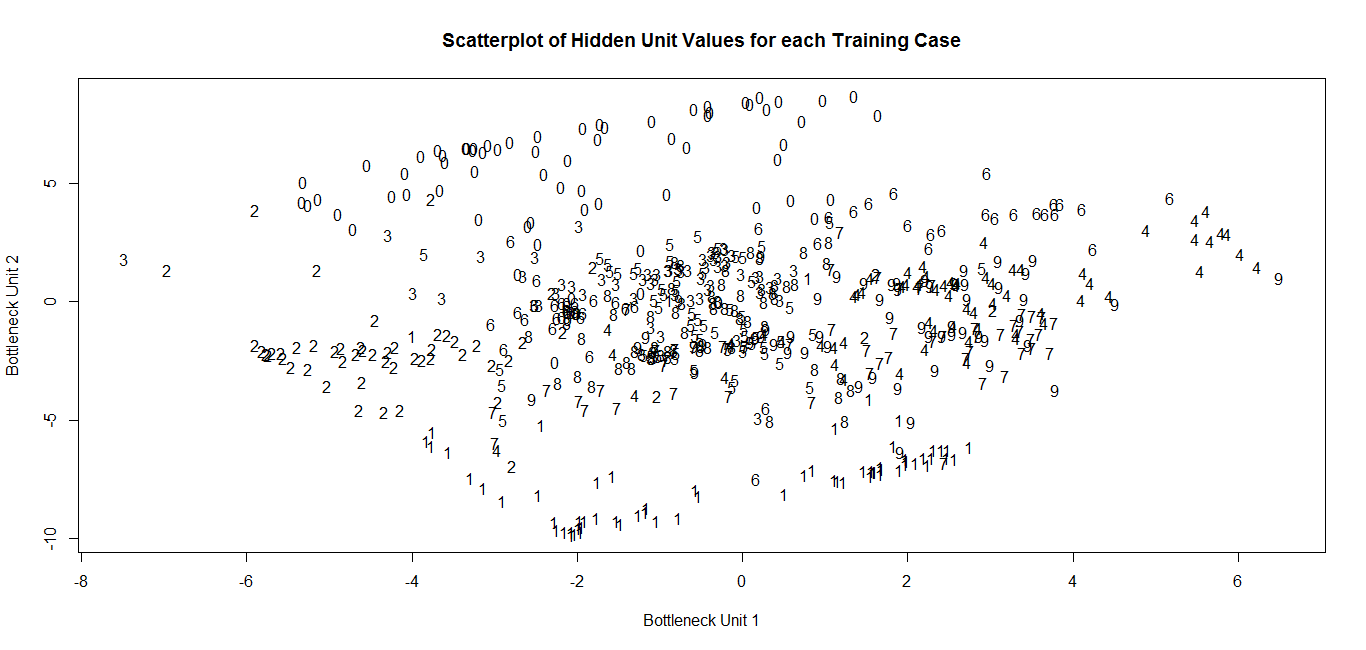
**With 1500 iterations and eta = 0.001**



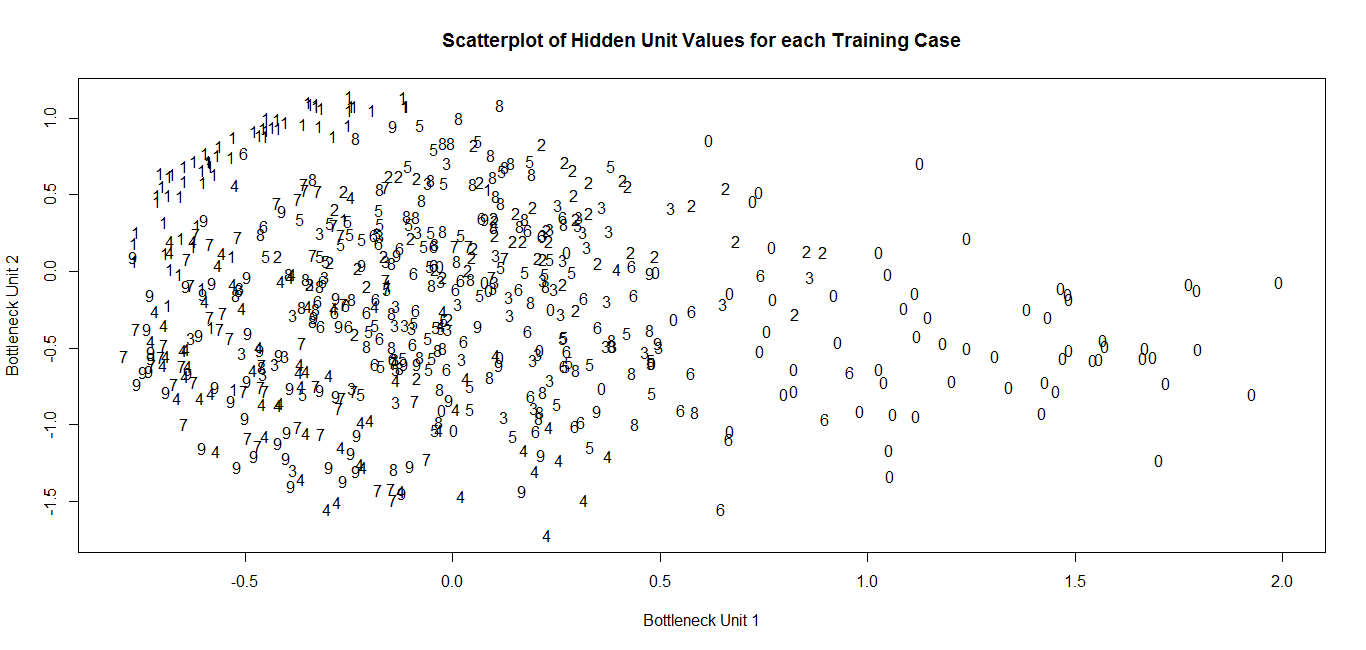


**With 750 iterations and eta = 0.001**

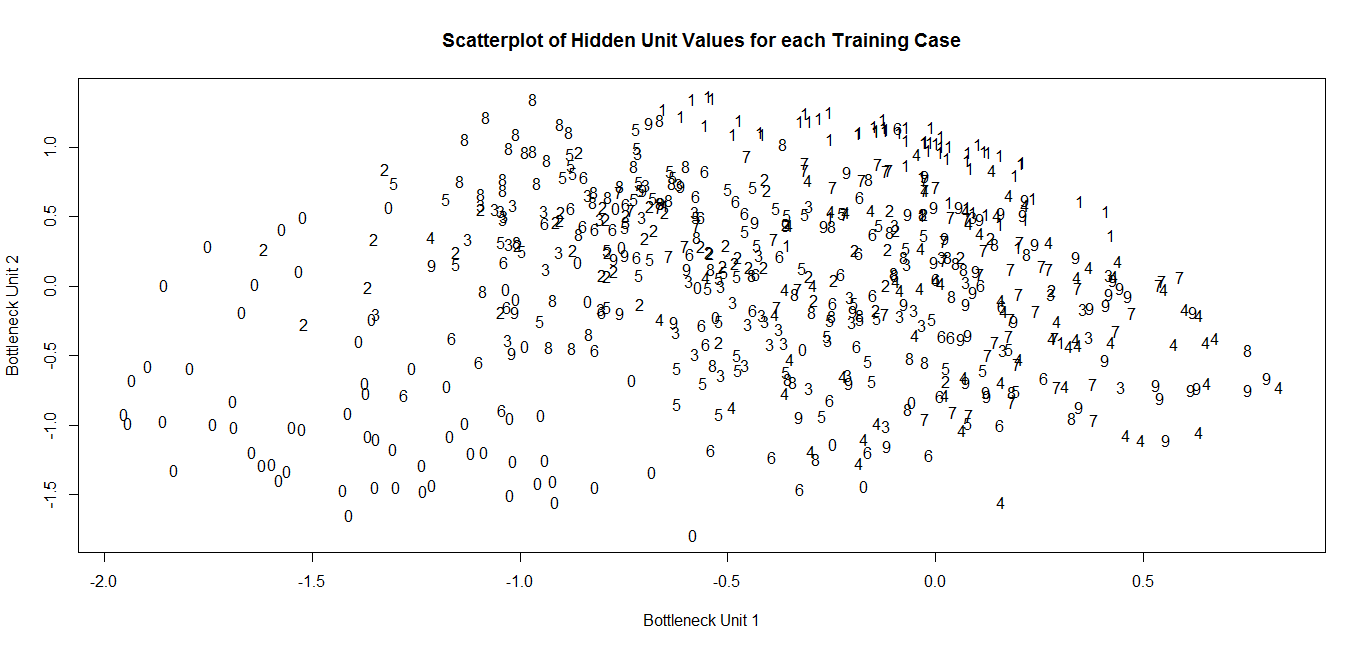
 This is when we attempt to do early stopping for data set 3 with 750 iterations, it seems that the clusters are harder to define than when we started with 1500 iterations. With 1500 iterations, we can clearly identify the clusters for the 0, 1, 2, 4, and 6 digits and to a lesser extent, the 3 and 8 digits as well, showing that the reduction to a two dimensional structure seems to have preserved the information needed for classifying an image as a digit fairly well.  
 If we increase the number of iterations to 2250 while holding eta constant at 0.001, we find that the results are not really any better than when we trained the data set just using 1500 iterations. We might argue that it is in fact even worse, as the 3 and 8 digits have no noticeable clusters at all .   
  
**Using 2250 iterations and eta = 0.001**  


  
Here, we try the original trial with 1500 iterations and an eta of 0.001 using a different random initialization to see whether the results differ or not. Then we try changing the eta value to 0.0001 by a multiple of 10 and correspond by increasing the number of the iterations by a multiple of 10.   
  
**Using 15000 iterations, eta = 0.0001**



  
When we decrease eta to 0.0001, and increase the number of iterations correspondingly to 15000, we can clearly see from the plot of the training error that we are indeed at a local maximum. The curve is also quite smooth and monotonically decreasing, and does not have the spikes/wiggles that appeared when we set eta to be 0.001, and so we obtain considerably better results for our MLP training when we decrease our eta values. However, we also see that with the exception of the 0 and 1 digits, all other clusters are mixed together quite well, and so it is hard to distinguish the digit clusters from one another. Despite good results for the plot of the training errors, we may find it prudent to do early stopping at around 5000 iterations and see what the results might be like. In the two graphs below, we see once again that the clusters for each digit are not too well defined either, with the exception of the 0, 1 and 8 digits.

**Using 5000 iterations, eta = 0.0001**

  
It seems the reduction to a two dimensional structure seems to have preserved the information better when we were dealing with 1500 iterations as opposed to any other number of iterations chosen, but the problem lies in the fact that the squared error plot was not as monotonic decreasing as we would have liked. We run one final trial with 800 iterations and an eta value of 0.0001 as opposed to the original 0.001, to see whether a lower number value of eta would give us a smooth curve for the training error plot, and whether or not a lower number of iterations (Earlier stopping) might give us better results in terms of how the digits cluster with one another. We find that this does almost just as well, with the 0, 1, 2, 3, and 4 digit clusters well defined, and so the information preserved is larger than in most of the other trials.

**Using 800 iterations, eta = 0.0001**

