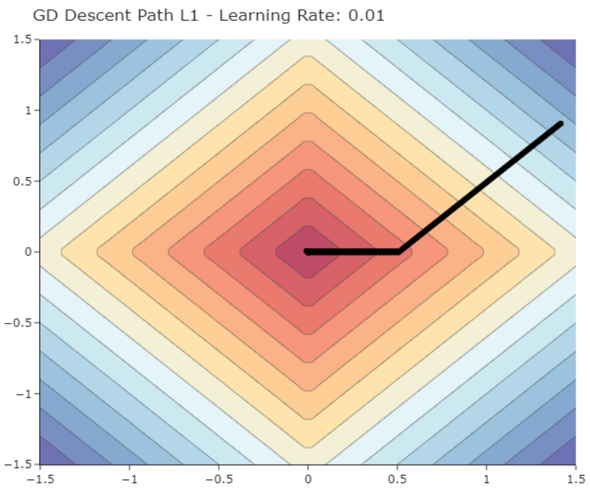
Practical Part

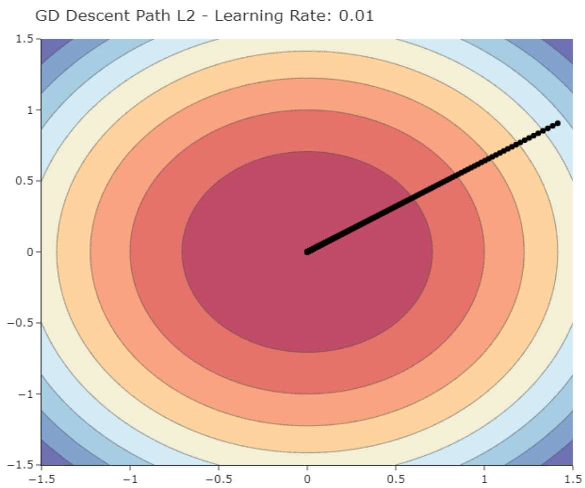
2.1. Gradient Descent

2.1.1. Comparing Fixed learning rates

1. For the ℓ_1 objective, the gradient descent path is irregular and stair-like due to the diamond-shaped contours, causing frequent direction changes. This occurs because not all points on the same level set are equidistant from the minimum.

In contrast, the ℓ_2 objective results in a smooth and direct descent path. The elliptical contours ensure equal distance to the minimum for any two points on the same level set, allowing consistent and efficient movement towards the minimum. This comparison shows that ℓ_2 offers a more stable convergence path than ℓ_1.



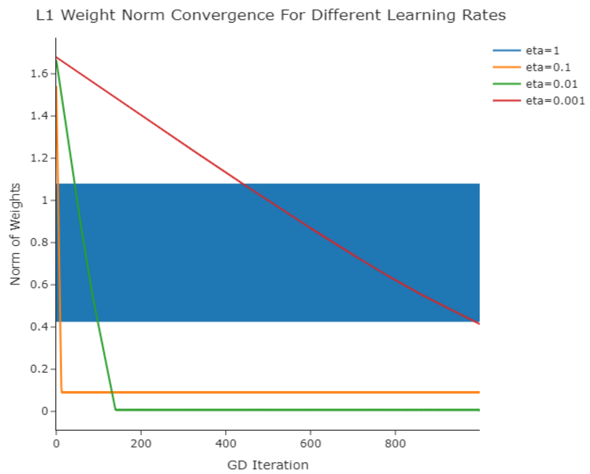


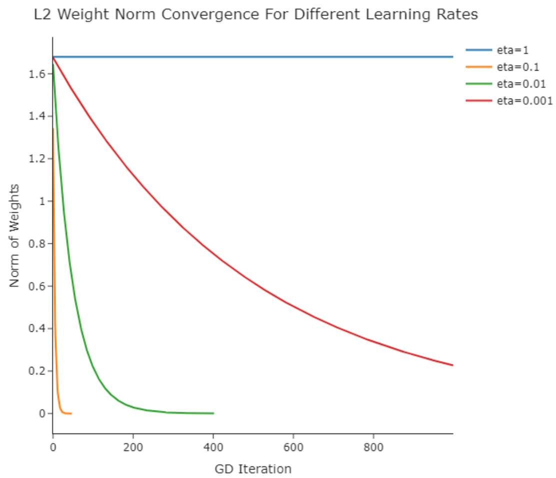
2. The two phenomena that can be seen in the descent path of the ℓ_1 objective are:

- **Sharp Direction Changes:** The ℓ_1 objective causes the descent path to change direction sharply due to the non-smooth, diamond-shaped optimization landscape.
- **Oscillation During Convergence:** The gradient descent algorithm often oscillates between two values as it converges, due to overshooting and correction in the non-smooth ℓ_1 objective.

3. For ℓ1 objective, a high learning rate (η = 1) causes instability, with the norm remaining high even as iterations increase. Moderate learning rates (η = 0.1) and (η = 0.01) lead to stable and quick convergence, while a low learning rate (η = 0.001) results in slow but steady convergence over many iterations.

Similarly, for ℓ2 objective, a high learning rate (η = 1) also causes instability. Moderate learning rates (η = 0.1) and (η = 0.01) result in rapid convergence, whereas a low learning rate (η = 0.001) leads to slower convergence requiring more iterations. Thus, selecting an appropriate learning rate is crucial for efficient and stable optimization as the number of iterations increases.

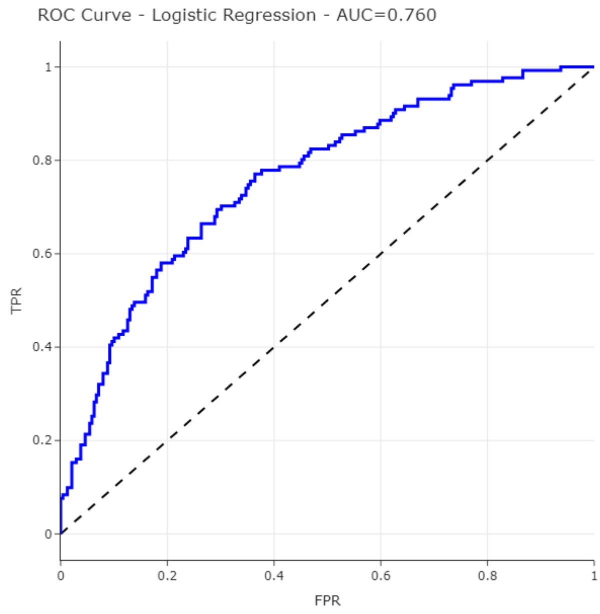




4. The obtained loss when minimizing the ℓ_1 objective was approximately 8.11962×e-3, while for the ℓ_2 objective, it was approximately 2.19211242×e-9. The significant difference in these values is due to the fixed norm of the ℓ_1 gradient, which leads to less efficient convergence with a fixed step size. In contrast, the ℓ_2 gradient decreases with the size of the weights, allowing for finer adjustments and resulting in a much lower final loss.

2.2. <u>Minimizing Regularized Logistic Regression</u>

5. The ROC curve for the logistic regression model shows an AUC of 0.760, indicating good performance in distinguishing between positive and negative instances. The curve lies above the diagonal, demonstrating that the model performs better than random guessing, with a higher true positive rate at lower false positive rates.



6. The value of α that achieves the optimal ROC value according to the criterion

$\alpha^* = \arg\max_\alpha \{TPR\alpha - FPR\alpha\}$ is approximately 0.325. Using this value of α*, the model's test error is approximately 0.337.

7. The optimal λ selected for the ℓ1-regularized logistic regression, based on cross-validation, is 0.02. Using this λ and α = 0.5, the model's test error is approximately 0.2826.