Similarities between Nonnested Regression Estimation and Ridge Regression

Caleb Leedy

May 4, 2024

1 Nonnested Regression

In the case of non-nested two phase sampling, we have $A_1 = (X_i)_{i=1}^{n_1}$ and $A_2 = (X_1, Y_i)_{i=1}^{n_2}$ with A_1 and A_2 being selected independently from the same sampling frame. The regression estimator is then

$$\begin{aligned}
\hat{\hat{Y}}_{reg} &= \hat{\hat{Y}}_{HT} + (\hat{\hat{X}}_c - \hat{\hat{X}}_2)'\hat{\beta}_2 && \text{for } \hat{\beta}_2 = \left(\sum_{i \in A_2} x_i x_i'\right)^{-1} \sum_{i \in A2} x_i y_i \text{ with } x_{1i} = 1 \\
&= \hat{\hat{Y}}_{HT} + (\hat{\hat{X}}_1 - \hat{\hat{X}}_2)'W'\hat{\beta}_2 && \text{if } \hat{\hat{X}}_c = W\hat{\hat{X}}_1 + (I - W)\hat{\hat{X}}_2 \\
&= \sum_{i \in A_1} x_i'W'\hat{\beta}_2 + \sum_{i \in A_2} (y_i - x_i'W'\hat{\beta}_2).
\end{aligned}$$

While the matrix W controls the interaction between A_1 and A_2 it also plays the role of shrinkage on $\hat{\beta}_2$.

2 Ridge Regression

Given a sample A, ridge regression solves the optimization problem,

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i \in A} (y_i - x_i'\beta)^2 + \beta^T(\lambda I_p)\beta.$$

Differentiating with respect to β and setting this equal to zero yields a solution,

$$\hat{\beta} = \left(\sum_{i\in A} x_i x_i' + \lambda I_p\right)^{-1} \sum_{i\in A} x_i' y_i = (X'X + \lambda I_p)^{-1} X'Y.$$

Let $\hat{\beta}_{OLS} = \hat{\beta}_2 = (X'X)^{-1}X'Y$, then using the Sherman-Morrison-Woodbury inverse formula,

$$\hat{\beta} = (I_p - (X'X)^{-1}(\lambda^{-1} I_p + (X'X)^{-1}))^{-1} \hat{\beta}_{OLS}.$$

3 Discussion

The previous two sections suggest that if we let

$$\begin{aligned} W &= (I_p - (X'X)^{-1}(\lambda^{-1} I_p + (X'X)^{-1}))^{-1} \\ \iff \lambda W &= (X'X)(I_p - W) \\ \iff \lambda &= (X'X)(W^{-1} - I_p) \end{aligned}$$

then we would have equivalent results. In this way, non-nested two phase sampling is like ridge regression.