Logistic Regression Model for Titanic Survival Prediction

Analysis of Training and Validation Losses

To evaluate the performance and convergence of the logistic regression model, it was monitored the changes in training and validation losses over 100 iterations of gradient descent. The graph presented in Figure 1 provides a visual representation of these losses throughout the iterative optimization process.

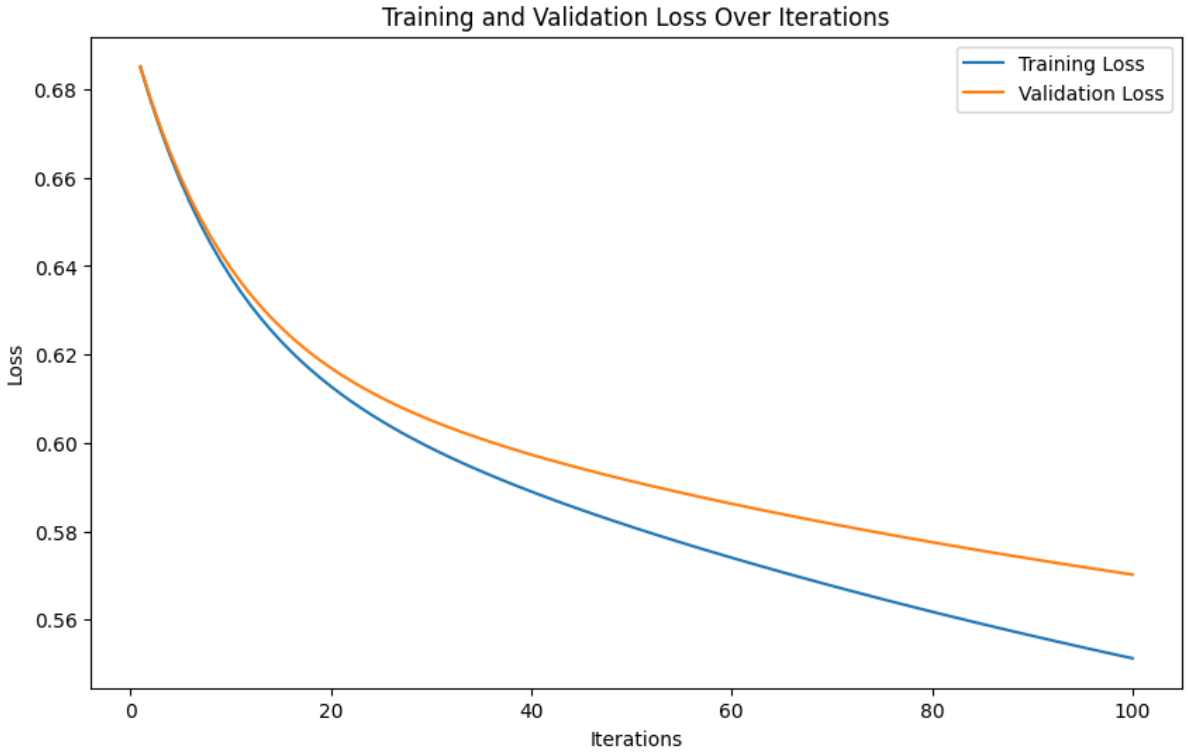

Figure 1: Training and Validation Loss Over Iterations. This graph illustrates the decline in both training and validation losses as the logistic regression model is optimized via gradient descent over 100 iterations. The x-axis represents the iteration count, while the y-axis indicates the loss. The convergence of the two curves suggests a model that fits well with the training data and generalizes effectively to validation data. As seen in the graph, both training and validation losses decrease significantly as the number of iterations increases, an indication of effective learning. The training loss exhibits a consistent decline, showing the model's improving fit to the training data. The validation loss demonstrates a similar downward trend. The parallel reduction of both training and validation losses is a positive indication that our model's performance is enhancing on familiar (training) and unfamiliar (validation) data.

Hyperparameter Tuning

The performance of the logistic regression model relies heavily on the selection of hyperparameters, particularly the learning rate and the number of iterations for the gradient descent algorithm. To this end, we executed a comprehensive grid search to determine the optimal hyperparameter combination. It evaluated four distinct learning rates—0.01, 0.05, 0.1, and 0.5—across four iteration counts: 50, 100, 200, and 500, resulting in sixteen unique configurations.

This experiment determined that a learning rate of 0.5 in conjunction with 500 iterations yielded the most favorable outcome, attaining the lowest validation loss of 0.5105. This setting also achieved a validation accuracy of approximately 74.72%, signifying that the model, with these hyperparameters, can make predictions with an appreciable level of confidence while effectively managing the bias-variance trade-off.

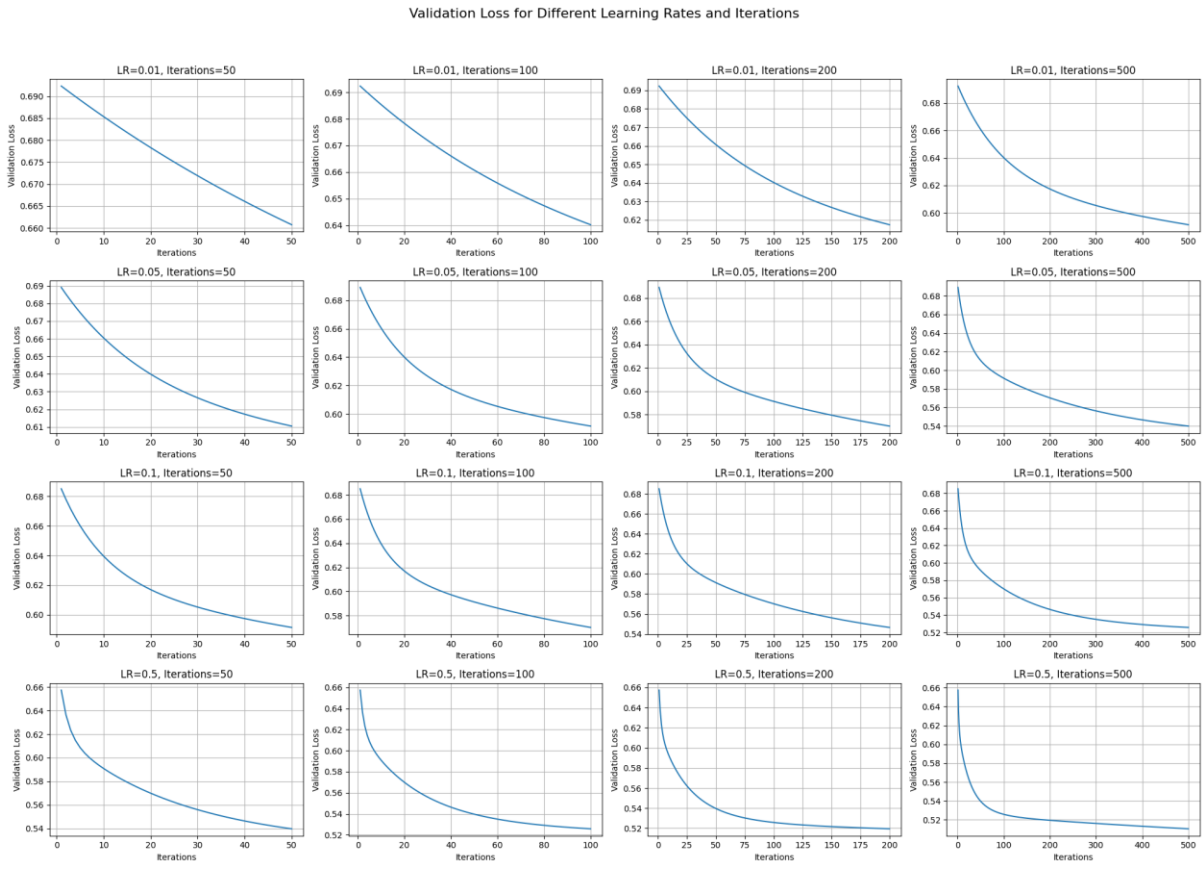

Figure 2: Validation Loss for Different Learning Rates and Iterations. This composite figure shows the validation loss across various iterations for each learning rate tested. It depicts the impact of learning rates and iteration counts on the convergence of the logistic regression model.

Final Model Performance

Following hyperparameter optimization, the final model was trained using the amalgamated training and validation datasets. This strategy aimed to leverage the larger pool of data to enhance the model's generalization capabilities. When evaluated on the test dataset, the refined logistic regression model attained an accuracy of approximately 79.32%. This level of accuracy is indicative of the model's strong generalization performance and substantiates the efficacy of the chosen hyperparameters for this binary classification task.