**AML – 64016 – Assignment 2**

*apeter@kent.edu*

GitHub link - <https://github.com/Aloysius95/apeter_64061/tree/4546a7de2a5e2b7b527a4f7a0e20fc0765add2ae/Assignment2>

1. **You used two hidden layers. Try using one or three hidden layers and see how doing so affects validation and test accuracy.**

* The base model with 2 hidden layers shows the highest validation accuracy and the lowest loss. Which outperforms both the model with 1 and 3 hidden layers.
* Model 1 with 1 hidden layer performs slightly worse than the base model in both validation and training accuracy.
* Model 3 with 3 hidden layer does not improve the performance, which shows that the additional layers do not always enhance the performance and also results in instability.
* Overall, the base model with two hidden layers provides the most stable and optimal performance.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **Base Model** | 0.28 | 0.89 |
| **1HL** | 0.28 | 0.89 |
| **3HL** | 0.37 | 0.87 |

*\*\* Note the code snipping’s will be attached at the end of the report.* A graph of different colored lines

Description automatically generated A graph with lines and numbers

Description automatically generated with medium confidence A graph of a number of different colored lines

Description automatically generated A graph of different colored lines

Description automatically generated

1. **Try using layers with more hidden units or fewer hidden units: 32 units, 64 units, and so on.**

* In validation accuracy the model with 32 units actually performs better than 16 and 64 units. The model with 64 units shows lower and more variable validation accuracy while the model with 32 units performed similar to the base model which is 16 units.
* The model with 64 hidden units has the lowest validation loss when compared to 32 and 16 hidden units model, especially at higher epochs which suggests that increasing the hidden units may lead to overfitting and suboptimal generalization.
* All the models with units 13, 32 and 64 performed similar to each other in training accuracy and loss but the base model performed slightly better as a whole.
* To conclude increasing the no of hidden units from 16 to 32 to 64 did not improve the model that much in these models but it gave a lot of volatility because of which the base model is a better option among them.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **Base Model** | 0.28 | 0.89 |
| **32HU** | 0.29 | 0.88 |
| **64HU** | 0.29 | 0.88 |

A graph of different colored lines

Description automatically generated

A graph with lines and numbers

Description automatically generated

A graph of a number of different colored lines

Description automatically generated

A graph of loss comparison across models

Description automatically generated

1. **Try using the MSE loss function instead of binary cross entropy.**

* BCE is better suited for binary classification problem, in this case the MSE performs significantly better than BCE.
* The MSE majorly outperforms in validation loss comparison, where it shows a lower val loss compared to the base model. This shows the MSE model is better able to minimize the error between predicted and true values.
* The MSE model gives better results with less training (with lesser EPOCHS )
* In conclusion the MSE model is clearly better than using BCE model for this specific model building.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **Base Model** | 0.28 | 0.89 |
| **MSE** | 0.09 | 0.88 |

A graph of a graph showing the value of a model

Description automatically generated with medium confidence

A graph with blue line and orange line

Description automatically generated

A graph of a curve

Description automatically generated

A graph of a loss

Description automatically generated

1. **Try using the tanh activation (an activation that was popular in the early days of neural networks) instead of relu**

* The relu activation (Base line model) performs better than tanh activation, where relu’s accuracy is better.
* Relu activation model shows a lower loss compared to the tanh model which shows that the relu model is better to minimize the error between predicted and true value.
* In conclusion, relu is a better option because it outperforms tanh in accuracy and has a lower loss.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **Base Model** | 0.28 | 0.89 |
| **Tanh** | 0.28 | 0.88 |

A graph of a graph

Description automatically generated with medium confidence

A graph with a line and a line

Description automatically generated with medium confidenceA graph with blue and orange lines

Description automatically generatedA graph of loss comparison across models

Description automatically generated

1. **Use any technique we studied in class, and these include regularization, dropout, etc., to get your model to perform better on validation.**

* Dropout optimization has the best accuracy compared to the base line and L2 regularization.
* Dropout again performs better at the loss where it has the lowest error rate, with L2 close behind and the base model being not that effective.
* Mostly compared to all the graphs the dropout is a better optimization method for this model because it performs the best compared to baseline and L2.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **L2** | 0.34 | 0.88 |
| **Dropout** | 0.32 | 0.88 |

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

A graph of a number of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

**Comparison of all the models.**

* For the validation accuracy model among all the models, the 32 hidden units performed the best.
* For the accuracy comparison, tanh and the base model together performed very similar, and they were the best in this comparison.
* For the validation loss and the loss comparison the 64 hidden units model came out to be the best when compared to other models.
* To conclude the model a good balance and performed equally acceptable on all fronts was the 64 hidden units.

|  |  |  |
| --- | --- | --- |
|  | **Test loss** | **Test Accuracy** |
| **Base Model** | 0.28 | 0.89 |
| **1HL** | 0.28 | 0.89 |
| **3HL** | 0.37 | 0.87 |
| **32HU** | 0.29 | 0.88 |
| **64HU** | 0.29 | 0.88 |
| **MSE** | 0.09 | 0.88 |
| **Tanh** | 0.28 | 0.88 |
| **L2** | 0.34 | 0.88 |
| **Dropout** | 0.32 | 0.88 |

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generatedA graph of different colored lines

Description automatically generated

**Following is the code explanation**

* Inputting the dataset and preparing the data – the code was given by professor.
* Each model was created differently for model building to make it easier to compare them, in total there were around 9 models.
* Then each model was trained and plotted to see its accuracy and loss.
* Each model was retrained as per the base model given by the professor.
* Based on the questions each model were put together to compare and all the models were compared together, the results are above.

**\*\*** *Because of the length of the model the google colab page was printed and attached at the end of the report.*