Advanced Statistical Methods
Homework 6

Brandon Hosley

University of Illinois - Springfield

November 3, 2020

Overview

1. Q1A: Mallow's C_P
2. Q1A: Akaike Information Criterion
3. Q1A: Bayesian Information Criterion
4. Q1A: Adjusted R^2
5. Q1B: Elastic Net Regularization
6. Q2: Hastie and Tibshirani Summary

Mallow's C_P

$$C_p = \frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2)$$

- Error is multiplied by number of predictors
- Biasing against models with larger numbers of predictors
- This method seeks to address over-fitting.

Akaike Information Criterion

$$\mathrm{AIC} = -2 \log L + 2 \cdot d$$

- This method emphasizes maximum likelihood
- Also biases against increased numbers of parameters; but in this case the bias increases strictly linearly

Bayesian Information Criterion

$$\text{BIC} = \frac{1}{n}(\text{RSS} + \log(n) d \hat{\sigma}^2)$$

- Similar to Mallows C_P
- Bias is higher for models using more than 7 predictors

Adjusted R^2

$$\text{Adjusted}\, R^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$$

- Unlike previous metrics, higher is better
- Similarly biases against models with more predictors
- Linear increase in denominator

Elastic Net Regularization

$$\hat{\beta} \equiv \text{argmin}_{\beta}(\|y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1)$$

- Improvement on LASSO and Ridge regression methods
- Adds a quadratic penalty, creating a unique minimum
- Biases towards the fewest number of good predictors

Hastie and Tibshirani Lecture: Model Selection

Possible approaches:

Subset Selection Perform regression on each subset of predictors; select model with best performance.

Forward/Backward stepwise selection builds the model one predictor at a time; using single step optimization.

Shrinkage Fitting a model containing all predictors; reducing coefficients to fit (Some to zero)

Dimension Reduction Combining original predictors together to reduce the total number of predictors that the model will fit to