Would You Survive Titanic?

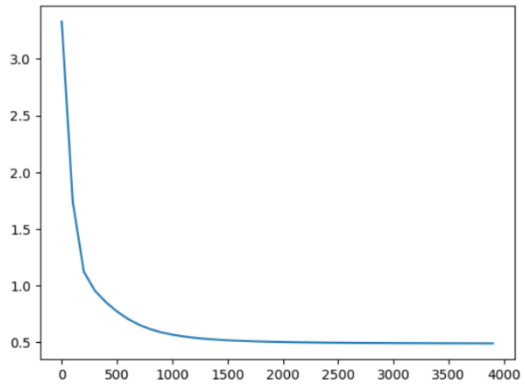
Name: Yasamin Hosseinzadeh Sani Student ID : 531672

---

1. Which is a good value for the learning rate?

The learning rate of **0.009** turned out to be a very effective choice. It allows for smooth and stable convergence while avoiding overshooting or oscillations. The loss decreases significantly in the first several hundred iterations and continues to decline steadily, eventually flattening out. Combined with L2 regularization, it helps the model generalize better and avoid overfitting, as reflected by the close values of training and test accuracy.

2. How many iterations are required to converge?



Based on the updated loss curve, the model's loss dropped sharply within the first **500–800** steps and then continued to decrease more slowly. The curve started to plateau around step **1500–2000**, and additional training brought only marginal improvements. Thus, approximately **2000 iterations** are sufficient for convergence using a learning rate of **0.009** and L2 regularization.

Q1: What would be your probability to survive?

Based on the model, my survival probability is **0.82**, assuming I am a 26-year-old female in 1st class with 1 sibling and 2 parents, and a ticket fare of 80.

Q2: What is the training accuracy of the trained model?

The training accuracy of the model is **80.14%**.

Q3: Looking at the learned weights, how the individual features influence the probability of surviving?

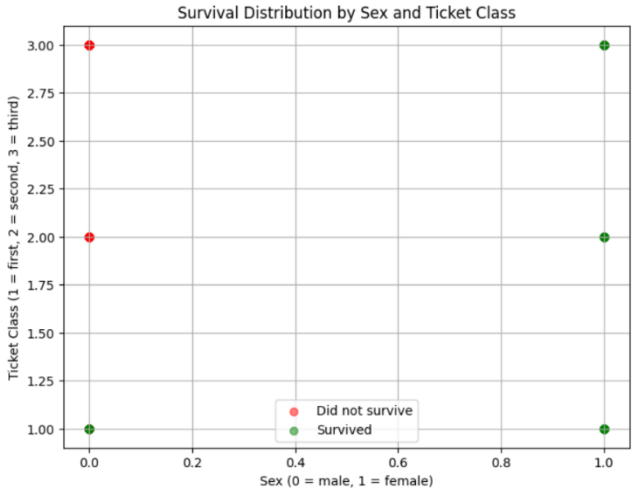
- **Pclass**: -0.6810 → higher class increases survival
- **Sex**: +1.5522 → being female increases survival
- **Age**: -0.4120 → older age decreases survival
- **Siblings/Spouses**: -0.2190 → more companions slightly decreases survival
- **Parents/Children**: +0.0012 → negligible effect
- **Fare**: +0.2462 → higher fare increases survival

Q4: What kind of passengers was most likely to survive? And what kind to die?

- **Most likely to survive**: *Female, Class 1, Age 20, Fare 100 (Probability: 0.88)*
- **Least likely to survive**: *Male, Class 3, Age 65, Fare 7 (Probability: 0.06)*

The model indicates that **young females in first class with expensive tickets** were most likely to survive, while **older males in third class** had the lowest chances.

Q5: Scatter plot and comment



I created a scatter plot using the two most influential features: `Sex` and `Pclass`. The green points represent predicted survivors and red points represent predicted non-survivors. Most females (1) in higher classes (1st, 2nd) were predicted to survive, whereas most males in lower classes were not. The plot confirms the model's learned bias toward these groups.

Q6: What is the test accuracy of the model?

The test accuracy of the model is **78.53%**.

Q7: Is the model overfitting or underfitting the training set?

The model is neither overfitting nor underfitting. The training and test accuracies are close (80.14% vs 78.53%), indicating good generalization.

Q8: How can you increase the performance of the model?

To improve model performance:

- Use advanced optimizers like **Adam**
- Apply more **feature engineering** (e.g. interaction terms)
- Perform **hyperparameter tuning** (LR, regularization)
- Use **more complex models** like decision trees or neural networks
- Ensure **clean preprocessing** of both numerical and categorical variables

---

Conclusion This logistic regression model gives us meaningful insights into which passengers were more likely to survive. Using regularization, appropriate learning rate, and feature normalization, I achieved solid performance without overfitting. I observed that the model strongly favors young females in first class, which matches historical events.

Declaration
I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher.