Lecture Summary: March 3, 2023

• Adjusted R^2

$$R^2 = 1 - \frac{\text{SSE}}{\text{SSTO}}$$

cannot be directly used in variable selection, because, if so, "larger model always wins".

Some modification of R^2 is needed. The modification is

$$R_{\text{a}}^2 = 1 - \frac{\text{MSE}}{\text{MSTO}},$$

where MSTO = SSTO/$(n-1)$ [and MSE = SSE/$(n-p)$].

Select the model that maximizes R_{a}^2, or, equivalently, the model that minimizes MSE.

• Mallow's C_p

Suppose that $x_1, \ldots, x_{K-1}$ are all the candidate predictors.

For any subset of predictors $x_{i_1}, \ldots, x_{i_{p-1}} \in \{x_1, \ldots, x_{K-1}\}$, define

$$C_p = C_p(x_{i_1}, \ldots, x_{i_{p-1}}) = \frac{\text{SSE}(x_{i_1}, \ldots, x_{i_{p-1}})}{\text{MSE}(x_1, \ldots, x_{K-1})} - (n - 2p),$$

where $\text{SSE}(x_{i_1}, \ldots, x_{i_{p-1}})$ = SSE of fitting the regression with $x_{i_1}, \ldots, x_{i_{p-1}}$ being the predictors, and $\text{MSE}(x_1, \ldots, x_{K-1})$ = MSE of fitting the regression with all of the candidate predictors.

The best subset of predictors (model) corresponds to the one such that C_p is small and close to p.

Note: C_K is always equal to K.

• AIC and BIC (SBC) criteria

$$\begin{aligned} \text{AIC}_p &= n \log(\text{SSE}_p/n) + 2p, \\ \text{SBC}_p &= n \log(\text{SSE}_p/n) + (\log n)p. \end{aligned}$$

Choose a subset of predictors (model) that minimizes AIC_p (SBC_p).

- Forward stepwise selection:

1. Choose the first predictor, $x_{(1)}$, that has the largest $|t|$ for the slope under a simple linear regression with the predictor.