Logistic Regression with Gradient Descent: Capstone Project Report

Cheera Uday Kiran

January 18, 2025

Abstract

This report presents the implementation of logistic regression using gradient descent with backtracking line search and Leave-One-Out Cross-Validation (LOOCV) for tuning the regularization parameter λ. The model is trained and tested on the MNIST dataset, specifically for classifying digits 3 and 8. The results demonstrate the effectiveness of the approach, and potential improvements are discussed.

1 Introduction

Logistic regression is a widely used classification algorithm that models the probability of a binary outcome. In this project, we apply gradient descent to solve the optimization problem arising in logistic regression. Key components include:

- Gradient descent with backtracking line search for efficient optimization.
- Leave-One-Out Cross-Validation (LOOCV) for selecting the regularization parameter λ.
- Evaluation of the model on the MNIST dataset.

2 Algorithm Description

2.1 Logistic Regression

Given a training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \{-1, +1\}$, logistic regression solves the optimization problem:

$$\min_{\beta \in \mathbb{R}^{n+1}} F(\beta),$$

where

$$F(\beta) = \sum_{i=1}^{N} \ln\left(1 + e^{-y_i(\mathbf{a}^T \mathbf{x}_i + b)}\right) + \lambda \|\mathbf{a}\|_2^2.$$

Here, $\beta = [\mathbf{a}; b]$ includes the weights $\mathbf{a}$ and bias b, and $\lambda > 0$ is the regularization parameter.

2.2 Gradient Descent

Gradient descent is used to minimize $F(\beta)$. The update rule is:

$$\beta_{k+1} = \beta_k - \alpha_k \nabla F(\beta_k),$$

where α_k is the step size determined by backtracking line search.

2.3 Backtracking Line Search

Backtracking line search ensures sufficient decrease in the objective function. The step size α_k is chosen such that:

$$F(\beta_k - \alpha_k \nabla F(\beta_k)) \leq F(\beta_k) - c\alpha_k \|\nabla F(\beta_k)\|^2,$$

where $c \in (0, 1)$ is a constant.

2.4 Leave-One-Out Cross-Validation (LOOCV)

LOOCV is used to select the optimal λ. For each candidate λ, the model is trained on $N - 1$ samples and validated on the remaining sample. The λ with the lowest average validation error is selected.

3 Experimental Setup

3.1 Dataset

The MNIST dataset is used, with 1000 examples each of digits 3 and 8. The labels are relabeled as -1 and $+1$, respectively. The dataset is split into training and test sets.

3.2 Preprocessing

The features are normalized by subtracting the mean and dividing by the standard deviation. A bias term is added to the feature matrix.

3.3 Implementation Details

- The logistic loss function and its gradient are implemented.
- Gradient descent with backtracking line search is used for optimization.
- LOOCV is performed to select λ from $\{0.001, 0.01, 0.1, 1.0, 10.0\}$.
- The model is evaluated on the test set using the success rate.

4 Results

4.1 Optimal λ

The optimal λ selected by LOOCV is $\lambda = 0.01$.

4.2 Success Rate

The success rate on the test set is 94.71%, demonstrating the effectiveness of the model.

5 Discussion

The results show that logistic regression with gradient descent and LOOCV performs well on the MNIST dataset. Potential improvements include:

- Using stochastic gradient descent for faster convergence.
- Exploring advanced regularization techniques like elastic net.
- Incorporating feature engineering or dimensionality reduction.

6 Conclusion

This project successfully implements logistic regression with gradient descent and LOOCV for binary classification on the MNIST dataset. The results are promising, and future work can explore advanced optimization and regularization techniques.

References

- Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Backtracking Line Search - Wikipedia