The Vector Geometry of Linear Models - Part 2

Multiple Regression

For multiple regression, it will be easiest to illustrate vector geometric ideas by using only two covariates, and again excluding the intercept term. Our initial fitted model is:

$$\mathbf{y} = A\mathbf{1}_n + B_1\mathbf{x}_1 + B_2\mathbf{x}_2 + \mathbf{e},$$

and again we subtract the mean $\overline{Y} = A + B_1\overline{x}_1 + B_2\overline{x}_2$ from the fitted model in scalar form to get:

$$\mathbf{y}^* = B_1\mathbf{x}_1^* + B_2\mathbf{x}_2^* + \mathbf{e}$$

using mean-deviation vectors. Some of the concepts that can be explained geometrically now are i) the effect of collinearity, ii) the ANOVA decomposition for multiple regression, iii) multiple correlation as the simple correlation between observed and fitted values and as the cosine of the angle between them, iv) the incremental F test for a new variable, v) the relationship between simple and multiple regression, and vi) the ease of interpretation that occurs when covariates are uncorrelated. Degrees of freedom again are interpreted as the dimensionality of the subspaces to which $\mathbf{y}^*$, $\widehat{\mathbf{y}}^*$, and $\mathbf{e}$ are confined.

The Vector Geometric Approach and Projection Matrices

We have seen that least-squares estimators work by projecting the data vector $\mathbf{y}$ onto the space spanned by the columns of the $\mathbf{X}$ matrix, yielding the predicted values $\widehat{\mathbf{y}}$. This projection is accomplished by projection matrices, the symmetric idempotent matrices that we encountered in the previous chapter. We see this in the expression $\widehat{\mathbf{y}}= \mathbf{Xb} =\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, where the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ projects the vector $\mathbf{y}$ onto a linear combination of $\mathbf{x}'s$, called $\widehat{\mathbf{y}}$. For this reason the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the hat matrix, because 'it puts the hat on $\mathbf{y}$'. This geometric approach can be used to understand the idempotency property of these matrices, and also a property where, when projecting onto a smaller subspace, we obtain results like $\mathbf{PP}_C = \mathbf{P}_C\mathbf{P} = \mathbf{P}_C$, which were observed when testing subsets of coefficients.

Estimation of Error Variance

The text points out that although $\varepsilon \sim N_n(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_n)$, the residuals are correlated and have different variances, $\mathbf{e} \sim N_n(\mathbf{0}, \sigma_\varepsilon^2\mathbf{Q})$. In fact, $\mathbf{Q} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is nondiagonal, singular, and has rank $n-k-1$. Of course, we already know that $\mathbf{Q}$ is singular with rank $n-k-1$, because we have shown that it is idempotent and we know that for those matrices we have $\text{rank}(\mathbf{Q}) = \text{trace}(\mathbf{Q})$, which for $\mathbf{Q}$ is $\text{trace}(\mathbf{Q}) = n-k-1$. They then state that a matrix $\mathbf{G}$ (of dimension $n-k-1$ x n) can be found to transform the (n x 1) vector $\mathbf{e}$ to a vector $\mathbf{z}$ of dimension ($n-k-1$ x 1) via $\mathbf{z} = \mathbf{Ge}$. Furthermore, the matrix $\mathbf{G}$ will satisfy $\mathbf{GG}' = \mathbf{I}_{n-k-1}$ and $\mathbf{GX} = \mathbf{0}$ (of dimension $n-k-1$ x $k+1$). We can then show that this transformation has the following properties:

$$\begin{aligned}
\mathbf{z} &= \mathbf{Ge} = \mathbf{G}(\mathbf{y}-\widehat{\mathbf{y}}) = \mathbf{Gy} - \mathbf{GXb} = \mathbf{Gy} - (\mathbf{GX})\mathbf{b} = \mathbf{Gy}, \\
E(\mathbf{z}) &= E(\mathbf{Ge}) = \mathbf{G}E(\mathbf{e}) = \mathbf{0}, \\
V(\mathbf{z}) &= V(\mathbf{Gy}) = \mathbf{G}V(\mathbf{y})\mathbf{G}' = \mathbf{G}\sigma_\varepsilon^2\mathbf{I}_n\mathbf{G}' = \sigma_\varepsilon^2\mathbf{GG}' = \sigma_\varepsilon^2\mathbf{I}_{n-k-1}.
\end{aligned}$$

The text uses these results to find that

$$\frac{(n-k-1)S_E^2}{\sigma_\varepsilon^2} \sim \chi^2_{n-k-1}.$$

We had previously demonstrated that S_E^2 has $n-k-1$ degrees of freedom by calculating $E(\mathbf{e}'\mathbf{e}) = E(\mathbf{y}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y})$.

ANOVA models

The geometric approach is used to show that when an overparametrized model is used, $\mathbf{y}$ can still be orthogonally projected onto the column space of $\mathbf{X}$, but parameter estimates are not uniquely determined.