Model Architecture and Performance

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Architecture** | **Activation** | **Regularization** | **Drop out** | **Loss Function** | **Validation Acuuracy** | **Validation Acuuracy** | **Test Accuracy** | **Test loss** |
| Model 1 | 2 hidden layer,16 units | ReLU | None | None | Binary crossentropy | 86.81% | 36.89 | 86.52% | 28.57 |
| Model 2 | 2 hidden layers,16 units | ReLU | None | None | MSE | 87.51 | 32.37 | 88.42 | 28.75 |
| Model 3 | 1 hidden layer,16 units | Tanh | None | None | Binary crossentropy | 87.17 | 37.37 | 86.47 | 34.8 |
| Model 4 | 3 hidden layers,32 units | ReLU | None | None | MSE | 87.91 | 30.73 | 88.55 | 28.13 |
| Model 5 | 2 hidden layer, 64 units | Tanh | None | Drop out (0.5) | Binary crossentropy | 87.94 | 31.56 | 88.05 | 29.76 |
| Model 6 | 1 hidden layer,64 units | ReLU | L2 (0.01) | None | MSE | 88.62 | 8.41 | 88.58 | 8.77 |
| Model 7 | 3 hidden layer,32 units | Tanh | None | Dropout (0.5) | Binary crossentropy | 88.47 | 8.56 | 88.43 | 8.68 |
| Model 8 | 2 hidden layer, 64 units | ReLU | None | None | MSE | 86.74 | 0.14 | 86.09 | 1.48 |
| Model 9 | 1 hidden layer, 16 units | ReLU | None | None | Binary Crossentropy | 88.77 | 0.82 | 88.82 | 0.86 |

**Model Architecture's Effect**

The performance of the model is largely determined by its architecture. Models with different numbers of hidden layers and units are included in the dataset. Models with more neurons and deeper topologies typically exhibit somewhat higher validation accuracy.

Compared to the others, Models 4 (3 hidden layers, 32 units) and 5 (2 hidden layers, 64 units with dropout) had good test accuracy (~88.5%).

The test accuracy of shallower networks (one hidden layer, 16 units) was marginally lower (~86.4%).

Although generalization was enhanced by adding additional layers and units, calculation time rose as a result.

Therefore, in this case, a deeper network with optimum regularization produces superior outcomes.

**Functions of Activation**

The way information spreads throughout the network is influenced by several activation functions. The dataset's characteristics ReLU and Tanh:

Most models use the Rectified Linear Unit (ReLU), which speeds up convergence and lessens the vanishing gradient issue.

Tanh: It may experience saturation at extreme values, although it generally performs better when the data is centered around zero, as seen in Models 3 and 5.

Results:

In this dataset, ReLU-based models (Model 1, Model 2, and Model 4) outperformed Tanh models in terms of validation accuracy.

Because it can avoid vanishing gradients, ReLU is typically chosen for deep networks.

Therefore, the suggested activation function for this issue is ReLU.

**Loss Functions**

Two different loss functions were used: Binary Crossentropy and Mean Squared Error (MSE).

Binary Crossentropy: Helps with faster and more stable convergence, ideal for classification issues.

Usually utilized for regression, MSE was also incorporated in several of these models.

Notes:

Models (e.g., Model 1, Model 3, Model 5) that were trained using Binary Crossentropy regularly demonstrated high test accuracy.

MSE-using models (e.g., Model 2, Model 4) had comparable test accuracy but marginally higher validation accuracy.

The suggested loss function for this classification problem is Binary Crossentropy.

**Regularization (Dropout and L2)**

Overfitting can be avoided with regularization techniques:

Dropout: Only Model 5 used dropout (0.5), which was helpful for generalization and helped maintain competitive test accuracy (88.05%).

The dataset did not specifically use L2 regularization.

**Results:**

**A graph of a line graph

AI-generated content may be incorrect.**

**A graph with blue dots

AI-generated content may be incorrect.**

**A graph of a graph with blue dots

AI-generated content may be incorrect.**

A graph with blue dots

AI-generated content may be incorrect.

A graph of a graph

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

A graph of training and validation loss

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

A graph of a graph with blue dots

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

A graph of training loss

AI-generated content may be incorrect.

A graph with blue lines and dots

AI-generated content may be incorrect.

A graph with blue lines

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

A graph of a loss

AI-generated content may be incorrect.

A graph of a line graph

AI-generated content may be incorrect.

A graph of a training loss

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

As demonstrated in Model 5, dropout improved generalization and avoided overfitting.

While their accuracy was similar, models without dropout were more likely to overfit.

For improved generalization, dropout (0.5) is therefore advised.

**Last Model Suggestion**

According to the findings, the optimal compromise between accuracy and generalization is demonstrated by Models 4 (3 hidden layers, 32 units, ReLU, MSE loss) and 5 (2 hidden layers, 64 units, Tanh, dropout 0.5, Binary Crossentropy).

**Model Suggestion:**

Three buried layers, each with 32–64 units

Activation of ReLU

Crossentropy loss in binary

Dropout for regularization (0.5)

With fresh, untested data, this model should function fine.