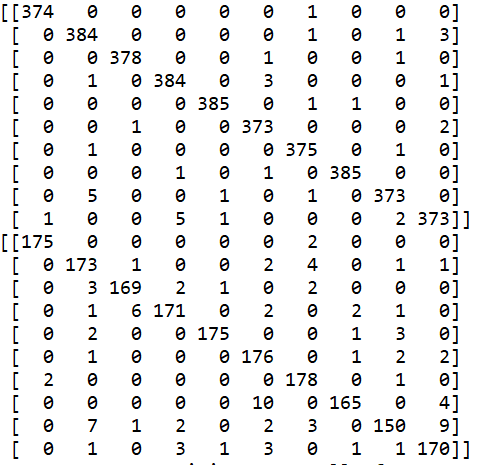
**LAB ASSIGNMENT 2**

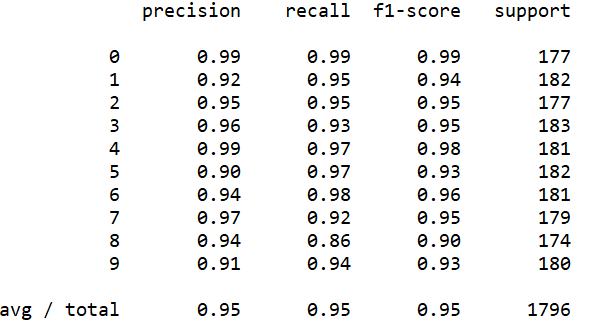
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S.no | Number of hidden layers | Number of units in each hidden layer | Loss function | Activation  Function  (Input) | Optimizer | Learning  Rate | Momentum | Input Scaling | Accuracy | Convergence  speed |
| 1 | 1 | One layer:16 | Cross-  entropy | ReLU | SGD | 0.01 | 0 | Yes | 95.92 | 25.88 |
| 2 | 1 | One layer:16 | Cross-  entropy | Tanh | SGD | 0.01 | 0 | Yes | 93.21 | 29.00 |
| 3 | 1 | One layer:16 | MSE | ReLU | Adam | N/A | N/A | No | 93.48 | 28.91 |
| 4 | 1 | One layer:16 | MSE | Tanh | Adam | N/A | N/A | Yes | 92.81 | 27.37 |
| 5 | 2 | 1st: 16  2nd:16 | Cross-  entropy | ReLU | Adam | N/A | N/A | No | 99.04 | 28.21 |
| 6 | 2 | 1st: 16  2nd:16 | Cross-  entropy | Tanh | SGD | 0.001 | 0.5 | No | 90.003 | 11.95 |
| 7 | 2 | 1st: 32  2nd:16 | MSE | ReLU | Adam | N/A | N/A | Yes | 95.43 | 22.22 |
| 8 | 2 | 1st: 32  2nd:16 | MSE | Tanh | Adamax | 0.002 | N/A | Yes | 95.48 | 31.66 |
| 9 | 2 | 1st: 48  2nd:16 | Cross-entropy | Tanh | Adamax | 0.001 | N/A | No | 98.88 | 27.24 |
| 10 | 2 | 1st: 48  2nd:16 | MSE | ReLU | Adamax | 0.001 | N/A | No | 94.48 | 26.61 |

Experiments are repeated with different hyperparameters such as different activation functions such as tanh and relu; different optimizers, adam, SGD, Adamax and different loss functions such as cross entropy error and mean square error function.

Best Model Learned for Cross-entropy with ReLU as activation unit and Adam optimizer has two hidden layers each with 16 units and has an accuracy of 99.04

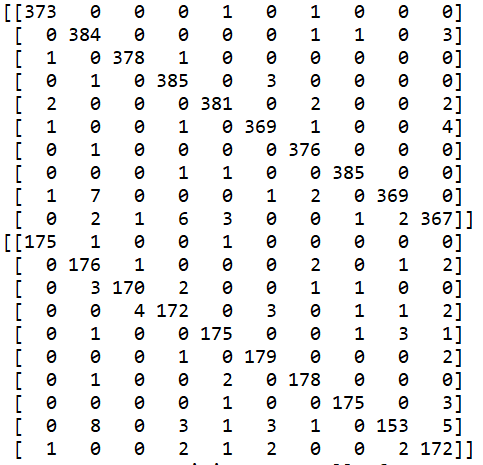
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 2 | 1st: 16  2nd:16 | Cross-  entropy | ReLU | Adam | N/A | N/A | No | 99.04 | 28.21 |

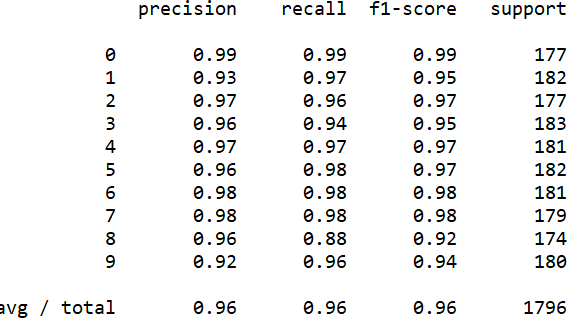




Best Model Learned for Cross-entropy with TanH as activation unit and Adamax optimizer has two hidden layers one with 48 units and the other with 16 units and has an accuracy of 98.88

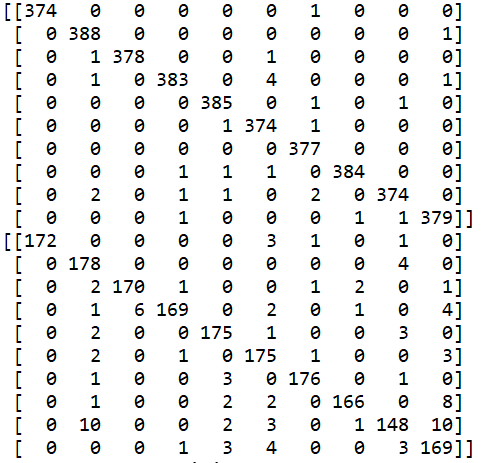
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9 | 2 | 1st: 48  2nd:16 | Cross-entropy | Tanh | Adamax | 0.001 | N/A | No | 98.88 | 27.24 |

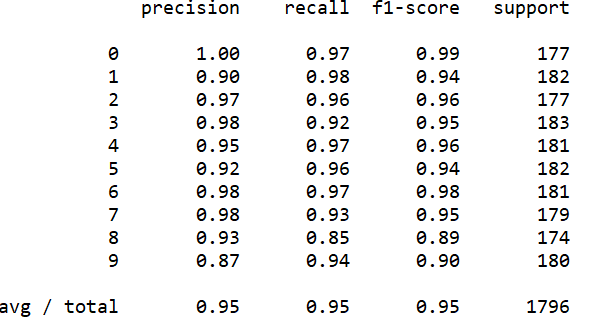




Best Model Learned for MSE with ReLU as activation unit and Adam optimizer has two hidden layers one with 32 units and the other with 16 units and has an accuracy of 95.43

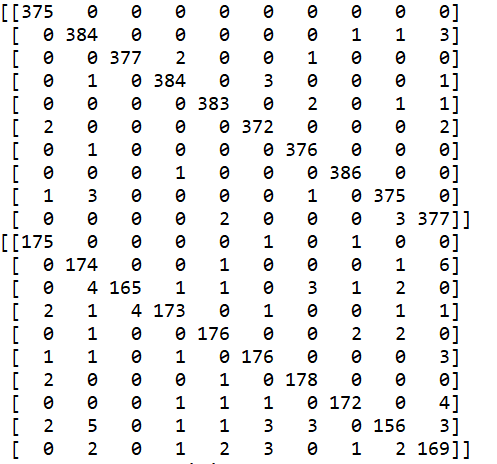
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 2 | 1st: 32  2nd:16 | MSE | ReLU | Adam | N/A | N/A | Yes | 95.43 | 22.22 |

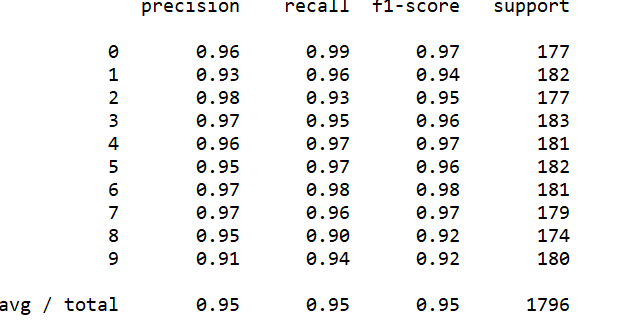




Best Model Learned for MSE with TanH as activation unit and Adamax optimizer has two hidden layers one with 32 units and the other with 16 units and has an accuracy of 95.48

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 2 | 1st: 32  2nd:16 | MSE | Tanh | Adamax | 0.002 | N/A | Yes | 95.48 | 31.66 |





**Conclusion for Results 1:** Cross-entropy performs better than Mean-squared error function using any activation units because it got an accuracy with 98% compared to 95% of MSE.

ReLU performs slightly better than TanH, but there are possibilities in which tanh can perform better based on parameters such as increasing hidden layers or hidden units or learning rate etc.

ReLU is non-saturation of its gradient, which accelerates the convergence speed while using optimizers compared to the tanh functions. Also, tanh neurons involve expensive operations such as exponentials, while the ReLU can be implemented by simply thresholding a matrix of activations at zero.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| S.no | Number of filters | Filter size | Dropout  probability | Dropout  probability | Accuracy | Convergence  Speed |
| 1 | 32 | 3x3 | 0.3 | 0.5 | 97.93 | 44.12 |
| 2 | 32 | 3x3 | 0.1 | 0.9 | 96.88 | 51.38 |
| 3 | 32 | 2x2 | 0.4 | 0.6 | 97.60 | 38.40 |
| 4 | 32 | 2x2 | 0.2 | 0.9 | 96.82 | 45.38 |
| 5 | 32 | 4x4 | 0.3 | 0.8 | 96.83 | 36.90 |
| 6 | 32 | 5x5 | 0.5 | 0.5 | 96.71 | 33.72 |
| 7 | 32 | 6x6 | 0.8 | 0.9 | 95.37 | 43.39 |

**Conclusion for results for question 2:**

Best model learned is CNN with number of filters as 32 for first half and 32 for second half of image in Convolution 2D and filter window size as 3x3 matrix with dropout probabilities of 0.3 and 0.5 has an accuracy of 97.93

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 32 | 3x3 | 0.3 | 0.5 | 97.93 | 44.12 |

