ASSIGNMENT 1

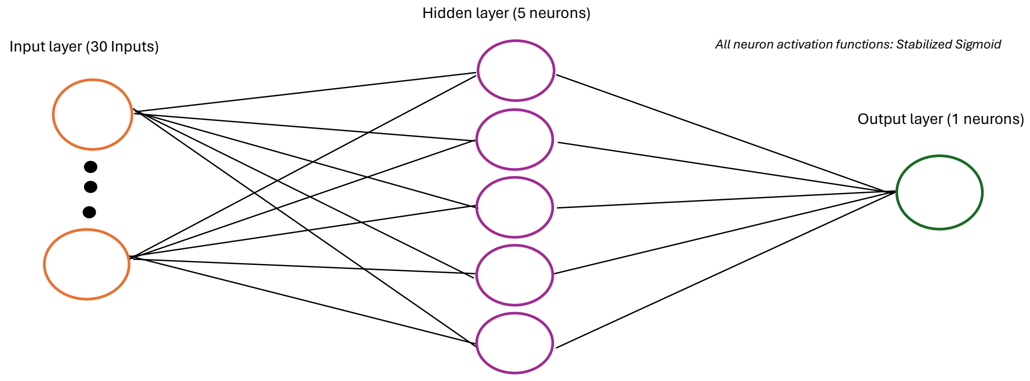
CSC3066 – Deep Learning

Date: 23/02/2024

Student Name: Ryan McKee
Student Number: 40294886

Introduction

This project endeavours to construct an artificial neural network for diagnosing breast cancer using the Wisconsin dataset. The architecture comprises an input layer with 30 nodes, a hidden layer with 5 neurons employing sigmoid activation functions, standard normalization in preprocessing, and an output layer providing classification probabilities. This report evaluates model performance, proposes enhancements, implements them, and analyses outcomes to optimize the classification model.



Baseline Model Evaluation

To comprehensively assess the initial model's performance, I implemented 5-fold cross-validation on the training and testing data. This method systematically rotates data subsets for training and validation, effectively utilizing the limited Wisconsin dataset, which contains only 569 observations. It generates testing on unseen data subsets, enhancing understanding of the model's bias-variance trade-off and detecting underfitting or overfitting, thus aiding in the assessment, refinement, and selection of the best model for more accurate and robust classification.

Various evaluation techniques, including cross-validation and confusion matrices, were utilized to assess performance and class-specific metrics in this binary classification task. These tools facilitated the analysis of imbalance, errors, threshold adjustments, and model comparisons. Moreover, error rates and accuracy across epochs for each fold during training were monitored. By aggregating the fold results, I obtained a precise overview of the model's performance. Additionally, ROC curves were generated to effectively compare baseline and final models, while also gaining insights into how classification thresholds impact model performance.

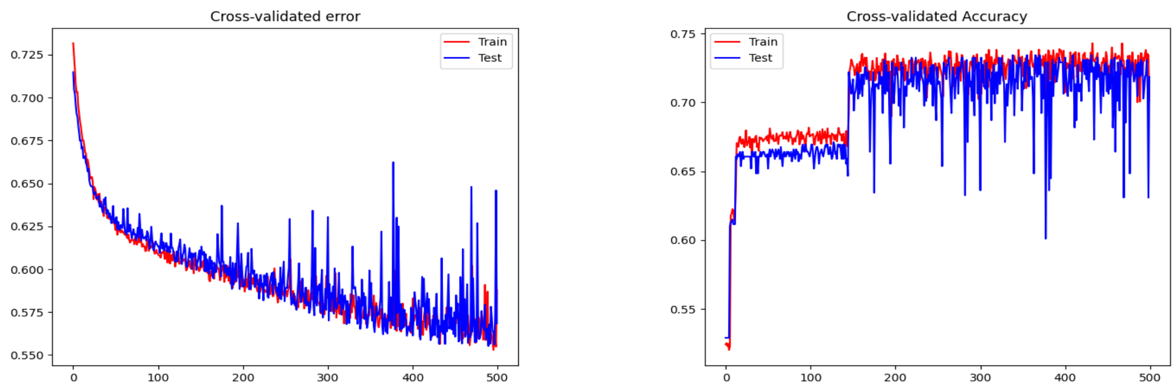


Figure 1: Cross-validated Error and Accuracy graphs plotted against epochs.

The initial evaluation of the neural network model reveals intriguing insights into its performance dynamics. Notably, the training process exhibits a decrease in error by 0.162

and an increase in accuracy by 0.203, while the testing phase mirrors similar improvements. However, despite these advancements, the model's performance nuances warrant deeper investigation.

Training Statistics			Testing Statistics		
	Error	Accuracy		Error	Accuracy
Pre-trained	0.731742	0.524296	**Pre-trained**	0.714878	0.529110
Post-trained	0.568751	0.728212	**Post-trained**	0.554755	0.732619
Improvement	0.162991	0.203916	**Improvement**	0.160123	0.203509

One prominent observation lies in the model's convergence behaviour over epochs. The graphical representation (figure 1) underscores a plateau in training accuracy after approximately 150 epochs, accompanied by significant fluctuations in both training and testing error rates. Such erratic behaviour hint at the underlying challenges that may be hindering the model's ability to reach global optima efficiently.

A critical factor contributing to these fluctuations is potentially the lack preprocessing and normalization of the dataset may be rendering the model overly sensitive to input feature scales. Normalization techniques can mitigate this issue by stabilizing the training process and preventing the model from getting trapped in suboptimal solutions or saddle points therefore may improve the performance significantly.

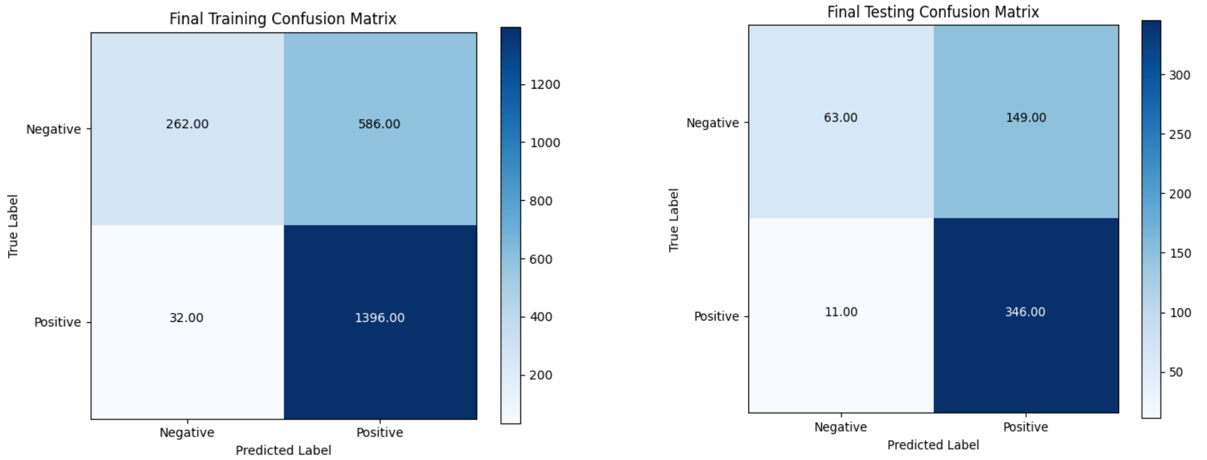


Figure 2: Cross-Validated Confusion matrices for Training and Testing

	Precision	Recall	F1 Score (Macro)	Accuracy
Training	0.725738	0.977577	0.825514	0.728512
Testing	0.723247	0.969014	0.817535	0.718584

Further analysis of confusion matrix statistics sheds light on the model's precision, recall, F1 score, and overall accuracy. In both the training and testing datasets, the model exhibits commendable precision and recall scores, indicating its ability to accurately identify positive instances while minimizing false positives and false negatives. The F1 scores, which represent the harmonic mean of precision and recall, underscore the balance between the model's ability to make precise predictions and its capacity to capture all positive instances. Additionally, the confusion matrix accuracies provide a holistic view of the model's classification accuracy across both datasets. Based on the confusion matrix results the model is currently performing quite well.

To gain a more holistic view of the performance of this model and add more context to the previous statistics, An ROC curve was also created. The ROC curve below shows that the current model has an Area under curve of 0.5, indicating the need to improve the model, as essentially the model has poor discriminatory ability despite what the confusion matrix statistics where signifying.

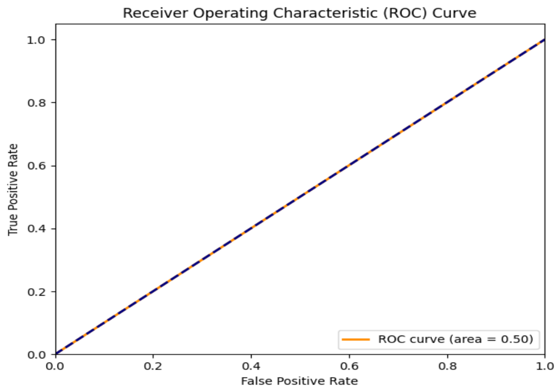


Figure 3: ROC curve for baseline model.

Despite the model's promising performance, additional improvements can be made. The model's architecture and complexity emerge as pivotal considerations in its performance optimization. Exploring alternative activation functions like ReLU or Leaky ReLU could enhance convergence properties and alleviate gradient saturation issues inherent in sigmoid functions. Furthermore, fine-tuning hyperparameters such as learning rates and optimization algorithms presents avenues for improving training stability and convergence.

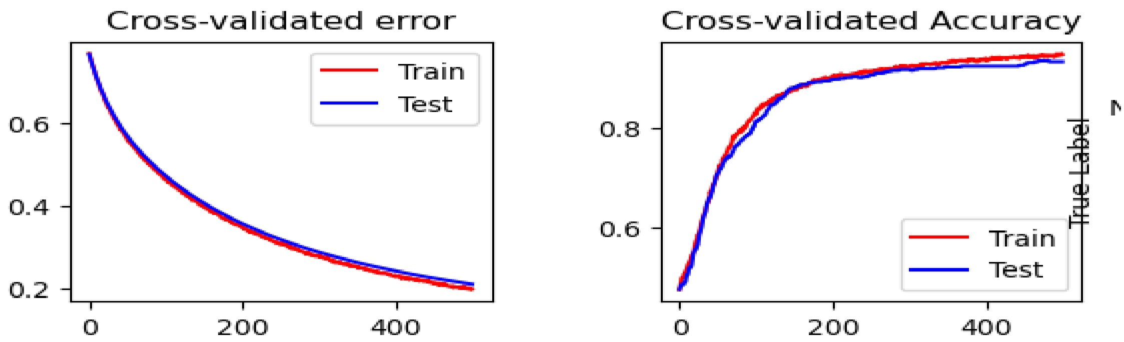
Baseline Model Improvements

To enhance the classification performance of the neural network model, a systematic approach was adopted to address key limitations identified during the evaluation process. The strategy involved iterating through preprocessing techniques, adjusting model architecture, exploring different activation functions, and optimizing hyperparameters.

Pre-processing:

Initial observations revealed fluctuations in error and accuracy graphs, indicating instability in the feature vector data used for training. To address this issue, standard scaling was applied to both the training and test data during cross-validation. This simple adjustment yielded significant improvements across performance metrics, indicating that data instability was a primary factor contributing to the initial model's poor performance.

	Precision	Recall	F1 Score	Accuracy	AUC
Training	0.947723	0.953881	0.950533	0.937612	0.98
Improvement	0.221985	-0.023696	0.125019	0.2091	0.48
Testing	0.936667	0.986186	0.955691	0.945819	0.98
Improvement	0.21342	0.017172	0.138156	0.227235	0.48



Activation Functions, Model Architecture, and Hyperparameters:

Further enhancements were explored by experimenting with different activation functions. Transitioning the hidden layer from sigmoid to ReLU to accelerate convergence therefore lower training time not only improved training time but also overall model performance. Adjustments to hyperparameters, such as decreasing batch size (128 to 100) to better improve chances of finding global minimum and tweaking the learning rate (0.01 to 0.03) to increase the gradient changes each epoch, contributed to further enhancements in training time and model performance.

Increasing the number of neurons in the hidden layer (5 to 15) also proved beneficial, as it allowed for capturing more complex patterns in the data, resulting in improved classification accuracy.

Final Model:

The culmination of these improvements resulted in a final model that exhibited remarkable performance gains across all key metrics and improved the training time substatially therefore providing a feasible classification model for the Wisconsin Breast Cancer dataset:

	Precision	Recall	F1 Score	Accuracy	AUC
Training	0.985437	0.996508	0.990939	0.988577	1
Improvement	0.259699	0.01861	0.165425	0.260065	0.5
Testing	0.977452	0.980992	0.978994	0.973622	1
Improvement	0.254205	0.011978	0.161459	0.255038	0.5

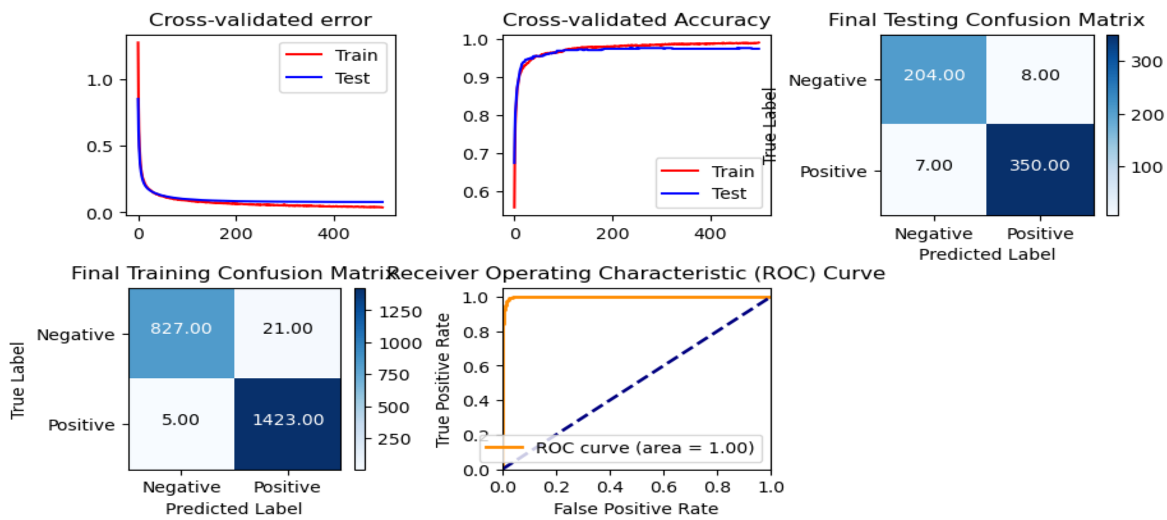


Figure 4: Final model performance visualisation.