Performance Analysis of the Elastic Net Regularization Method in High-Dimensional Regression Models

Thierry Kubwimana

1. Introduction

As machine learning and data analytics become increasingly integral in both academic research and industry applications, the ability to extract meaningful insights from high-dimensional datasets has become crucial. High-dimensional data refers to datasets with a large number of features (predictors), often surpassing the number of observations. While these datasets offer a wealth of information, they also present unique challenges, particularly multicollinearity and overfitting. Regularization methods have emerged as essential tools for addressing these issues.

This report focuses on **Elastic Net regularization**, a method that combines the benefits of both **Lasso (L1)** and **Ridge (L2)** regression. We evaluate its performance under various simulated conditions, analyze its behavior with different values of the mixing parameter α, and conduct a Monte Carlo simulation to understand its robustness across a range of correlation scenarios.

---

2. Background and Objective

Regularization techniques aim to constrain or shrink the coefficient estimates in a linear regression model to prevent overfitting and enhance generalizability.

- **Lasso Regression (L1):** Encourages sparsity, potentially setting some coefficients to zero.

- **Ridge Regression (L2):** Shrinks coefficients toward zero but does not eliminate them, useful for multicollinearity.

Elastic Net combines both penalties using a mixing parameter α:

- $\alpha = 0$: Pure Ridge
- $\alpha = 1$: Pure Lasso
- $0 < \alpha < 1$: A blend of both

Objective: To assess how Elastic Net performs with varying dimensions (p), correlation levels (ρ), and values of α. We evaluate predictive accuracy, feature selection, and robustness.

---

3. Methodology

3.1 Simulation-Based Case Study

Model Used: $Y = \beta_0 + X\beta + \varepsilon$

```r
# Define positive definite covariance matrix
rho <- 0.8
Sigma <- matrix(c(1, rho, 0,
                  rho, 1, 0,
                  0, rho, 1), nrow = 3, byrow = TRUE)

# Validate and generate data
if (all(eigen(Sigma)$values > 0)) {
  L <- chol(Sigma)
  X <- matrix(rnorm(300), nrow = 100) %*% L
  Y <- X %*% c(2, 1, 3) + rnorm(100)
} else {
  stop("Covariance matrix is not positive definite.")
}
```

Small Dataset ($p = 3$):

- Generated X from a multivariate normal distribution.
- Two predictors are strongly correlated; the third is independent.
- Used Cholesky decomposition for generation.
- Ran Elastic Net regression for $\alpha = 0.1,\ 0.5,\ 0.9$.
- For each α, used cross-validation to select λ and computed:
  - Squared error: $\|Y - \hat{\beta}_0 - X\hat{\beta}_{EN}\|^2$
  - Number of non-zero coefficients (q)

Large Dataset ($p = 10$):

- Mixed correlated and uncorrelated variables.
- Repeated the same process as above.

```r
alphas <- c(0.1, 0.5, 0.9)
results_small <- run_elastic_net(X, Y, alphas)
print(results_small)
```

3.2 Monte Carlo Simulation

- Fixed $p = 3$.
- Defined covariance matrix Σ:

$$\Sigma = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & \rho & 1 \end{bmatrix}$$

- Vary ρ over 19 values: $\{-0.9, \ldots, 0.9\}$
- For each ρ and $\alpha \in \{0.1,\ 0.5,\ 0.9\}$:
  - Run Elastic Net 100 times
  - Record mean squared error (MSE) into a 19×3 matrix
- Plot MSE vs ρ to assess sensitivity

```r
mse_matrix <- monte_carlo_sim(3, 100, rho_values, alphas, beta_small, reps = 100)
mse_df <- melt(mse_matrix)
```

---

4. Results and Discussion

4.1 Case Study Performance

Results Summary Table (Small Dataset):

Table 1: Elastic Net Performance on Small Dataset ($p = 3$)

α	λ	MSE	Selected Predictors
0.1	0.0266	1.188793	4
0.5	0.0215	1.179291	4
0.9	0.0184	1.269211	4

Large Dataset ($p = 10$):

Table 2: Elastic Net Performance on Large Dataset ($p = 10$)

α	λ	MSE	Selected Predictors
0.1	0.2548	1.419199	10
0.5	0.0891	1.377309	10
0.9	0.0173	1.482455	4

Observations:

- Best performance (lowest MSE) at $\alpha = 0.5$.
- Elastic Net retains relevant predictors and handles correlation effectively.

4.2 Monte Carlo Simulation

Trends:

- For $\rho \approx 0$: best performance at $\alpha = 0.9$ (Lasso-like)

- For $\rho \approx \pm 0.9$: $\alpha = 0.5$ performs best
- For $\rho \approx -0.9$: $\alpha = 0.1$ (Ridge-like) is more stable

These results confirm that Elastic Net's adaptive blending of L1 and L2 penalties helps it maintain robustness under varying levels of predictor correlation.

---

5. Conclusion

Elastic Net effectively balances variable selection and coefficient shrinkage, making it well-suited to high-dimensional data with correlated features.

- Outperforms Lasso and Ridge in correlated settings
- $\alpha = 0.5$ generally achieves best overall results
- Monte Carlo simulations confirm adaptability across correlation structures

Elastic Net is a flexible, powerful tool for regression in complex data environments.

---

6. Appendix

Complete R code for data generation, model fitting, and visualization is available in the accompanying files.