Predicting Student Performance

Linear Models (PHP2601), Prof. Ani Eloyan

Daniel Posmik, Jizhou Tian, Aristofanis Rontogiannis

2024-12-10

Table of contents I

EDA and the Linear Model

Introduction

We will be analyzing educational data to understand the predictors of student performance. Specifically, we seek to **understand whether five predictors – as a subset of an exhaustive list of potential predictors – are significant predictors of student performance**.

Testing the significant of a subset of predictors is becoming increasingly important in modern statistical questions, especially with more information becoming available.

We will be using a publicly available dataset from Kaggle that contains information about students and their exam scores.

Hypothesis to be Tested

We are interested in:

- Hours Studied
- Attendance
- Sleep Hours
- Previous Scores
- Tutoring Sessions

Elastic Net

Why Use Elastic Net?

- **Limitations of Lasso**: May select only one variable from a group of highly correlated predictors.
- **Limitations of Ridge**: Cannot produce sparse models (i.e., no feature selection).
- **Elastic Net Advantage**:
  - Encourages group selection.
  - Balances sparsity and multicollinearity handling.

Elastic Net Formula

Elastic Net adds two penalty terms:

$$\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right)$$

- $\|\beta\|_1$: Lasso penalty (L1).
- $\|\beta\|_2^2$: Ridge penalty (L2).
- λ_1, λ_2: Regularization parameters.

Tuning Parameters in Elastic Net

1. α: Controls the mix between Ridge and Lasso.
   - $\alpha = 0$: Ridge.
   - $\alpha = 1$: Lasso.
   - $0 < \alpha < 1$: Elastic Net.
2. λ: Controls the overall strength of regularization.

Grid Search:

- Perform cross-validation to find optimal values of α and λ.