ANALYSIS:

I tried many kinds of hyperparameters while training the neural
network for this situation. The best accuracy i could reach up to
was 97.58 percent and this was obtained with a 4 layer(2 hidden)
neural network with ReLU activation function and Adam optimizer with
a learning rate of 0.0001. All the models i trained were with 30
epochs and 50 batch size.

First, lets see the changes with no of change of layers in the
neural network:
The accuracy remained almost same for 1,2 and 3 hidden layers with
values of 97.32,97.58 and 97.58 repectively. For no hidden layer
too, the network was pretty accurate with 96 percent accuracy, but
when u make it 0 layers, i.e basically a regression, the accuracy
comes down to around 51 percent. This is because regression requires
a lot more training time and isnt very efficient as neural networks.

Next, I tried using different optimizers to see how the result
differs, here using SGD gave around 95 percent accuracy and AdaGrad
gave around 89 percent, both lesser than Adam which gave a whooping
97.32 percent accuracy.

Now, lets see how the accuracy changed with the change in activation
function. ReLU function was more accurate with 97.32 compared to 94
percent accuracy in sigmoid function. This is why ReLU is more
widely used than sigmoid nowadays.

Lastly, i tried altering the learning rate and seeing how the result
changed. Learning rates of 0.001,0.0001 and 0.00001 gave pretty good
accuracies of 96.32,97.32 and 96.0 respectively. As i lessened the
learning rate more to 0.000001, accuracy dropped to 89 and when i
increased it to 0.01 or more, it dropped drastically to 9.8 percent
due to overfitting.

Graph of loss vs no of epochs for sigmoid activation function
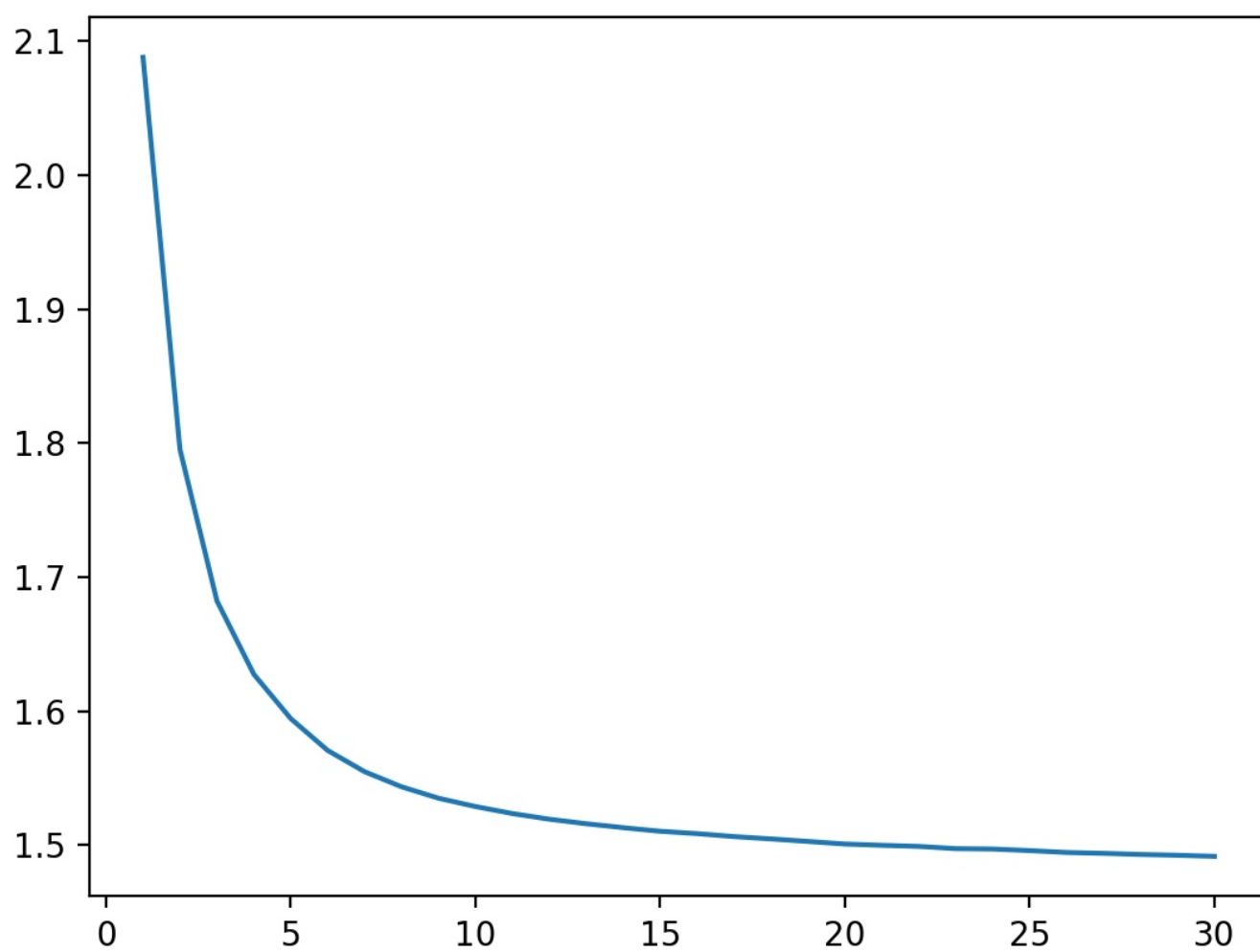
Graph of loss vs no of epochs for ReLU activation function,
Note how it decays faster then the one with sigmoid.