Name: RAGHVENDRA PAL

College/graduating year/GPA: GOVERNMENT ENGINEERING COLLEGE (JABALPUR, M.P.)

Phone number: 8319903602

Email ID: raghvendrapal65@gmail.com

1. To Run the assignment, user need to run **Weight_and_Neuron_Pruning_the_network.ipynb** file**.**

CONCLUSION OF ASSIGNMENT :

2.Why pruning Important?

Pruning is a way to reduce the evaluation time to predict the result of the data. Pruning optimizes the actual model and make resultant model which takes less storage in disk and also takes less time for prediction.

In the mobility world, mobile devices can not able to store large amount of data like some model gives 500MB to 1GB trained model and we need to store them in out mobile disk for further predicting result on future data. Like Face detection model could have 1GB of data but we can not use it store on mobile thats why pruning comes. Pruning uses weight and neuron pruning to reduce the time and space complexity of model.

I had not known about this technology and i am very happy to be part of this assignment. Through this assignment i got to know that through pruning we can optimize our model.
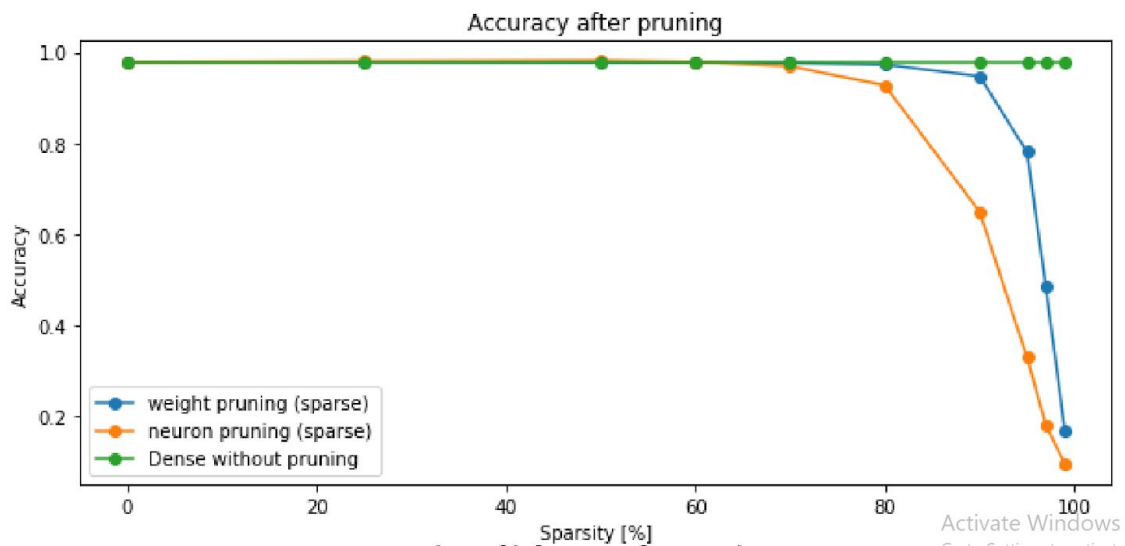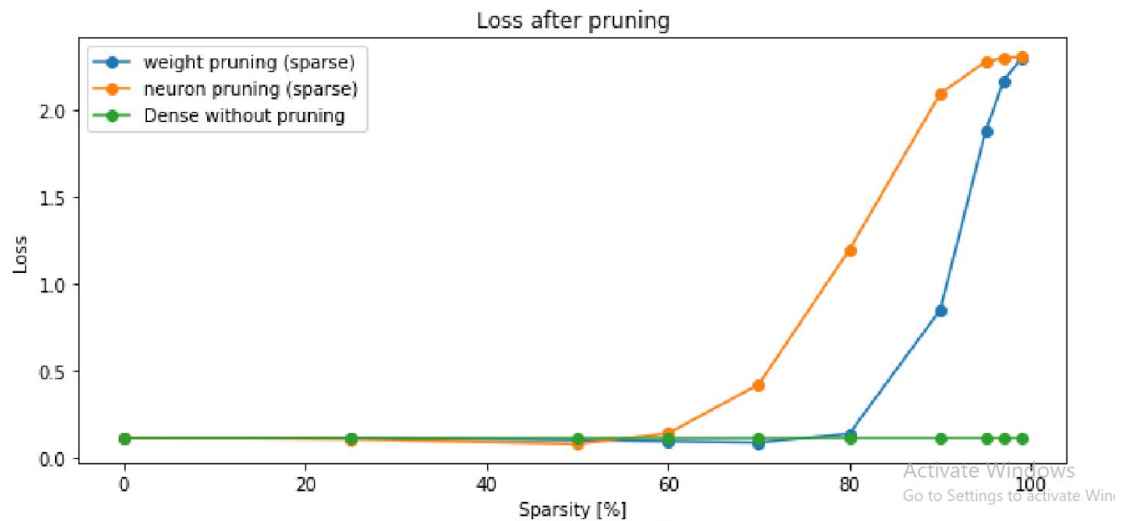
3. Why do you think that is/isn't? Do you have any hypotheses as to why we are able to remove so much of the network without hurting performance (this is an open research question)?

From the graph we can see that for weight pruning model accuracy is going down when sparsity crosses 80% and for neuron pruning model it is reducing after 70% and loss increasing. Accuracy for weight pruning is reducing and loss increasing, because after 80% sparsity we are making important weight to zero, and for neuron pruning we can say imp neuron are deleted.

Before 80% in weight pruning and 70% in neuron pruning, i think in weight pruning we are making those values to zero or removing those neurons in neuron pruning which are not impacting so much in performance of model.