STAT 200B 2019 Week13

Soyeon Ahn

1 Model selection overview

- Variable selection
  - All possible subsets: compare all possible combinations. It is computationally intensive and may result in overfitting with multiple testing. RSS and R^2 natually improves as the number of input variable increases. We want to minimize the test error, rather than the training error.
  - Forward selection: from intercept-only to full model. Although there is no absolute stopping rule for inclusion or exclusion of predictors, usually the p-value (say <0.05), AIC, or BIC.
  - Backward selection: from full model to intercept-only model, elimination of the least significant candidate predictor. A backward selection is preferred in general, since it allows a modeller to judge the effects of all candidate predictors simultaneously. It is also able to include correlated variables, which might not be entered in a forward selection.
  - Regularization based selection: lasso, elastic-net.
- Model selection
  - Hypothesis testing: A likelihood ratio test can be performed by adding a single variable to the current model (nested models). It is straightforward but is not applicable for different model specifications.
  - Information criteria: Both AIC and BIC consider fit to the data, but penalize for the complexity of the model
- Model validation and diagnostics
  - Overall performance: Mallow's C_p, adjusted R^2, Brier
  - Calibration: Calibration-in-the-large, Hosmer-Lemeshow test
  - Discrimination, accuracy, precision, recall, and F1
  - Other measures: reclassification (reclassification index, integrated discrimination index), usefulness (decision curve, or net benefit),

2 Variable selection: topic-based

- In linear regression, adding predictors always decreases the training error or RSS.

minimizing the residual sum of squares

$$RSS = \sum_{i=1}^{n} \hat{\epsilon}_i^2$$

where $\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$

In the simple linear regression model fitting,

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right),$$

where $\hat{\beta}_0$ and $\hat{\beta}$ are least squares estimates, which minimize

$$f(\beta_0, \beta) = \sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 X_i)\right]^2.$$

The first-order partial derivatives of $f(\beta_0, \beta)$ are

$$\frac{\partial f(\beta_0, \beta)}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 X_i)\right];$$
$$\frac{\partial f(\beta_0, \beta)}{\partial \beta} = -2 \sum_{i=1}^{n} X_i \left[Y_i - (\beta_0 + \beta_1 X_i)\right].$$

$\hat{\beta}_0$ and $\hat{\beta}$ should satisfy that

$$\frac{\partial f(\hat{\beta}_0, \hat{\beta})}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)\right] = 0,$$
$$\frac{\partial f(\hat{\beta}_0, \hat{\beta})}{\partial \beta} = -2 \sum_{i=1}^{n} X_i \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)\right] = 0$$

Thus,

$$\bar{Y} - \left(\hat{\beta}_0 + \hat{\beta}_1 \bar{X}\right) = 0$$
$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

The least squares estimates are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

2.1 Expectation and Variance

Rewrite the equation as follows:

$$Y_i = \beta_0 + \beta_1 \bar{X} + \beta_1 (X_i - \bar{X}) + \epsilon_i \tag{1}$$

and the least squares estimates are

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} - \frac{\sum_{i=1}^n (X_i - \bar{X})\bar{Y}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \sum_{i=1}^n w_i Y_i
\end{aligned}$$

where $w_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$ Note that $\sum_{i=1}^n w_i = 0$ and $\sum_{i=1}^n w_i (X_i - \bar{X}) = 1$.

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= \sum_{i=1}^n \frac{1}{n} Y_i - (\sum_{i=1}^n w_i Y_i) \bar{X} \\
&= \sum_{i=1}^n \left(\frac{1}{n} - w_i \bar{X} \right) Y_i \\
&= \sum_{i=1}^n \left(\frac{1}{n} - \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) Y_i
\end{aligned}$$

$$
\begin{aligned}
E[\hat{\beta}_1] &= E[\sum_{i=1}^{n} w_i Y_i] \\
&= \sum_{i=1}^{n} w_i E[Y_i] \\
&= \sum_{i=1}^{n} \frac{(X_i - \bar{X})(\beta_0 + \beta_1 \bar{X} + \beta_1 (X_i - \bar{X}))}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
&= \beta_1 \\
V[\hat{\beta}_1] &= V[\sum_{i=1}^{n} w_i Y_i] \\
&= \sum_{i=1}^{n} w_i^2 \sigma^2 \\
&= \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right)^2 \sigma^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sigma^2}{n s_X^2}
\end{aligned}
$$

$$
\begin{aligned}
E[\hat{\beta}_0] &= E[\bar{Y} - \hat{\beta}_1 \bar{X}] \\
&= \frac{1}{n} E[\beta_0 + \beta_1 X_i] - \beta_1 \bar{X} \\
&= \beta_0 \\
V[\hat{\beta}_0] &= V[\sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) Y_i] \\
&= \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right)^2 \sigma^2 \\
&= \left(\sum_{i=1}^{n} \frac{1}{n^2} - \sum_{i=1}^{n} \frac{2}{n} \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2 \bar{X}^2}{(\sum_{i=1}^{n}(X_i - \bar{X})^2)^2} \right) \sigma^2 \\
&= \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) \sigma^2 \\
&= \left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) \sigma^2 = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{n s_X^2} \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
Cov[\hat{\beta}_0, \hat{\beta}_1] &= Cov[\sum_{i=1}^{n} w_i Y_i, \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) Y_i] \\
&= \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \times \big(\frac{1}{n} - \frac{(X_i - \bar{X})\bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \big) \right) \sigma^2 \\
&= \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X})}{s_X^2} \times \big(1 - \frac{(X_i - \bar{X})\bar{X}}{s_X^2} \big) \right) \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n s_X^2} (-\bar{X})
\end{aligned}
$$