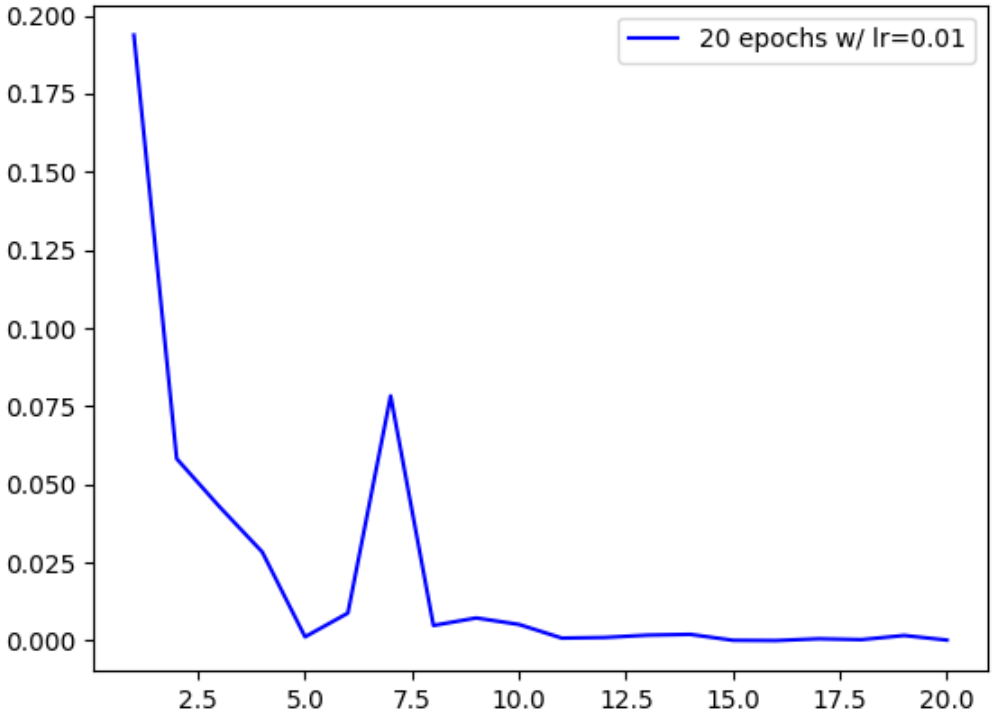
Model Explaination

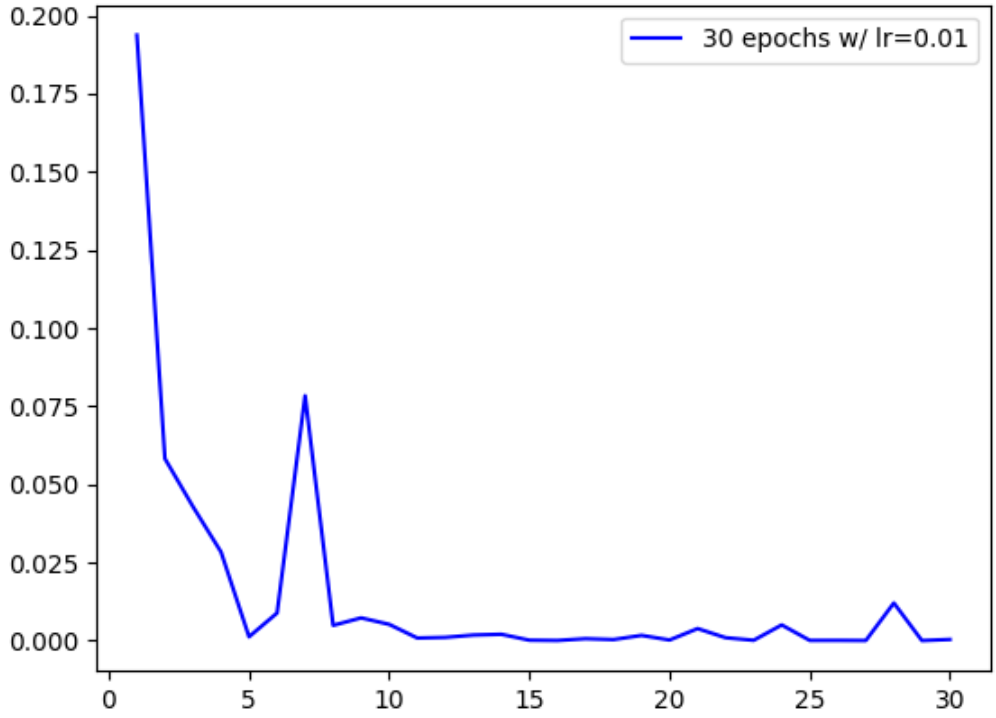
Aniket Pokle

July 30, 2021

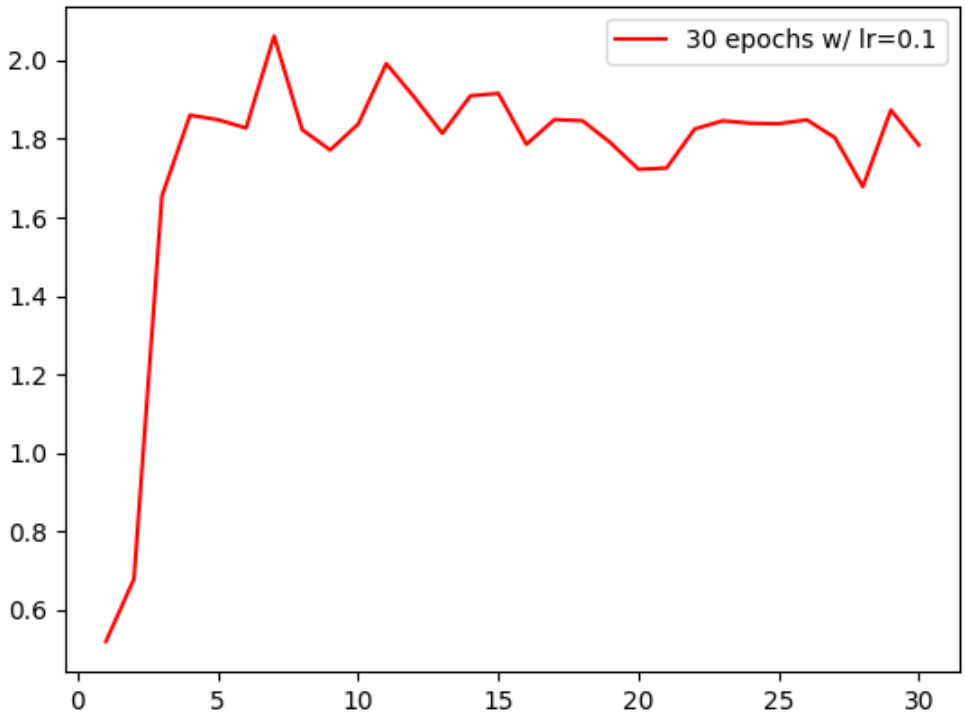
The model used for training is a deep learning model with with a input layer taking in inputs of length 784 (the pixels of a single image). The model has 2 hidden layers both having 64 hidden neurons and ReLu activation function. The output layer has 10 neurons to predict the probability of each number.



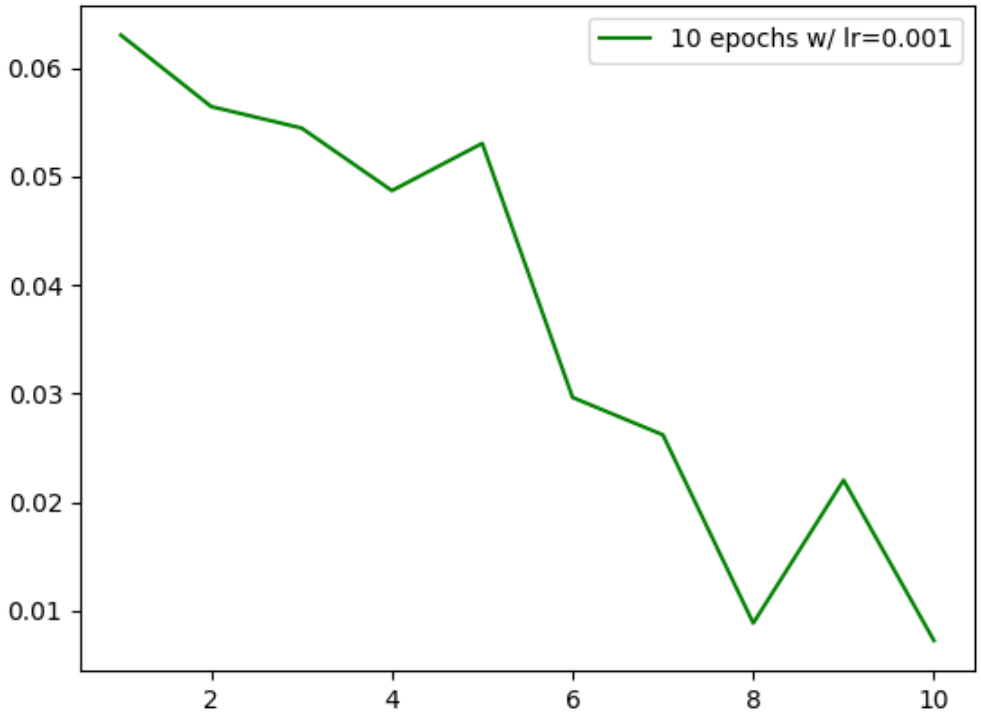
Increasing the number of epochs with the same learning rate just makes the model overfit the training data and the overall accuracy of the model on testing data decreases from 96.63% to 96.29%.



Experimenting further with the hyperparameters of the model and changing the learning rate to 0.1 from the earlier learning rate of 0.01 keeping the no. of epochs 30, takes a blow on the prediction of the model, now the model producing only 17.13% accuracy on testing data.
As the learning rate is increased 10 times the original the model tries to learn the data faster taking major blows on its accuracy and hence increasing its loss.
Here is the figure on how the loss varies

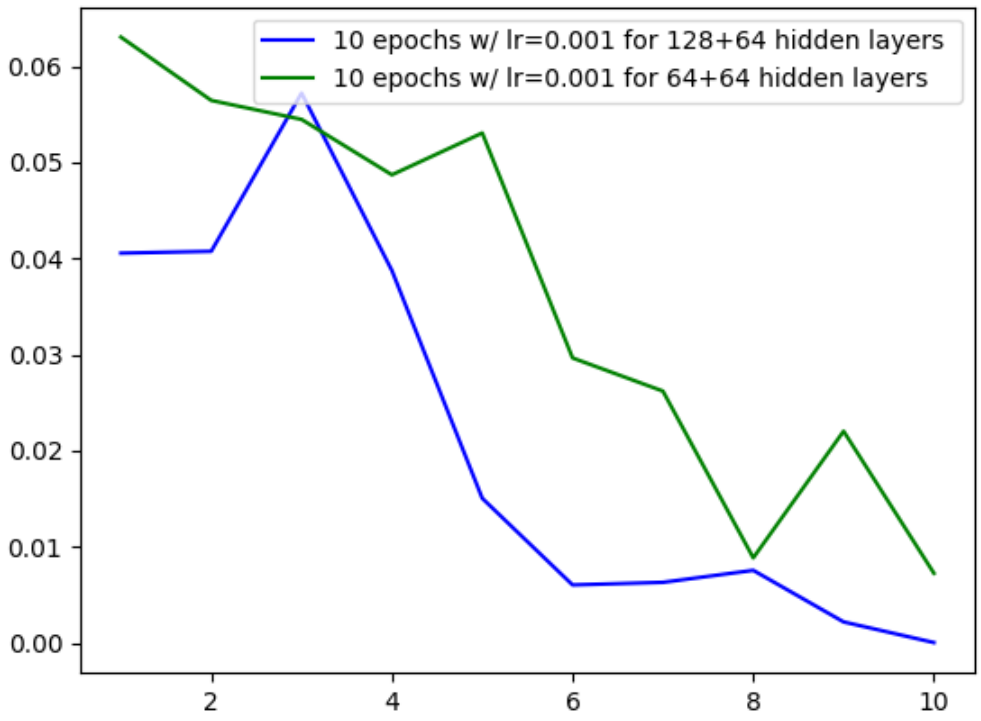


Now, inorder to improve the accuracy of the model I further changed the number neurons in the hidden layers of the model from 64+64 to 128+64 and keeping the number of epochs 10 and learning rate decresed 10 times from the original model to 0.001.
This minor improvements in the number of neurons and learning rate increased the accuracy of the model from 96.63% to 97.89%.



We can see how the losses in the model drop drastically and further the steep drop in the losses suggests that increasing the number of epochs will saturate the losses of the model.

Here is the comparison of the 2 models with 128+64 neurons in the hidden layer and 64+64 neurons in the hidden layer of the other model. I have kept the other hyperparameters of both the models same i.e learning rate is 0.001 and number of epochs are 10.



We can see that the losses in the first model decreases faster than the losses in the second model.