**Name: Aishwarya Lingam**

**Course: Advance Machine Learning**

**Assignment: Neural Network**

**Git Hub link: https://github.com/alingam9/aishwaryalingam\_MachineLearning/tree/main/Assignment2**

**Neural Network Summary Report**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Models Designed** | **Layers and Units Used** | **Activation** | **Regularization** | **Drop out** | **Loss Function** | **Validation Accuracy** | **Validation loss** | **Test Accuracy** | **Test loss** |
| Model 1 | 2 Hidden layers, 16 units | ReLU | None | None | Binary Crossentropy | 87.74 | 43.09 | 88.52 | 28.89 |
| Model 2 | 1 Hidden layers, 16 units | ReLU | None | None | Binary Crossentropy | 87.98 | 30.28 | 88.46 | 28.61 |
| Model 3 | 3 Hiden layers, 16units | ReLU | None | None | Binary Crossentropy | 86.96 | 38.52 | 87.93 | 31.13 |
| Model 4 | 2 Hidden layers, 32 units | ReLU | None | None | Binary Crossentropy | 87.07 | 37.07 | 88.22 | 29.45 |
| Model 5 | 1 Hidden layers, 32 units | ReLU | None | None | Binary Crossentropy | 86.08 | 41.12 | 88.22 | 29.45 |
| Model 6 | 2 Hidden layers, 16 units | Tanh | L2(0.01) | None | MSE | 88.39 | 37.36 | 88.21 | 42.04 |
| Model 7 | 3 Hidden layers, 16 units | ReLU | None | Drop out(0.5) | MSE | 85.17 | 18.64 | 81.69 | 22.96 |
| Model 8 | 1 Hidden layers, 32 units | Tanh | None | Drop out(0.5) | MSE | 85.88 | 14.6 | 84.68 | 15.38 |
| Model 9 | 2 Hidden layers, 64 units | ReLU | None | None | Binary Crossentropy | 88.71 | 8.27 | 88.75 | 8.74 |

1. **Model’s Architecture:**

The effectiveness of the models is influenced by the quantity of hidden layers and units used.

Good generalization is demonstrated by

1. Model 2 (1 hidden layer, 16 units), which has the lowest validation loss (30.28) and the highest validation accuracy (87.98%).
2. Model 1 (2 hidden layers, 16 units) achieves a similar validation accuracy (87.74%) despite having a higher validation loss (43.09), which might be a sign of overfitting.
3. Model 3 (3 hidden layers, 16 units) has the largest validation loss (38.52) and the lowest validation accuracy (86.96%), suggesting that it is too complicated without providing any value.
4. Model 4's performance is comparable to that of Model 2 but not much better (2 hidden layers, 32 units).
5. **Insights Based on Activation Functions used**

Activation functions play a critical role in neural networks by introducing non-linearity, enabling the model to learn complex patterns and relationships in the data. Without activation functions, the network would behave like a linear model, regardless of the number of layers, severely limiting its learning capacity.

**ReLU (Rectified Linear Unit):**

All models in the dataset utilize the ReLU (Rectified Linear Unit) activation function. This is one of the most commonly used activation functions in deep learning due to its simplicity and effectiveness.

* Model 2 performs the best, with the highest validation accuracy (87.98%) and the lowest validation loss (30.28).
* All models use ReLU, so the differences in performance are due to architectural choices rather than activation functions.
* ReLU is a strong default choice—but if performance plateaus, trying Leaky ReLU or adding Batch Normalization may help.

**Final Insights and Recommendations**

* ReLU is the best choice for these models given their depth and structure.
* Potential improvement: Consider using Leaky ReLU if models experience dying ReLU issues.
* If overfitting occurs: Try Batch Normalization or a small dropout to stabilize training.

1. **Loss Functions:**

Loss functions play a critical role in training deep learning models by quantifying the difference between predicted and actual values. The chosen loss function determines how the model updates its weights during training.

**Binary Crossentropy**

All models in the dataset utilize Binary Crossentropy as the loss function. This is the standard choice for binary classification problems, where the goal is to categorize inputs into one of two classes (e.g., positive/negative, spam/not spam, etc.).

* Lower loss values indicate better performance since they reflect how accurately the model's predictions align with actual labels.
* Model 2 has the lowest validation loss (30.28) and test loss (28.61), suggesting better generalization compared to the other models.
* Model 1 has a significantly higher validation loss (43.09) compared to its test loss (28.89), which might suggest some overfitting.
* Model 3 and Model 4 have relatively higher losses, indicating that increasing model complexity (more layers or units) does not necessarily lead to better performance.

**Potential Alternatives to Binary Crossentropy**

Although Binary Crossentropy is a solid choice, alternative loss functions may sometimes offer advantages:

**Hinge Loss (for SVM-like behavior):**

* Encourages larger margins between classes.
* Useful for models that need more robustness against misclassified points.

Focal Loss (for imbalanced datasets):

* Addresses class imbalance by focusing more on hard-to-classify examples.
* Can be useful if one class is much more frequent than the other.
* Mean Squared Error (MSE) or Mean Absolute Error (MAE)

These are generally used for regression, but in some cases, they can help with classification when treating probabilities as continuous values.

**4. Regularization (L2 and Dropout):**

Model 1 indicates modest overfitting, which might be lessened by dropout or L2 regularization, since it has a greater validation loss than Model 2.

In deeper models (such as Model 3), adding dropout (e.g., 0.2–0.3) may improve generalization.

**To improve the model even more:**

To avoid overfitting, use L2 regularization or dropout (0.2–0.3).

Try somewhat larger architectures (two hidden layers, 16–32 units, etc.) while maintaining regularization.

**Last Model Suggestion:**

The best-performing model, according to the results, is Model 2 (1 hidden layer, 16 units, ReLU activation, Binary Crossentropy loss). It exhibits the best balance between learning and generalization, with the highest validation accuracy (87.98%), the lowest validation loss (30.28), and the lowest test loss (28.61).