The Matrix Approach to Regression

This document describes a matrix approach to regression. Throughout, we will let $Y$ be an $n \times 1$ vector that is the response variable, $X$ be the $n \times (p + 1)$ matrix of the intercept and $p$ explanatory variables, $\beta$ be the $(p + 1) \times 1$ vector of parameters, and $b = \hat{\beta}$ be the least squares estimate of $\beta$. The model is that $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i$ for $i = 1, 2, \ldots n$ and $\epsilon_i$ is the $i$th error. In matrix notation, the model is $Y = X \beta + \epsilon$, so the vector of errors is $Y - X \beta$. The sum of squared errors is $(Y - X \beta)^T (Y - X \beta)$, because in matrix notation a sum of squares is the product of the transpose of a vector with the vector itself. For example,

$$(Y - X \beta)^T (Y - X \beta) = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}))^2$$

We can reexpress the sum of squared errors by carrying out the multiplication.

$$(Y - X \beta)^T (Y - X \beta) = Y^T Y - (X \beta)^T Y - Y^T X \beta + (X \beta)^T X \beta$$

The transpose of a product is the product of the transposes in reverse order. Each term in the sum is a real number, and hence equal to its transpose. Using these two facts, we can rewrite the sum of squared errors as

$$Y^T Y - 2 \beta^T X^T Y + \beta^T X^T X \beta$$

This expression is minimized by setting $\beta = b$. The solution for $b$ is attainable by calculus.

Begin by taking (partial) derivatives with respect to each $\beta_i$ and then set each derivative to 0. Here, $Y$ and $X$ are constants as far as the calculus is concerned. The partial derivative of $\beta$ with respect to $\beta_i$ is one in the $i$th position and 0 in each other position. Let $e_i$ be the $n \times 1$ column vector with a 1 in the $i$th position and a 0 everywhere else. We will also need to use the analogue of the calculus rule

$$\frac{d}{dx} f(x)g(x) = f(x)g'(x) + f'(x)g(x)$$

Thus,

$$\frac{\partial}{\partial \beta_i} Y^T Y - 2 \beta^T X^T Y + \beta^T X^T X \beta = -2e_i^T X^T Y + e_i^T X^T X \beta + \beta^T X^T X e_i$$

$$= -2e_i^T X^T Y + 2e_i^T X^T X \beta$$

$$= 0$$

By dividing by 2 and moving the first term to the other side of the equation, we find this equation.

$$e_i^T X^T X \beta = e_i^T X^T Y$$

This says that the $i$th element of the $X^T X \beta$ is equal to the $i$th element of $X^T Y$. Since this is true for each $i$, we have the equality

$$X^T X \beta = X^T Y$$

This matrix equation is known as the normal equations. The matrix $X^T X$ is a square, $(p + 1) \times (p + 1)$ matrix. If it is invertible, we can solve for $\beta$ by premultiplying by its inverse, so the least squares estimate is

$$b = (X^T X)^{-1} X^T Y$$