An Exact Formula for Linear Regressions

(This section is very notation intensive; skip over it if you need to)

Now, we introduce the transpose of a matrix, which is what you get by swapping the rows with columns i.e. $a_{ij} \to a_{ji}$. For example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 \\ a & b & c \end{bmatrix}^T = \begin{bmatrix} 1 & a \\ 2 & b \\ 3 & c \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 & 9 \end{bmatrix}$$

Most importantly, for a vector **v**, multiplying the vector's transpose with it gives the sum of the squares of the items in the vector.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \Rightarrow \mathbf{v}^T\mathbf{v} = v_1^2 + v_2^2 + \cdots + v_n^2$$

This means we can show $||\varepsilon||^2$ as $||e||^2 = e_1^{\ 2} + e_2^{\ 2} + ... + e_n^{\ 2} = e^T e$. This gives us a nice shorthand for the magnitude of ε (or rather, the magnitude squared).

Using the fact that transposing and matrix multiplication are distributive over addition (showing that transposing is distributive over addition isn't too hard to prove; I leave it as an exercise for the reader), we have:

$$||\varepsilon||^2 = \varepsilon^T\varepsilon = (y - X\beta)^T(y - X\beta) = (y^T - (X\beta)^T)(y - X\beta) =$$

$$y^T y - y^T(X\beta) - (X\beta)^T y + (X\beta)^T(X\beta)$$

Notice that $w^T v = w \cdot v$. Since the dot product is commutative, the middle two terms in our expansion are equal. To find the minimum of this when varying β, we take the derivative and set it to 0. This becomes:

$$\frac{d}{d\beta}(y^T y - 2(X\beta)^T y + (X\beta)^T(X\beta)) = -2\frac{d}{d\beta}(X\beta)^T y + \frac{d}{d\beta}(X\beta)^T(X\beta) =$$

$$-2X^T y + 2X^T(X\beta) = 0$$

With a bit of algebraic manipulation and because matrix multiplication is associative:

$$2X^T(X\beta) = 2X^T y \implies (X^T X)\beta = X^T y \implies \boxed{\beta = (X^T X)^{-1} X^T y}$$