# Tutorial 6 – Ioannis Manousaridis

In all experiments the learning rate is 0.001.

1. **MnistMLP – Experiment**

| No | Activation function | Optimizers | No of neurons in hidden layer | Without PCA | With PCA |
|---|---|---|---|---|---|
| 1 | Tanh | Sgd | 40 | 0.885 | 0.883 |
| 2 | Tanh | adam | 40 | 0.883 | 0.923 |
| 3 | Relu | sgd | 40 | 0.348 | 0.891 |
| 4 | Relu | adam | 40 | 0.811 | 0.932 |
| 5 | Tanh | sgd | 30 | 0.885 | 0.875 |
| 6 | Tanh | adam | 30 | 0.861 | 0.915 |
| 7 | Relu | sgd | 30 | 0.438 | 0.888 |
| 8 | Relu | adam | 30 | 0.836 | 0.937 |
| 9 | Tanh | sgd | 20 | 0.875 | 0.877 |
| 10 | Tanh | adam | 20 | 0.826 | 0.902 |
| 11 | Relu | sgd | 20 | 0.281 | 0.877 |
| 12 | Relu | adam | 20 | 0.652 | 0.925 |
| 13 | Tanh | sgd | 10 | 0.846 | 0.851 |
| 14 | Tanh | adam | 10 | 0.847 | 0.881 |
| 15 | Relu | sgd | 10 | 0.121 | 0.835 |
| 16 | Relu | adam | 10 | 0.416 | 0.892 |

**Conclusions:**

For the MNIST dataset the results with the PCA were all over 87% and are considered as excellent. Without the PCA some results were not very good. The TANH activation function provide stable and good results in this case but the RELU gave unstable and bad results. Also, the Adam optimizer provided better results than the SGD almost in every case. Finally, the increasement of number of neurons did not provide the expected improvement in the performance. In some occasions, the performance was better when a smaller number of neurons was used.

## 2. FMnistMLP – Experiments

| No | Activation function | Optimizers | No of neurons in hidden layer | Without PCA | With PCA |
|---|---|---|---|---|---|
| 1 | Tanh | sgd | 40 | 0.797 | 0.808 |
| 2 | Tanh | adam | 40 | 0.729 | 0.835 |
| 3 | Relu | sgd | 40 | 0.355 | 0.819 |
| 4 | Relu | adam | 40 | 0.785 | 0.84 |
| 5 | Tanh | sgd | 30 | 0.744 | 0.817 |
| 6 | Tanh | adam | 30 | 0.717 | 0.826 |
| 7 | Relu | sgd | 30 | 0.39 | 0.815 |
| 8 | Relu | adam | 30 | 0.697 | 0.833 |
| 9 | Tanh | sgd | 20 | 0.695 | 0.815 |
| 10 | Tanh | adam | 20 | 0.725 | 0.83 |
| 11 | Relu | sgd | 20 | 0.099 | 0.808 |
| 12 | Relu | adam | 20 | 0.44 | 0.833 |
| 13 | Tanh | sgd | 10 | 0.435 | 0.806 |
| 14 | Tanh | adam | 10 | 0.63 | 0.815 |
| 15 | Relu | sgd | 10 | 0.099 | 0.835 |
| 16 | Relu | adam | 10 | 0.098 | 0.815 |

**Conclusions:**

For the FMNIST dataset the results with the PCA were all over 80% which is considered as a good performance. Without the PCA some results were not very good. The TANH activation function provide again more stable and better results comparing to the RELU, which gave unstable and bad results. The TANH had poor performance only when 10 neurons were used. Furthermore, the Adam optimizer provided better results than the SGD almost in every case. Finally, the increasement of number of neurons did not provide the expected improvement in the performance when a lot of neurons were used. The differences in performance when 40 and 30 neurons are used are slightly. However, when 10 or 20 neurons were used the performance in the models without the PCA was extremely poor.