SDS 383D Multiple Regression

Jan-Michael Cabrera, JC7858

February 11, 2019

A: Beta Estimator

Least squares

Least squares involves finding the estimator $\hat{\beta}$ that minimizes the sum of squared errors.

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{R}^p} \left\{ \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \right\} \tag{1}$$

The term within the parenthesis can be rewritten in vector notation as

$$f(\beta) = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 = (y - x\beta)^T (y - x\beta). \tag{2}$$

This function must then be differentiated and the resultant function set to zero to find the estimator $\hat{\beta}$.

$$\frac{\partial}{\partial \beta} f(\beta) = 0 \tag{3}$$

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} \left((y - x\beta)^T (y - x\beta) \right) &(4)\\ &= -x^T (y - x\beta) &(5)\\ &= -x^T y + x^T x \beta &(6) \end{aligned}$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \tag{7}$$

Maximum Likelihood

Maximum likelihood involves finding the estimator $\hat{\beta}$ that maximizes the likelihood function for a given model. Our model is of the form $y_i = x_i^T \beta + \epsilon_i$.

$$\hat{\beta} = \arg\max_{\beta \in \mathcal{R}^p} \left\{ \prod_{i=1}^{n} p(y_i \mid \beta, \sigma^2) \right\} \tag{8}$$

For our model we assume that the errors are mean zero normally distributed samples each with all the same variances, $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$. This gives our error the form $\epsilon_i = y_i - x_i^T\beta$.

The likelihood function is then of the form

$$
\begin{aligned}
f(\beta) &= \prod_i^n p(y_i|\beta, \sigma^2) && (9) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{n}{2\sigma^2}\sum_i^n (y_i - x_i^T\beta)^2\right] && (10)
\end{aligned}
$$

We can take the log of the above and maximize the resulting function because the logarithm is a monotonically increasing function.

$$
\frac{\partial}{\partial\beta}\ln(f(\beta)) = 0 \qquad (11)
$$

Omitting terms that do not depend on β for brevity,

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\beta}\left[-\frac{n}{2\sigma^2}\sum_i^n (y_i - x_i^T\beta)^2 + ...\right] && (12) \\
&= \frac{\partial}{\partial\beta}\left[-\frac{n}{2\sigma^2}(y - x\beta)^T(y - x\beta) + ...\right] && (13) \\
&= -x^T(y - x\beta) && (14) \\
&= -x^Ty + x^Tx\beta && (15)
\end{aligned}
$$

$$
\hat{\beta} = (x^Tx)^{-1}x^Ty \qquad (16)
$$

Method of Moments

Here we want to choose $\hat{\beta}$ so that the sample covariance between the errors and each predictor is exactly zero, $\mathrm{cov}(\epsilon, x_j) = 0$.

$$
\begin{aligned}
0 &= \sum_i^n (\epsilon_i - \bar{\epsilon})(x_{ij} - \bar{x_j}) && (17) \\
&= \sum_i^n \epsilon_i x_{ij} - \bar{x_j}\sum_i^n \epsilon_i - \bar{\epsilon}\sum_i^n x_{ij} + \bar{\epsilon}\bar{x_j} && (18)
\end{aligned}
$$

We observe that $\bar{\epsilon} = \frac{1}{n}\sum_i^n \epsilon_i$ and $\bar{x}_j = \frac{1}{n}\sum_i^n x_{ij}$. We can also center our data such that $\bar{x}_j = 0$.

$$\begin{aligned}
0 &= \sum_{i}^{n} \epsilon_i x_{ij} && (19)\\
&= \epsilon^T x && (20)\\
&= (y - x\beta)^T x && (21)\\
&= y^T x - \beta^T x^T x && (22)\\
&= x^T y - x^T x\beta \quad \text{with} \quad \beta^T x^T x = x^T x\beta \quad \text{and} \quad y^T x = x^T y && (23)\\
& && (24)
\end{aligned}$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \tag{25}$$