**Experiment 1 :**

Flatten input layer ;

First hidden layer; **128** densely connected neuron with **rectified linear function** as activation function

Second hidden layer ; **128** densely connected neuron with **rectified linear function** as activation function

Output layer; 4 neurons with **softmax** function as activation function .

|  |  |  |
| --- | --- | --- |
| Epoch=50 | Epoch =100 | Epoch = 200 |
| loss(train)=0.0497 | Loss(train) = 0.0567 | loss(train)=0.012 |
| accuracy(train)= 0.9788 | Accuracy(train)=0.98166 | accuracy(train) = 0.996 |
| RMSE(test)=0.18836 | RMSE(test) = 0.18039095 | **Rmse(test)=0.1691** |

**Experiment 2 :**

Flatten input layer;

First hidden layer; **64** densely connected neuron with rectified linear function as activation function

Second hidden layer ; **64** densely connected neuron with rectified linear function as activation function

Output layer; 4 neurons with **softmax** function as activation function .

|  |  |  |
| --- | --- | --- |
| Epoch=50 | Epoch =100 | Epoch = 200 |
| loss(train)= 0.00004 | Loss(train) = 0.00004 | loss(train)= 0.00003 |
| accuracy(train)= 0.999999999 | Accuracy(train)= 0.9999999 | accuracy(train) = 0.999999 |
| RMSE(test)= 0.1971 | RMSE(test) = 0.196826 | Rmse(test)= 0.19545865 |

**Experiment 3 :**

Flatten input layer ;

First hidden layer; **64** densely connected neuron with rectified linear function as activation function

Second hidden layer ; **64** densely connected neuron with rectified linear function as activation function

Output layer; 4 neurons with **sigmoid** function as activation function .

|  |  |  |
| --- | --- | --- |
| Epoch=50 | Epoch =100 | Epoch = 200 |
| loss(train)= 0.08972 | Loss(train) = 0.0053 | loss(train)= 0.000001 |
| accuracy(train)= 0.97490346 | Accuracy(train)= 0.999 | accuracy(train) = 0.99999 |
| RMSE(test)= 0.1894 | RMSE(test) = 0.1839 | Rmse (test)= 0.2718 |

**Findings :**

* The best network among the three experiments is the one with two hidden layers of 128 neurons activated by the rectified linear function and 4 output neuron activated by softmax function . It is because the root mean square error for completely unseen data is the lowest for this network structure
* The lesser the number of neurons in the hidden layer used to train , the more the network tends to over fit .
* For multiclass classification problem , softmax function as activation function for the output neuron works better .

.