

CS412 Homework 3 – Titanic Logistic Regression Model Report

1. Introduction

In this report, we present the evaluation results of a logistic regression model trained on the Titanic dataset. The goal of the model is to predict whether a passenger survived the Titanic disaster based on features such as age, sex, and passenger class.

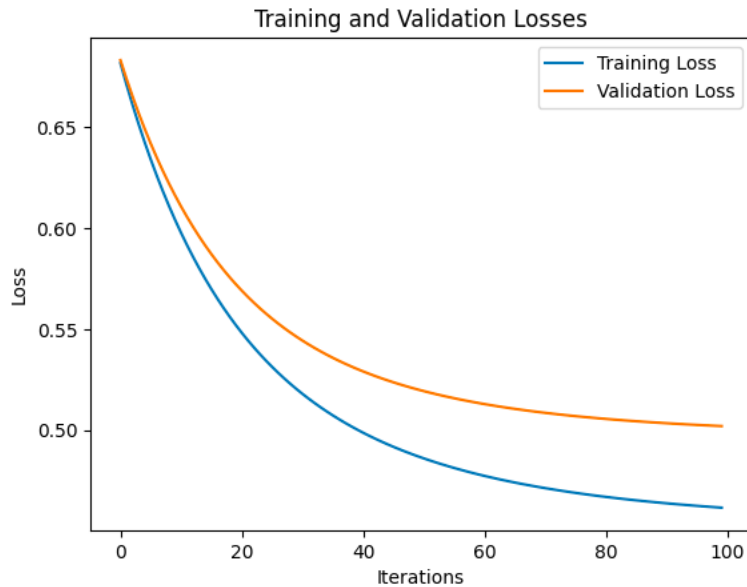
2. Data Preprocessing

The dataset was preprocessed as follows:

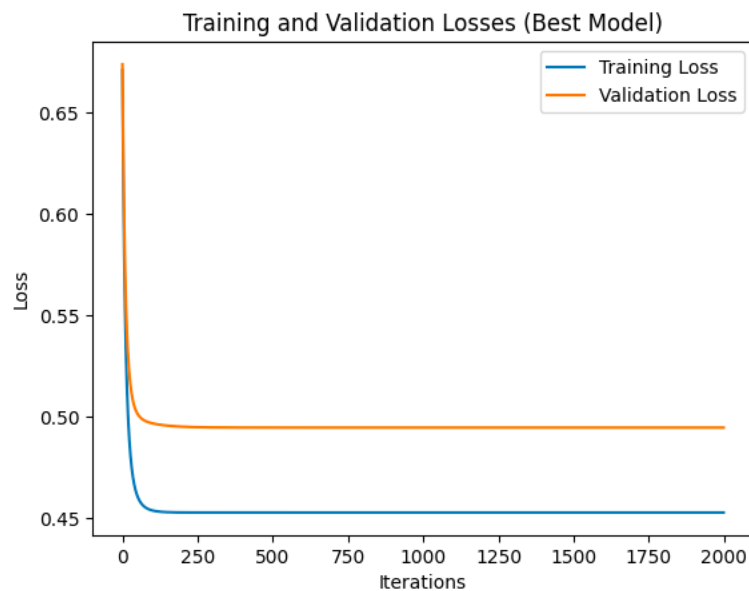
- Random seed was set to 42 for reproducibility.
- The data was split into training (60%), validation (20%), and test (20%) sets.
- Features were scaled linearly within the 0-1 range using StandardScaler.

3. Model Training and Hyperparameter Tuning

- Initially, we trained our model with 0.1 step size and 100 iterations using the training data. Validation Loss: 0.5020058575164894



- We trained multiple logistic regression models with varying learning rates and numbers of iterations.
 - Learning Rates: [0.001, 0.01, 0.1, 0.2, 0.5]
 - Number of Iterations: [100, 200, 500, 1000, 2000]
- The best performing model was selected based on the lowest validation loss.
- The chosen hyperparameters for the final model were:
 - Learning Rate: 0.2
 - Number of Iterations: 2000
 - Validation Loss: 0.494793753313271



4. Model Evaluation

- Final test loss after combining the validation and training data and retraining the final model with best hyperparameters: 0.413151685404943
- Accuracy on testing data: 0.8156424581005587 (81%)

5. Conclusion

Based on the given features, the logistic regression model was able to predict survival outcomes with an accuracy of 81% on the testing data. Additional optimization and feature engineering may enhance the model's functionality.