ECGR 5106 – Real Time Machine Learning (Spring 2023)

Nahush D. Tambe – 801060297

Problem 1, Part a:

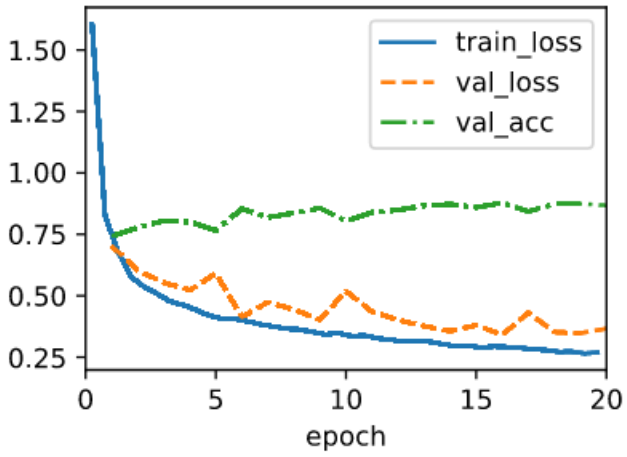

Figure 1: Results for MLP Model with 20 Epochs

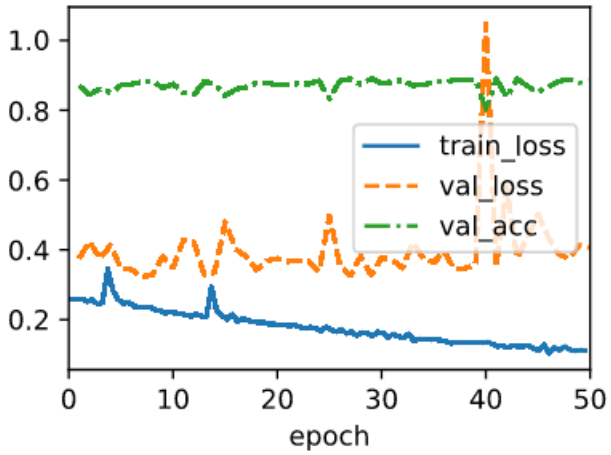

Figure 2: Results for MLP Model with 50 Epochs

For Problem 1, part a, the MLP model was built from scratch using the sample code from the book. First, the model was run over 20 Epochs and Figure 1, shows the same. It can be seen that both the training and validation losses are in the range of 0.25 to 0.6 and the validation accuracy is around 0.75.

Next, for comparison, the same model was run over 50 Epochs, and it can be seen that the training and validation losses have slightly decreased, and the validation accuracy has slightly increased. But, at the same time, the gap between training and validation loss has widened.

Also, it can bee seen that with the change in number of Epochs, the overfitting scenario is still significant as was during the run with 20 Epochs.

Problem 1, Part b:

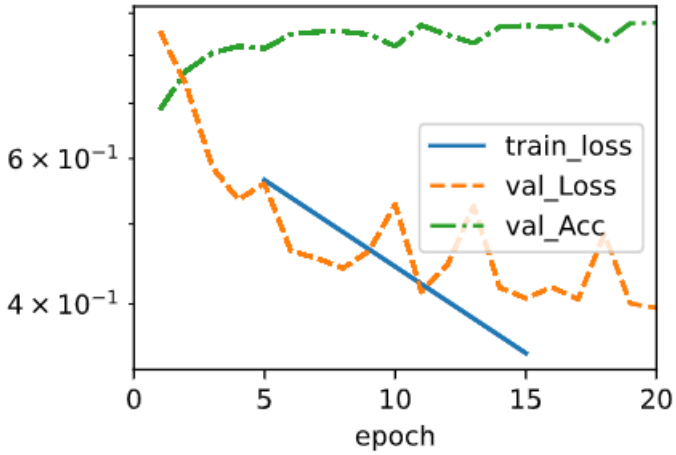

Figure 3: Results for MLP Model with Weight Decay

For Problem 1, part b, the MLP model from part a was used and few changes were made to accommodate the new requirements of Weight Decay. For that, a L2 penalty function was defined and the penalty was added in the Loss Function. The model was run over 20 Epochs and Figure 3, shows the same. It can be seen that the training loss is a linear line between 0.3 to 0.6 and validation loss is in the range of 0.4 to 0.8 and the validation accuracy is around 0.85. Even with the addition of Weight Decay, it can be seen that the Overfitting issue is still prominent. Comparing to the baseline model, the validation loss is showing a more decreasing change while the validation accuracy is slightly higher.

Problem 1, Part c:

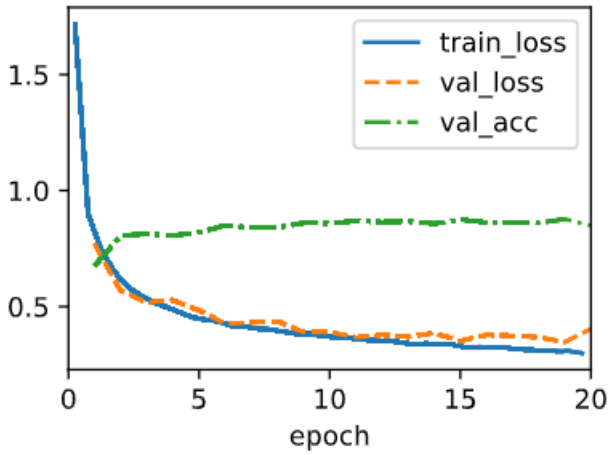

Figure 4: Results for MLP Model with Dropout of 30%

For Problem 1, part c, the MLP model from part a was used and few changes were made to accommodate the new requirements of Dropout. For that, 3 Dropout Layers with 30% dropout were added to the original model. The model was run over 20 Epochs and Figure 4, shows the same. It can be seen that both the training loss and the validation loss are in a range of 0.35 to 0.7 and the validation accuracy is around 0.85. With the addition of Dropout, it can be seen that the Overfitting issue has completely been solved as the generalization gap between the training and validation loss is almost very little to none. Comparing to the baseline model as well as the model with Weight Decay added this model has been the best one with the highest validation accuracy and the smallest generalization gap.

Problem 1, Part d:

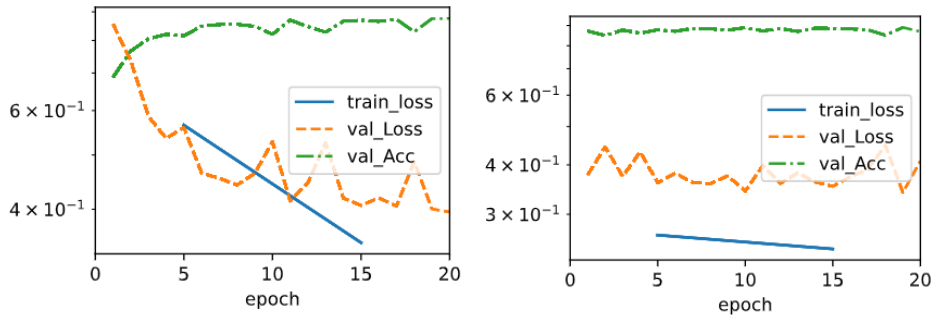

Figure 5: Comparison for Base (LEFT) Vs. Loaded (RIGHT) MLP Model with Weight Decay

For Problem 1, part d, the MLP model with Weight Decay from part b was saved and loaded for retraining the model. In Figure 5 shows, on left side, is the baseline model with Weight Decay added, and on the right is the loaded and re trained model with Weight Decay. It can be clearly seen that the training and validation losses are slightly lower in the loaded model and also the validation accuracy is slightly higher. But, at the same time, it looks like the generalization gap

with the loaded model is slightly bigger. The comparison with respect to training times, both the models took around the same amount of time to train. The two Figures below show the same.

```
Total Training Time (From Scratch):  255.93493556976318

Total Training Time (From Saved Model):  255.7039134502411
```

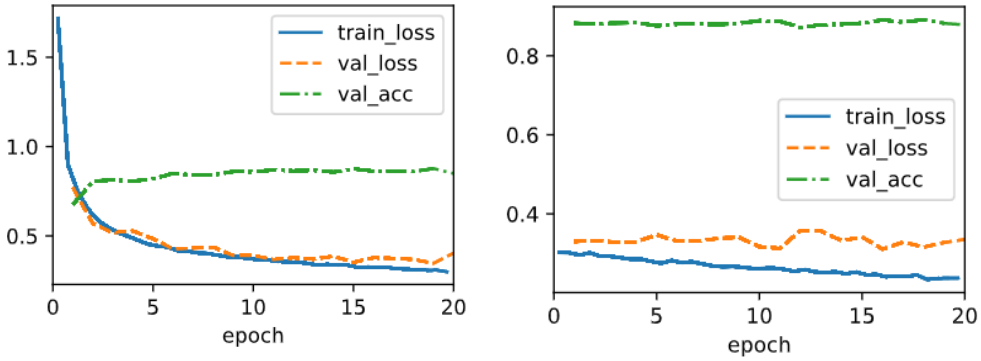

Figure 5: Comparison for Base (LEFT) Vs. Loaded (RIGHT) MLP Model with Dropout

For Problem 1, part d, the MLP model with Dropout of 30% from part c was saved and loaded for retraining the model. In Figure 6 shows, on left side, is the baseline model with Dropout of 30% added, and on the right is the loaded and re trained model with Dropout. It can be clearly seen that the training and validation losses are slightly lower in the loaded model while the validation accuracy is almost the same. But, at the same time, it looks like the generalization gap with the loaded model is slightly bigger. The comparison with respect to training times, the loaded model took around 2 seconds less to train. The two Figures below show the same.

```
Total Training Time (From Scratch):  268.2879922389984

Total Training Time (From Saved Model):  266.3024890422821
```

Problem 2, Part a:

Question:What happens if we need to standardize the continuous numerical features like what we have done in this section?

Answer: Standardization is performed on continuous numerical features if the features in the data set have differences between their ranges or if the features are measured in different units. So, to avoid the data from producing misleading results, standardization should be performed. For the

Housing dataset, features like the LotFrontage and LotArea, both have different ranges of the val ues and to avoid this data from producing the wrong output, standardization should be perfomed.

Problem 2, Part b:

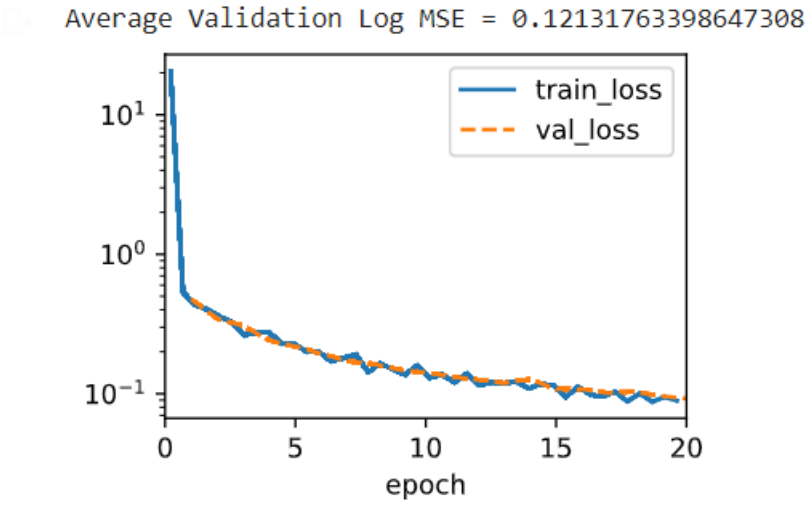

Figure 7: Complex Model of Linear Regression

With the baseline Linear Regression model provided, the model was made more complex by further adding 2 Lazy Linear and ReLU layers and changing the weight to 0.5. Figure 7 shows the same. It can be seen the Mean Squared Error is around 0.12 and both the training and validation losses are almost equal to each other, so no overfitting issue is present and the generalization gap is minimal to none.

Problem 2, Part c:

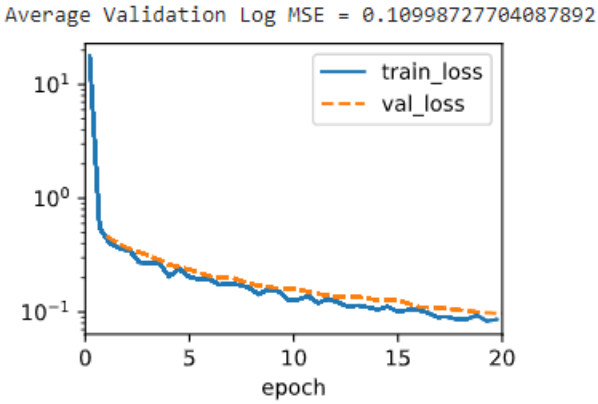

Figure 8: Complex Model of Linear Regression with Dropout and Weight Decay Added

The complex model from part b was used to create a new model with 2 added Dropout layers with 70%. On top of the Dropout Layers, the Weigh Decay was also added on the same model.

For adding the Weigh Decay, a L2 Penalty function was defined, and the original Loss function was modified to take into consideration the parameter penalty. Figure 8, above, shows the same. It can be seen the Mean Squared Error is around 0.10 and both the training and validation losses are almost equal to each other, so no overfitting issue is present, and the generalization gap is minimal to none. With comparison to the previous model without any dropout layers, this model has a lower loss value.

Problem 2, Part d:

After exploring the Dropout and Weight Decay options, the initial predictions were submitted to Kaggle and the first prediction's score was 0.33758 which put me at about 3500th position in the Leaderboard. So, after experimenting with different values for learning rate, number of layers, number of k values, and number of Epochs, the set which gave me the best score of 0.17026 using learning rate of 0.01, a K value of 15, 4 layers, and 250 Epochs put me at about 2728th position in the Leaderboard. Figure 9 below shows the decreasing graph of the Training and Validation loss. It can be seen from the graph that the generalization gap is also almost minimal to none. The Average Validation Log MSE is about 0.028. Figure 10, below, shows my ranging on the Leaderboard.

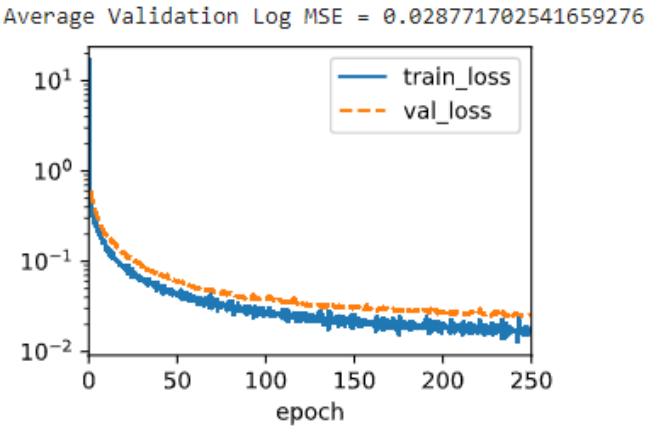

Figure 9: Result of the Best Model

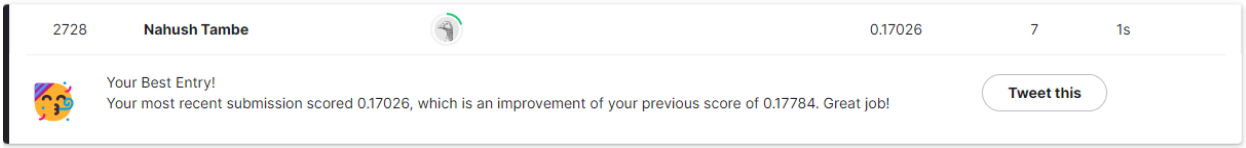

Figure 10: Ranking on the Leadership Board of Kaggle
