John Benedict M. Hornilla

BSCS 3-B

Regularization – Ridge vs Lasso Regression

In machine learning, the development of predictive models often encounters the challenge of overfitting, where a model becomes too closely aligned with the training data and fails to generalize to new datasets. Regularization is a statistical technique used to mitigate this issue by introducing a penalty to the model's complexity. This report details two primary forms of regularization: Ridge Regression and Lasso Regression.

Ridge Regression, also known as L2 Regularization, is designed to reduce errors by correcting multicollinear variables within a regression analysis. Multicollinearity occurs when independent variables are highly correlated, leading to unreliable coefficient estimates. Lasso Regression stands for Least Absolute Shrinkage and Selection Operator. Like Ridge, it is a technique used to correct overfitted training data, but it introduces a distinct penalty structure known as L1 Regularization.

The accuracy of a model is measured by the Residual Sum of Squared (RSS), which calculates the difference between the observed and predicted values. Ridge Regression modifies this by adding a penalty term at the end of the RSS function. This penalty, known as the L2 penalty, is the sum of the squares of the coefficients. The impact of the penalty is controlled by the hyperparameter lambda. If lambda is zero, the model performs standard OLS regression without regularization. As lambda increases, the penalty forces the high-value coefficients to shrink, reducing the model's variance at the cost of some bias.

The primary advantage of Lasso is its ability to handle high-dimensional data by performing automatic feature selection. While Ridge shrinks coefficients towards zero, Lasso can shrink less important coefficients to exactly zero. This effectively removes irrelevant variables from the model, making it highly interpretable. Before applying Lasso, it is critical to analyze the dataset for missing values, high feature counts, and correlated independent variables to prevent further overfitting. The goal is to choose a lambda value that minimizes the Mean Squared Error (MSE) while maintaining optimal model complexity.

The differences between the two methods to guide selection based on data characteristics; Shrinkage: Ridge (L2) does not shrink coefficients to absolute zero, whereas Lasso (L1) does. Variable Retention: Ridge does not perform feature selection and retains all variables; Lasso performs feature selection and can drop entire variables. Correlation: Ridge is best used when most features are correlated; Lasso is superior when only a few features are significant.

The utility of regularization is demonstrated through various case studies and research papers. In logistics, Ridge Regression can analyze supply chain datasets where delivery distances are predicted based on inventory size. The model ensures that long-distance deliveries are appropriately weighted against package volume without overfitting the noisy delivery data. Lasso is effective in predicting student exam scores. By applying L1 regularization, a model can identify "Study Hours" as a significant predictor while automatically zeroing out "noise" variables such as a student's physical height or gender, which have no impact on academic performance. Regularization is vital in genetic studies, such as research into riboflavin (Vitamin B2) production. Ridge estimators allow researchers to handle datasets where thousands of genes (features) are present, providing stable predictions even with small sample sizes. Advanced models combining Lasso with "Relief" feature selection techniques have been used to predict heart disease. These hybrid models achieve up to 99.05% accuracy by filtering through medical attributes like age, cholesterol, and blood sugar to isolate the most critical indicators for diagnosis.

In conclusion, regularization is a fundamental component of modern machine learning that balances the bias-variance tradeoff to produce robust models. While Ridge Regression provides stability in the presence of correlated variables, Lasso Regression offers a streamlined, efficient approach through automated feature selection. Choosing the correct method depends on whether the researcher values the inclusion of all data (Ridge) or the identification of a few key drivers (Lasso).

References:

Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies.

Ghosh, P., et al. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques.

IBM. (2025). Lasso Regression: What is Lasso Regression?

Murel, J., & Kavlakoglu, E. (2025). Ridge Regression: What is ridge regression?