SDS 383D Multiple Regression

Jan-Michael Cabrera, JC7858

February 11, 2019

B: Weighted Errors

Least squares

Here we wish to evalue $\hat{\beta}$ for data with idiosyncratic precisions.

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

As in part A, the term within the parenthesis can be rewritten as

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

We differentiate the function and set the derivative to zero to find the estimator that minizes the function.

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

$$\begin{aligned}
0 &= \frac{\partial}{\partial \beta} \left((y - x\beta)^T W (y - x\beta) \right) && (4)\\
&= -x^T (W + W^T)(y - x\beta) && (5)\\
&= -x^T 2W (y - x\beta) && (6)\\
&= -x^T W y + x^T W x \beta && (7)
\end{aligned}$$

We recognize in the above that $\frac{\partial}{\partial x}(x^T A x) = (A + A^T)x$ and because W is a symmetric matrix $(W + W^T) = 2W$

$$\hat{\beta} = (x^T W x)^{-1} x^T W y \tag{8}$$

Maximum Likelihood

Similarly in terms of the maximimum likelihood, we have

$$\hat{\beta} = \arg\max_{\beta \in \mathcal{R}^p} \left\{ \prod_{i=1}^{n} p(y_i \mid \beta, \sigma^2) \right\} \tag{9}$$

Our model as before is $y_i = x_i^T \beta + \epsilon_i$, however each error has it's own idiosyncratic variance, $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2)$. The resultant likelihood is then

$$\begin{aligned}
f(\beta) &= \prod_{i}^{n} p(y_i|\beta, \sigma^2) && (10)\\
&\propto \exp\left[-\frac{1}{2} \sum_{i}^{n} \frac{(y_i - x_i^T \beta)^2}{\sigma_i^2} \right] && (11)\\
&\propto \exp\left[-\frac{1}{2} (y - x\beta)\Sigma^{-1}(y - x\beta) \right] && (12)
\end{aligned}$$

We again take the maximimize the logarithm of the likelihood,

$$\frac{\partial}{\partial \beta} \ln(f(\beta)) = 0. \tag{13}$$

$$\begin{aligned}
0 &= \frac{\partial}{\partial \beta} \left[-\frac{1}{2} (y - x\beta)\Sigma^{-1}(y - x\beta) + ... \right] && (14)\\
&= -x^T(\Sigma^{-1} + (\Sigma^{-1})^T)(y - x\beta) && (15)\\
&= -x^T 2\Sigma^{-1}(y - x\beta) && (16)\\
&= -x^T\Sigma^{-1}y + x^T\Sigma^{-1}x\beta && (17)
\end{aligned}$$

$$\hat{\beta} = (x^T\Sigma^{-1}x)^{-1}x^T\Sigma^{-1}y \tag{18}$$

The result is identical to what was derived above with $w_i = 1/\sigma_i^2$.