HOMEWORK 2

TEAM: ARATRIK CHANDRA 23M0786
SAYANTAN BISWAS 23M0806

TITLE : ANALYZING THE EFFECTS OF HYPERPARAMETERS ON LOSS AND ACCURACY FOR FEED FORWARD NEURAL NETWORKS

SECTION 1: SIMPLE DATASET

METHODOLOGY:

We are training a feed forward neural network on the **"simple"** dataset. The feed forward neural network has ***1 hidden layer*** with ***20 neurons*** *in it*. We have used **Rectified Linear Unit** (ReLU) as the activation function of all layers. The optimizer used for training of the neural network is **Adam** Optimizer. The main objective of this is to train this neural network and analyze the loss and accuracy of the model with changes in hyperparameters like Learning Rate and Number of Epochs.

DATA PREPROCESSING: Not Done

RESULTS:

LEARNING RATE: The Learning Rate determines the rate of learning by limiting the amount of change of weights on each parameter such that the loss function decreases neither slowly or rapidly avoiding Exploding Gradient and Vanishing Gradient Problem. The Learning Rates we have used in our experiments are 0.001, 0.00033 and 0.0001.

Sl. No.	Learning Rate	Iterations	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy
1.	0.001	50	0.0773	0.1397	1.0	0.98
2.	**0.00033**	**100**	**0.1077**	**0.1437**	**1.0**	**0.98**
3.	0.0001	150	0.3224	0.3134	0.9375	0.95

For learning Rate 0.001, as the learning rate is very high the plot for the training accuracy and training graph is fluctuating a lot and the validation loss is increasing after a certain

number of epochs. For Learning Rate 0.0001, as the learning rate is too small, the training is too slow and we get almost linear graphs for validation loss and training loss. Therefore, we choose the learning Rate as **0.00033**.

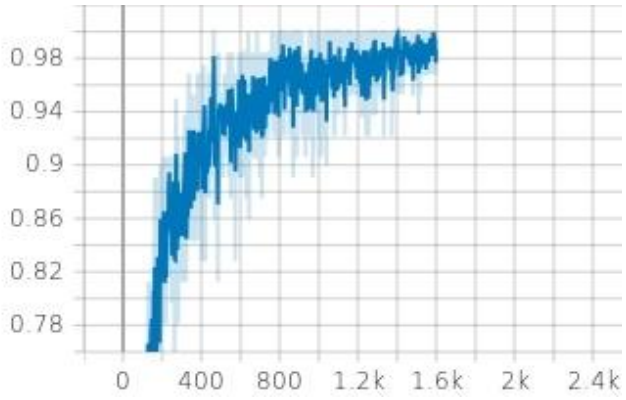
NUMBER OF EPOCHS: The number of Epochs determines the number of iterations i.e. how many times the weights on the corresponding parameters are updated in order to find the best fit model. The number of epochs must not be too small to result in poor training of the model and also very little change in accuracy of a very large number of epochs after a satisfactory accuracy is achieved can be ignored. In this report, for the best learning rate found, we have experimented with 75, 100 and 125 epochs for this models in order to get the best epoch.

Sl. No.	Best Learning Rate	Number of Epochs	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy
4.	**0.00033**	**100**	**0.1077**	**0.1437**	**1.0**	**0.98**
5.		75	0.1705	0.1942	0.9063	0.97
6.		125	0.09876	0.1454	1.0	0.98

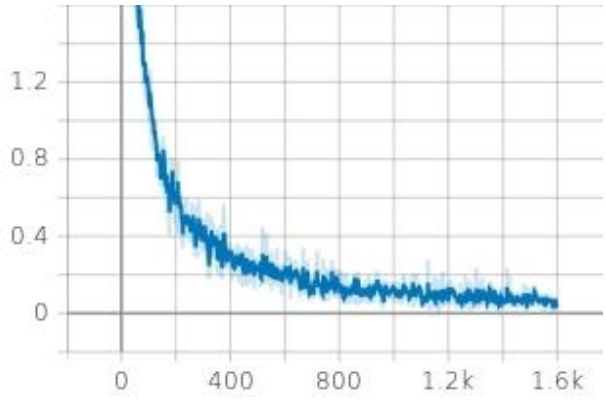
If the number of epochs is taken 75, we can see that the model is not completely trained. But the validation accuracy remains constant after 100 epochs and we see a very minute change in validation loss over some number of epochs. Therefore, to avoid overfitting, we have taken the number of epochs to be **100**.

OVERFITTING:

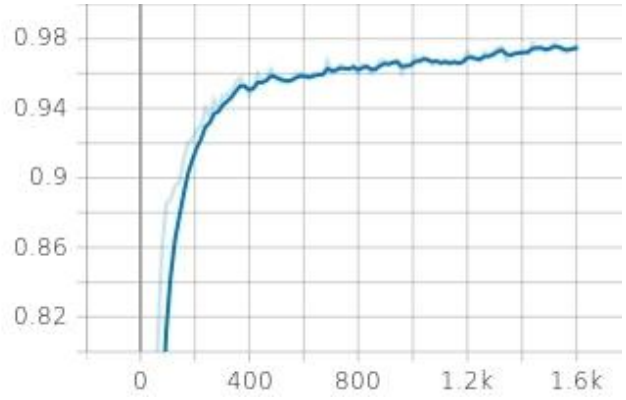
If the model is trained for too many epochs, it may start to memorize the training data rather than learning the underlying patterns. Thus, it leads to overfitting and to control it we use **Early Stopping**. Therefore, rather than going for 125 epochs, we cut off the training at 100 epochs.



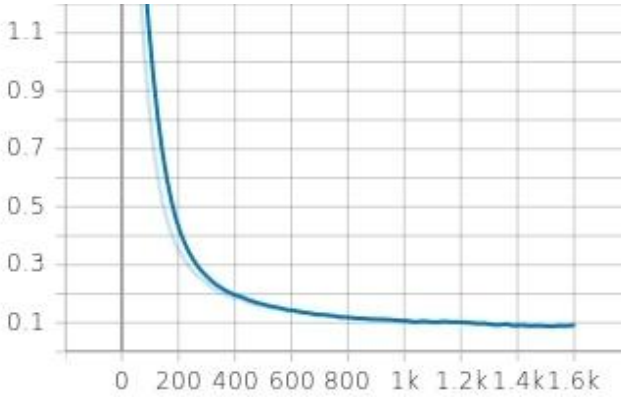
a)Training Accuracy



b)Training Loss



c)Validation Accuracy



d)Validation Loss

Fig: Accuracy and Loss wrt Training and Validation for Learning Rate=10^{-1} and Epochs = 50

CONCLUSION:

The main objective of this report is to bring forward how tuning the hyperparameters results in the change in loss and accuracy for a Feed Forward Neural Network. Along with this, the aim is to find the best fit hyperparameters for this model on the given dataset so that the model gives the best prediction on the validation dataset.
On performing this experiment, we have obtained that the trained Feed Forward Neural Network on the “simple” dataset is performing most efficiently for **Learning Rate=0.00033** and **Number of Epochs= 100.**
It gives astonishing results of a **Validation Accuracy = 98%** and **Validation Loss = 0.1437**

SECTION 2: DIGITS DATASET

METHODOLOGY:

We are training a feed forward neural network on the **"simple"** dataset. The feed forward neural network has *3 hidden layers* with *256 neurons in the first, 128 neurons in the second layer and 64 neurons in the third.* We have used **Rectified Linear Unit** (ReLU) as the activation function of all layers. The optimizer used for training of the neural network is **Adam** Optimizer. We have also used a **Dropout of 0.5** in all the hidden layers as it is a common practice for removing over fitting when the model is too complex. The main objective of this is to train this neural network and analyze the loss and accuracy of the model with changes in hyperparameters like Learning Rate and Number of Epochs.

DATA PREPROCESSING: Not Done

RESULTS:

LEARNING RATE: The Learning Rate determines the rate of learning by limiting the amount of change of weights on each parameter such that the loss function decreases neither slowly or rapidly avoiding Exploding Gradient and Vanishing Gradient Problem. The Learning Rates we have used in our experiment are 0.001, 0.0003 and 0.0001.

Sl. No.	Learning Rate	Epoch	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy
1.	0.001	75	0.02662	0.101	1.0	0.9825
2.	**0.0003**	**125**	**0.03563**	**0.07414**	**0.975**	**0.985**
3.	0.0001	200	0.1015	0.1147	0.975	0.975

learning rate controls the step size during gradient descent. For Learning Rate= 0.001,the model sometime converges rapidly and sometime overshoots the local minima.As a result the training and validation accuracy graph is fluctuating massively. For learning rate 0.0001, The training is very slow.Even after 200 epochs the model is not trained fully.So, We have observed that Validation loss in minimum and Validation accuracy is maximum for Learning Rate=**0.0003**.

NUMBER OF EPOCHS: The number of Epochs determines the number of iterations i.e. how many times the weights on the corresponding parameters are updated in order to find

the best fit model. The number of epochs must not be too small to result in poor training of the model and also very little change in accuracy of a very large number of epochs after a satisfactory accuracy is achieved can be ignored. In this report, for the best learning rate found, we have experimented with 10, 50, 100 and 200 epochs for both the models in order to get the best epoch.

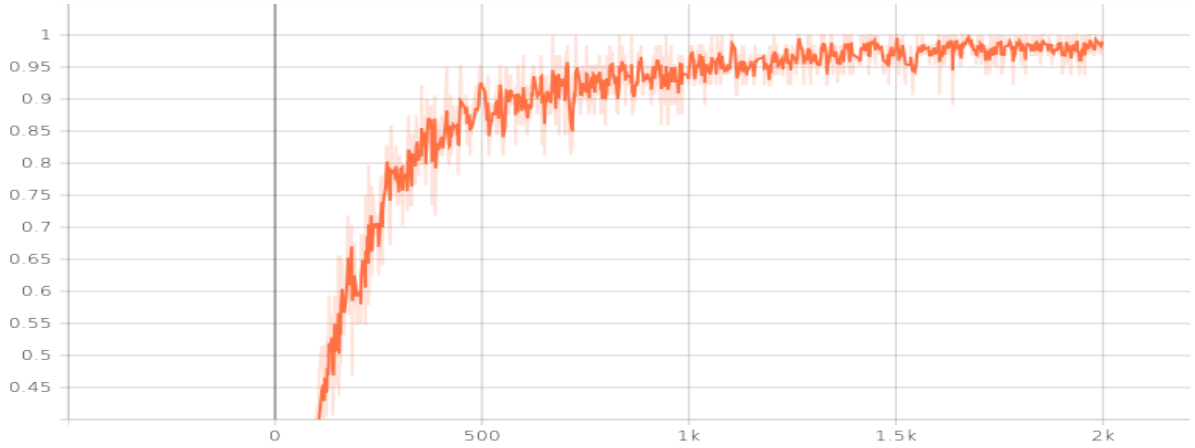
Sl. No.	Best Learning Rate	Number of Epochs	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy
4.	**0.0003**	100	0.06457	0.08222	0.975	0.98
5.		**125**	**0.03563**	**0.07414**	**0.975**	**0.985**
6.		150	0.01334	0.07414	0.975	0.985

We have observed that the validation loss and validation accuracy does not increase with increasing the number of epochs after 125. We have rejected the number of epochs=150, to avoid overfitting.

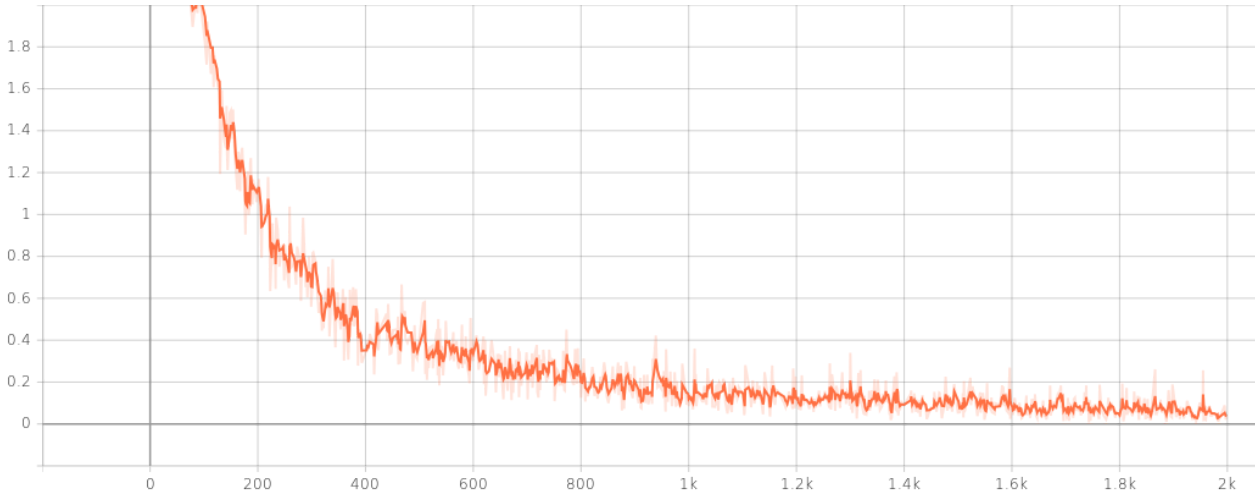
REGULARIZATION:

Regularizer	**Training Loss**	**Validation Loss**	**Training Accuracy**	**Validation Accuracy**
Dropout	**0.03563**	**0.07414**	**0.975**	**0.985**
Without Dropout	2.01e-4	0.1118	1.0	0.9725

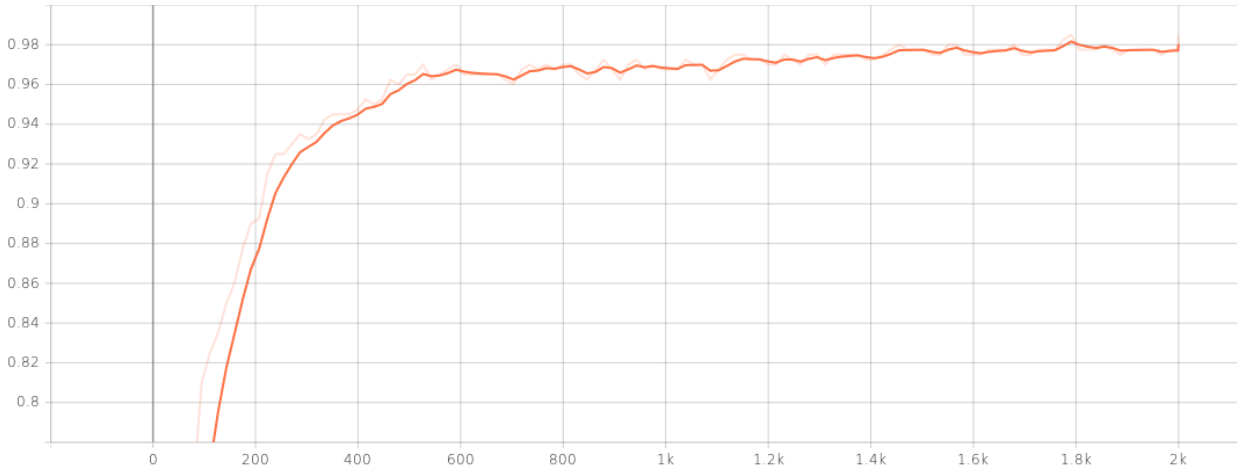
Without dropout, the validation loss is increasing after 600 steps and the validation accuracy remains constant after 1000 steps. Therefore, we use dropout to remove the overfitting from the model.



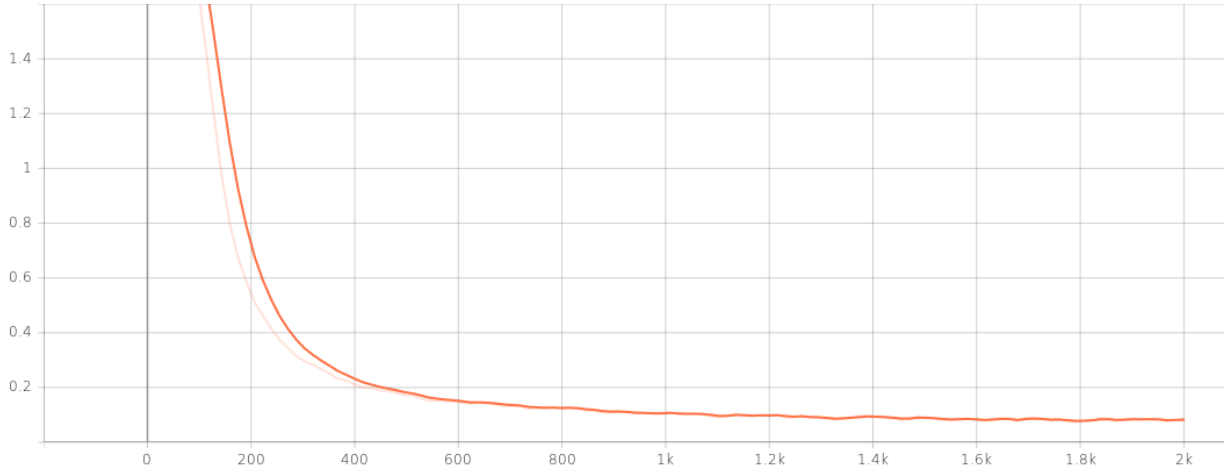
a)Training Accuracy



b)Training Loss



c)Validation Accuracy



d)Validation Loss

Fig: Accuracy and Loss wrt Training and Validation for Learning Rate=0.0003, Epochs = 100 and Dropout

CONCLUSION:

The main objective of this report is to bring forward how tuning the hyperparameters results in the change in loss and accuracy for a Feed Forward Neural Network. Along with this, the aim is to find the best fit hyperparameters for this model on the given dataset so that the model gives the best prediction on the validation dataset.

We have built a Feed Forward Neural Network, with 3 hidden layers having 256 neurons in the first layer, 128 neurons in the second layer and 64 neurons in the third layer. We have also used a Dropout of 0.5 in all the layers.

On performing this experiment, we have obtained that the trained Feed Forward Neural Network on the “digits” dataset is performing most efficiently for **Learning Rate=0.0003** and **Number of Epochs= 125.**

It gives astonishing results of a **Validation Accuracy = 98.5%** and **Validation Loss = 0.07414.**