Week 5 Learning Reflection

Summary:

This week, we moved beyond traditional linear regression and focused on linear model selection and regularization techniques to enhance model performance and interpretability. We studied various selection methods like best subset selection, forward and backward stepwise selection, and explored shrinkage methods like ridge regression and the lasso. These tools help in choosing a more effective model by reducing variance, improving prediction accuracy, and simplifying interpretation by eliminating irrelevant predictors.

Concepts:

- Best Subset Selection: Evaluates all possible combinations of predictors and selects the model with the best performance based on metrics like RSS, R2R^2R2, Cp, AIC, BIC, or cross-validation error.

- Forward Stepwise Selection: Begins with no variables and adds one predictor at a time, choosing the one that improves the model the most at each step.

- Backward Stepwise Selection: Starts with all predictors and removes the least useful one at each step.

- Model Selection Criteria:

    - Cp, AIC, BIC: Penalize model complexity to avoid overfitting.

    - Adjusted R2R^2R2: Modifies R2R^2R2 to account for the number of predictors.

    - Cross-Validation: A more robust approach to estimate model performance.

- Shrinkage Methods:

    - Ridge Regression: Shrinks coefficient estimates by adding an L2L_2L2 penalty, helpful when predictors are correlated or p>np > np>n.

    - Lasso (Least Absolute Shrinkage and Selection Operator): Adds an L1L_1L1 penalty, which can shrink some coefficients exactly to zero, effectively performing variable selection.

- Bias-Variance Tradeoff: Regularization introduces bias to reduce variance, which can lower the overall test error and improve model generalization.

Uncertainties:

I’m still unsure about when to prefer ridge regression over lasso, especially when dealing with multicollinearity or high-dimensional data. How do I determine which regularization method will perform better in a given scenario?