**THIS IS MY WORK. I HAVE NOT COPIED FROM ANYONE ELSE**

**ANIMESH SHARMA: 19901680**

3 different models were tried and run using keras packages on the MNIST data and models were assessed based on their accuracy and other factors.

**MODEL 1:**

For model 1, a very basic model was chosen with 2 convolutional layers, both with the 14 filters of dimensions 3x3. Maximum pooling was a part of the network along with flattening to have a fully connected layer were also included in the model and there are 512 hidden neurons in the network. Looking at the MNIST data, it was found that there wouldn’t be a lot of scope for overfitting so dropout was not included. Activation functions relu and softmax were used.

With this model three scenarios were tested. In one of the scenarios, the learning rate was kept very low at the value of 0.002 and it was run for 10 epochs with 60 as batch size.

Since the batch size and learning rate were kept low, the model did take some time to show results. It was found that both, training error and validation error had very small values and converged. Hence, there was no evidence of overfitting. The accuracy came out to be 99.6% which is close to perfect.

For the second scenario, the learning rate was increased to 0.02 from 0.002. The batch size was left unchanged. The error terms were larger than scenario 1 however this was not an entirely bad result because there was still very less evidence for overfitting. The accuracy reduced from almost perfect to 82%. Even in this scenario the neural network performed fairly well.

For the third scenario, the learning rate was set at 0.02 and the batch size was increased to 600 from 60. This time the error values turned out to be higher than scenario 1. The accuracy reduced from 82% to 81.78%. The neural network took a lot less time to show results and execute as compared to scenario 1 and scenario 2 which can be attributed to the 10-fold increase in batch size. Since there was no significant change in results due to increase in batch size, it is better to have a fairly large batch size to reduce time consumption and increase efficiency.

It can be concluded that clearly, the learning rate played a major role in determining the accuracy of the neural network. The batch size reduced the time taken by the network to show results.

**MODEL 2:**

For model 2, 2 additional convolutional layers were included with 28 filters of 3x3 dimensions with all other attributes of the network unchanged.

With a learning rate of 0.002 and batch size of 60, the neural network performed extremely well with the accuracy of 99.7% with very low training error and validation error. There was no evidence of overfitting.

Unlike model 1, this model showed high accuracy of 99.3% even when the learning rate was increased from 0.002 to 0.02. Even the errors, that saw a sudden increase in model 1 didn’t seem to change to such a great extent in model 2. The errors remained low.

These results prove that it is beneficial to add more layers to the neural network to get better results and low errors.

**MODEL 3:**

For model 3 the activation function of convolution layers was changed from relu to sigmoid. There was no addition of layers from model 1.

Although the model did end up performing very well when learning rate was set to 0.002, it didn’t perform as well as model 1. Accuracy reduced from 99.6% to 90% which still suggests a very good performance.

However, there was a contrast in performance from model 1 when the learning rate was increased to 0.02. In model 1, there was a decrease in accuracy when the learning rate was increased. In model 3, the accuracy increased to 96% from 90% when the learning rate was increased to 0.02.

From the evidence gained, it can be concluded that with sigmoid as the activation function, a higher learning rate gives better results i.e higher the learning rate, the better.

**RANKING: Based on the results obtained, it can be concluded that model 2, which had the highest number of layers showed the best results. Model 1 and model 3 showed almost the same results and showed the characteristics of relu and sigmoid as activation functions.**