# Hw 2 Report

**Introduction:**

The three deep learning models that were provided were run changing the parameters at each instance and the results were tabulated.

**Observation:**

Baseline:

The baseline approach without dropout was tested under various parameters. The best results with accuracy of 86.54 and loss of 4.14 were obtained by using the number of neurons as six with 4 layers,sigmoid activation function ,number of epochs 300 and batch size of 16. It can be noted that increasing the number of epochs from 300 to 500 and the batch size to 32 did not have much effect on the performance of the neural network

Dropout_hidden:

This approach makes use of dropout. Dropout approach is used to avoid overfitting of the models. It drops neurons with certain probability. This dropping of neurons help in increasing the generality of the network

The best results were accuracy of 84.56 and loss of 5.38.. This results were obtained by using 60 neurons with probability of dropping the neurons at 0.2,number of layers at 3,number of epochs 300 and batch size 32.Slightly better results were obtained by increasing the batch size from 16 to 32.Increasing the probability did not have any adverse effects in this case. However increasing the number of layers seems to have reduced the accuracy quite drastically.

Dropout_visible:

This approach uses dropout in the visible layers and was the best approach of the three. Accuracy of 86.92 and loss of 7.42 were achieved. This is pretty close to the upper accuracy cap. Increasing the number of epochs helped improve the accuracy. Here too increasing the number of layers reduced the accuracy drastically.

In all the cases ‘rlu activation functions in the upper layers and sigmoid activation function in the lowest layer worked best.’tanh’ activation function had very small accuracy and performed poorly.

**Conclusion:**

The experiments were performed using the given code with various changes in the parameter. All the experiments were conducted in the rambo servers making use of the GPU optimizations. All the experiments were quick to complete,with the experiments having large epochs taking longer.

**Table 1 : Results of the experiments**

| | prob | no. of neurons | no. of layers | activation fn | no.of epochs | batch size | | |
|---|---|---|---|---|---|---|---|---|
| baseline | | 60 | 3 | sigmoid | 300 | 16 | 83.61 | 6.24 |
| | | 60 | 3 | sigmoid | 500 | 16 | 83.61 | 5.03 |
| | | 60 | 3 | sigmoid | 500 | 32 | 83.59 | 5.06 |
| | | 60 | 3 | tanh | 500 | 32 | 54.36 | 10.33 |
| | | 60 | 4 | sigmoid | 300 | 16 | 86.54 | 4.14 |
| | | | | | | | | |
| dropout_hidden | 0.2 | 60 | 3 | sigmoid | 300 | 16 | 82.57 | 7.5 |
| | 0.2 | 60 | 3 | sigmoid | 500 | 16 | 81.23 | 7.26 |
| | 0.2 | 60 | 3 | sigmoid | 300 | 32 | 84.56 | 5.38 |
| | 0.2 | 60 | 3 | tanh | 300 | 32 | 32.81 | 21.5 |
| | 0.5 | 60 | 3 | relu | 300 | 32 | 3.81 | 11.43 |
| | 0.5 | 60 | 3 | sigmoid | 300 | 32 | 84.11 | 5.79 |
| | 0.2 | 60 | 4 | sigmoid | 300 | 32 | 53.38 | 1.23 |
| | | | | | | | | |
| dropout_visible | 0.2 | 60 | 3 | sigmoid | 300 | 16 | 83.07 | 6.52 |
| | 0.2 | 60 | 3 | sigmoid | 500 | 16 | 86.92 | 7.42 |
| | 0.2 | 60 | 3 | sigmoid | 1000 | 32 | 85.93 | 6.51 |
| | 0.2 | 60 | 3 | tanh | 500 | 16 | 24.07 | 24.2 |
| | 0.5 | 60 | 3 | sigmoid | 500 | 16 | 84.52 | 4.97 |
| | 0.2 | 60 | 4 | sigmoid | 500 | 16 | 59.24 | 13.21 |