Multiple Linear Regression

Matrix Formulation

Math 392

Notation

In the multiple regression framework, we consider not simply the joint distribution of $[Y, X]$, but the join distribution of $[Y, X_1, X_2, \ldots, X_{p-1}]$. In fact, since we are in the regression framework it is the conditional distribution of $[Y|X_1, X_2, \ldots, X_{p-1}]$ that is of interest.

It becomes far easier to express the higher dimension random variables using matrix notation and linear algebra operations.

Let:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & x_{n,1} & x_{n,2} & & x_{n,p-1} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}'$$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

The model

The matrix formula of the multiple linear regression model is

$$Y = X\beta + \epsilon; \quad \epsilon \sim N(0, \sigma^2 I)$$

which has a conditional mean of

$$E(Y|X) = X\beta$$

Estimating $\hat{\beta}$

As before, we estimate the regression coefficients that minimize the residual sum of squares, a loss function that can be written as

$$\begin{aligned} RSS(\beta) &= (Y - X\beta)'(Y - X\beta) && (1) \\ &= Y'Y + (X\beta)'X\beta - Y'X\beta - (X\beta)'Y && (2) \\ &= Y'Y + \beta'(X'X)\beta - 2\beta'X'Y. && (3) \end{aligned}$$

To find argmin$RSS(\beta)$, we differentiate with respect to β, set equal to zero,

$$\frac{\partial RSS}{\partial \beta} = 2(X'X)\beta - 2X'Y = 0$$

Then solve for β.

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Note that this requires that an inverse exists. This expression also leaves us with the following vectors. The *fitted values*

$$\hat{Y} = X\hat{\beta}$$

and the residuals

$$\epsilon = Y - \hat{Y} = Y - X\hat{\beta}.$$

Inference on $\hat{\beta}$

For the following, we assume that $Var(\epsilon) = \sigma^2 I$. We can find the expected value and variance of our vector of estimates as follows.

$$\begin{aligned} E(\hat{\beta}|X) &= E((X'X)^{-1}X'Y|X) && (4) \\ &= (X'X)^{-1}X'E(Y|X) && (5) \\ &= (X'X)^{-1}(X'X)\beta && (6) \\ &= \beta && (7) \end{aligned}$$

$$\begin{aligned} Var(\hat{\beta}|X) &= Var((X'X)^{-1}X'Y|X) && (8) \\ &= (X'X)^{-1}X'Var(Y|X)X(X'X)^{-1} && (9) \\ &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} && (10) \\ &= \sigma^2(X'X)^{-1} && (11) \end{aligned}$$

Note that this second expression is more accurately written as $Cov(\hat{\beta}|X)$, since it's a $p \times p$ matrix with variances down the diagonal and covariances off the diagonal (note this matrix is symmetric). When we're working in the particular case of simple linear regression ($p = 2$), we have

$$\sigma^2(X'X)^{-1} = \frac{\sigma^2}{SS_x}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$