Report Earthquake Alert Prediction Using Optimization Methods

1 Introduction

In this project, we worked on an *Earthquake Alert Prediction* dataset and applied different optimization techniques to predict earthquake alert levels. The main objective of this work was to study and compare the behavior of unconstrained and constrained optimization methods through real data.

The dataset contains the following features:

- Magnitude
- Depth
- CDI (Community Determined Intensity)
- MMI (Modified Mercalli Intensity)

The target variable is the alert level classified as `green`, `yellow`, `orange`, and `red`. These categories were encoded numerically as $0, 1, 2, 3$ and further scaled to the interval $[0, 1]$ for regression-based optimization.

All input features were normalized using standardization (zero mean and unit variance), and a bias term was added to improve the learning capability of the models.

2 Objective Function

For the unconstrained optimization case, we minimized the regularized mean squared error given by:

$$f(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^\top x_i)^2 + \lambda\|w\|^2,$$

where $\lambda = 0.01$ is the regularization parameter.

For the constrained optimization case, only the mean squared error was minimized:

$$f(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^\top x_i)^2,$$

subject to the constraints:

$$g(w) = \|w\|^2 - 1 \leq 0, \qquad h(w) = \sum_i w_i = 0.$$

3 Unconstrained Optimization

3.1 Gradient Descent

For Gradient Descent, we used:

- Step size (learning rate): $\alpha = 0.1$
- Stopping criterion: $\|\nabla f(w)\| < 10^{-6}$
- Maximum iterations: 100

The update rule is:

$$w_{k+1} = w_k - \alpha \nabla f(w_k).$$

From the convergence plot, the objective value decreases smoothly and the gradient norm decays steadily. This indicates stable but gradual convergence, which is characteristic of Gradient Descent.

3.2 Newton's Method

Newton's Method uses second-order derivative information through the Hessian matrix:

$$w_{k+1} = w_k - H^{-1} \nabla f(w_k).$$

Newton's Method converged extremely fast (within 1–2 iterations), which demonstrates the advantage of second-order methods in terms of speed and efficiency, although at higher computational cost.

4 Constrained Optimization Using the Penalty Method

In the constrained formulation of our optimization problem, the goal is to minimize the prediction error while simultaneously enforcing both an inequality and an equality constraint on the model parameters. The constrained optimization problem is mathematically defined as:

$$\min_{w} \quad f(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - w^\top x_i)^2$$

subject to the constraints:

$$g(w) = \|w\|^2 - 1 \leq 0$$

$$h(w) = \sum_{i=1}^{d} w_i = 0$$

Here, the inequality constraint $g(w)$ ensures that the squared Euclidean norm of the weight vector remains bounded by 1. This prevents excessively large weights and improves numerical stability. The equality constraint $h(w)$ enforces that the sum of all feature weights (excluding the bias term) is zero, which imposes a balance condition among the contributing features.

To solve this constrained problem using unconstrained optimization techniques, we employed the **Quadratic Penalty Method**. The constrained problem is transformed into an unconstrained one using the penalized objective function:

$$P(w, \rho) = f(w) + \rho \max(0, g(w))^2 + \rho h(w)^2$$

where $\rho > 0$ is the penalty parameter that controls the strictness of constraint enforcement.

The algorithm was implemented using a **two-level iterative structure**:

- Initial penalty parameter: $\rho_0 = 0.1$
- Penalty update rule: $\rho_{k+1} = 1.5\,\rho_k$
- Number of outer iterations: 15
- Number of inner iterations per outer loop: 20

At each outer iteration, the parameter vector w is optimized using gradient-based updates on the function $P(w, \rho_k)$, while keeping the penalty parameter fixed. After completing the inner updates, the value of ρ is increased so that constraint violations are penalized more severely in the next iteration.

During the early outer iterations, the penalty coefficient is small, and the optimizer focuses mainly on minimizing the prediction error. As ρ increases, the contribution of the penalty terms becomes dominant, and the optimizer strongly enforces feasibility with respect to both constraints.

It is theoretically expected that in later iterations, the penalized objective function may show a slight increase. This happens because even very small violations of the constraints get amplified by the large penalty parameter. Hence, a small rise in the penalized objective near the end does not indicate divergence but confirms that the optimizer is aggressively enforcing the constraints.

Image Output:
This is the image output we generated from our implementation, showing the convergence behavior of Gradient Descent, Newton's Method, and the Penalty Method, along with the comparison of feature weights across all three optimization approaches.

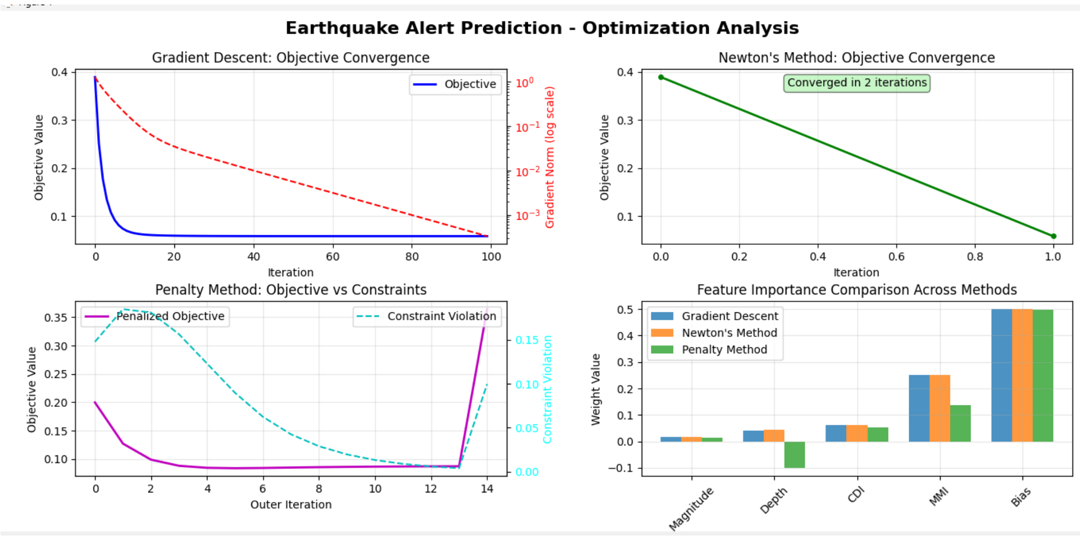

Figure 1: Optimization analysis showing convergence and feature importance comparison across methods.

5 Feature Importance Analysis

From the learned weights obtained using Gradient Descent, Newton's Method, and the Penalty Method, we performed a comparative feature importance analysis. The following observations were made:

- **MMI (Modified Mercalli Intensity)** consistently obtained the highest weight across all optimization methods, identifying it as the most influential predictor of earthquake alert severity.
- **Magnitude** and **CDI (Community Determined Intensity)** also showed significant contribution toward alert prediction.
- **Depth** was observed to have relatively lower influence compared to the other input features.

The consistency of feature importance across all three optimization techniques demonstrates the stability and reliability of the trained model and confirms that the physical characteristics of earthquakes strongly influence alert severity.

6 Conclusion

In this project, we successfully applied both unconstrained and constrained optimization techniques to an Earthquake Alert Prediction problem. Gradient Descent demonstrated stable and steady convergence, making it suitable for large-scale problems. Newton's Method achieved extremely rapid convergence due to the use of second-order derivative information, at the cost of increased computational complexity.

The Penalty Method enabled us to solve a constrained optimization problem by transforming it into a sequence of unconstrained problems. The method clearly demonstrated the trade-off between minimizing prediction error and enforcing constraint feasibility. The gradual increase of the penalty parameter ensured a smooth transition from unconstrained optimization to strict constraint satisfaction.

The feature importance analysis revealed that Modified Mercalli Intensity (MMI) is the most important determinant of earthquake alert severity, followed by magnitude and CDI, while depth has comparatively lower impact.

Overall, this study highlights the effectiveness of numerical optimization methods in solving real-world prediction problems and provides valuable insights into their convergence behavior, stability, and practical performance.

Done by:
Akshaya – BT2024215
Geethika – BT2024139