Chapter 1

Least squares when X is not of full rank

4 September 2016

Consider the least squares problem where the $n \times p$ matrix X of predictors has rank m, which is $< p$. For simplicity, suppose the columns of X have been rearranged so that $x_1, \dots, x_m$ are linearly independent, so that x_j for $j > m$ could be written as a linear combination of the first m. (The LINPACK QR algorithm (Dongarra et al., 1993, Chapter 9) permutes the order of the columns as it applies the Householder transformations.) Partition X and the matrices Q accordingly

$$X = \begin{array}{cc} & \begin{array}{cc} m & p-m \end{array} \\ n & \begin{bmatrix} X_1 & X_2 \end{bmatrix} \end{array} \quad \text{AND} \quad Q = \begin{array}{cc} & \begin{array}{ccc} m & p-m & n-p \end{array} \\ n & \begin{bmatrix} Q_1 & Q_2 & Q_3 \end{bmatrix} \end{array} = (q_1, q_2, \dots, q_n).$$

(The little numbers above and to the left of the matrices show the sizes of the blocks.) The QR algorithm arranges that the upper triangular matrix has a corresponding partition.

$$
Q^T X = \begin{array}{c} \\ m \\ p-m \\ n-p \end{array} \begin{array}{c} n \\ \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix} \end{array} \begin{bmatrix} X_1, X_2 \end{bmatrix} = \begin{array}{c} \\ m \\ p-m \\ n-p \end{array} \begin{array}{c} \begin{array}{cc} m & p-m \end{array} \\ \begin{bmatrix} Q_1^T X_1 & Q_1^T X_2 \\ Q_2^T X_1 & Q_2^T X_2 \\ Q_3^T X_1 & Q_3^T X_2 \end{bmatrix} \end{array}
$$

$$
= \begin{array}{c} \\ p \\ n-p \end{array} \begin{array}{c} p \\ \begin{bmatrix} R \\ 0 \end{bmatrix} \end{array} = \begin{array}{c} \\ m \\ p-m \\ n-p \end{array} \begin{array}{c} \begin{array}{cc} m & p-m \end{array} \\ \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{array}
$$

so that

$$
\begin{bmatrix} X_1 & X_2 \end{bmatrix} = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 R_1 & Q_1 R_2 \\ 0 & 0 \end{bmatrix}.
$$

That is, $X_1 = Q_1 R_1$ and $X_2 = Q_1 R_2$. Notice the effect of all the zeros in the R matrix.

The $m \times m$ matrix R_1 is upper triangular with nonzero elements down its diagonal; it has an inverse. The columns of Q_1 provide an orthonormal basis for the subspace $\mathcal{X}$ spanned by the columns of X, which is the same as the subspace spanned by the columns of X_1. The columns of Q_2 and Q_3 span the subspace $\mathcal{X}^\perp$. Thus

$$
X_2 = Q_1 R_2 = X_1 R_1^{-1} R_2. \qquad <1.1>
$$

As asserted, the columns of X_2 are linear combinations of the columns of X_1.

The matrix $H = Q_1 Q_1^T = \sum_{i \le m} q_i q_i^T$ projects $\mathbb{R}^n$ orthogonally onto the m-dimensional subspace $\mathcal{X}$ spanned by the columns of X. In particular $\widehat{y} = Hy = Q_1 Q_1^T y$.

The fitted vector can also be represented non-uniquely as a linear combination of the columns of X, that is, $\widehat{y} = Xb$. By equality <1.1>, we can also get away with a linear combination of just the X_1 columns. For some $m \times 1$ column vector $\widehat{d}$,

$$
Q_1 Q_1^T y = \widehat{y} = X_1 d = Q_1 R_1 \widehat{d}.
$$

Premultiply by Q_1^T, using the fact that $Q_1^T Q_1 = I_m$, then solve for $\widehat{d}$:

$$
\widehat{d} = R_1^{-1} Q_1^T y.
$$

This solution corresponds to setting the last $p-m$ components of b to zero. When displaying output (as in *summary*()), **R** inserts NA's (the symbol for missing data) instead of showing zeros. Why is this strategy better than just putting in zeros?

Note that $\widehat{d}$ is uniquely determined. The general solution for $\widehat{y} = Xb$ with b partitioned as $[B_1, B_2]$ is: choose $B_2 \in \mathbb{R}^{p-m}$ arbitrarily then put B_1 equal to $\widehat{d} - R_1^{-1}R_2B_2$:

$$X_1\widehat{d} = \widehat{y} = X_1B_1 + X_2B_2 = X_1\widehat{d} - X_1R_1^{-1}R_2B_2 + X_2B_2.$$

In general there is no compelling reason to display all the possible solutions for $\widehat{b}$. In some cases, it is helpful to see a solution $\widehat{b}$ that satisfies some extra constraints. For example, for models involving factors, can be parametrized in a number of useful ways. I'll say more about that case in a few weeks.

The lm object returned by the $lm()$ function contains everything we need. In particular, once we know y and the qr component we could calculate all the information returned by the summary function, and more.

The file RMDdemo.Rmd, which is in the Handouts directory, provides a numerical illustration. It is written using R Markdown, a simplified version of the **R** Sweave that I used for `least_squares.Rnw`. R Markdown has an incomplete understanding of how LaTeX works. However, if you are not familiar with LaTeX you will probably want to use R Markdown for some of your homework.

References

Dongarra, J. J., C. B. Moler, J. R. Bunch, and G. W. Stewart (1993). *Linpack Users' Guide*. Society for Industrial and Applied Mathematics.