Myrto Limnios, Marco Piccininni

Solution 1

(a) Here

$$y = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & x_{11} & 0 \\ 1 & 0 & x_{12} & 0 \\ 1 & 0 & x_{13} & 0 \\ 0 & 1 & 0 & x_{21} \\ 0 & 1 & 0 & x_{22} \\ 0 & 1 & 0 & x_{23} \end{pmatrix}, \quad \beta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

(b) With

$$X = \begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{12} & 0 & 0 \\ 1 & x_{13} & 0 & 0 \\ 1 & x_{21} & 1 & x_{21} \\ 1 & x_{22} & 1 & x_{22} \\ 1 & x_{23} & 1 & x_{23} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix},$$

we have model (i) with columns 1, 2, 3, model (ii) with columns 1, 2, 4, and model (iii) with columns 1 and 2.

Solution 2

(a) The first and second derivatives of

$$\|y - X\beta\|^2 = \sum_{j=1}^{n} (y_j - x_j^{\mathrm{T}}\beta)^2,$$

with respect to β (and β^{T} for the second) are

$$-2\sum_{j=1}^{n} (y_j - x_j^{\mathrm{T}}\beta)x_j = -2X^{\mathrm{T}}(y - X\beta), \quad 2\sum_{j=1}^{n} x_j x_j^{\mathrm{T}} = 2X^{\mathrm{T}}X.$$

If X has rank p then so too does $X^{\mathrm{T}}X$ and hence its inverse exists, and a little algebra after setting $-2X^{\mathrm{T}}(y - X\beta) = 0$ yields $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$. This gives the unique minimum because the second derivative matrix is positive definite.

(b) Symmetry and idempotency of H are simple to check. Note also that

$$(I - H)^2 = (I - H)(I - H) = I - 2H + H^2 = I - 2H + H = I - H,$$

so $I - H$ is also symmetric and idempotent. If v is an eigenvector of H, then $H^2 v = H(\lambda v) = \lambda^2 v$, but as $H^2 v = Hv = \lambda v$, the eigenvalues must satisfy $\lambda^2 = \lambda$, which implies that $\lambda = 1$ or $\lambda = 0$. But the trace of H is the sum of its eigenvalues, and this equals the trace of $\mathrm{tr}\{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X\} = \mathrm{tr}(I_p) = p$, so H has p eigenvalues equal to 1 and $n - p$ equal to 0.

(c) H' is a projection onto the vector subspace $\mathcal{V}'$ of $\mathbb{R}^n$ spanned by the columns of X', and this is a subspace of the space $\mathcal{V}$ generated by the columns of X, i.e., $\mathcal{V}' \subset \mathcal{V} \subset \mathbb{R}^n$.

Clearly $(I - H)H = H - H^2 = H - H = 0$, and likewise for H', and H and H' are projection matrices onto $\mathcal{V}$ and $\mathcal{V}'$.

Let $y \in \mathbb{R}^n$, and note that $H'y \in \mathcal{V}'$, so $H'y \in \mathcal{V}$, so $HH'y = H'y$, which implies that $HH' = H'$, because y was arbitrary. Hence

$$HH' = H' = (H')^{\mathrm{T}} = (HH')^{\mathrm{T}} = (H')^{\mathrm{T}}H^{\mathrm{T}} = H'H,$$

rearrangement of which gives the required

$$H'(H - H') = H'(I - H) = H'(I - H') = H(I - H) = 0.$$

Solution 3

(a) Let Q have spectral decomposition VDV^{T}, where the columns of the orthogonal matrix V are the eigenvectors of Q and the diagonal matrix D contains its eigenvalues (which are all positive). Recall that $VV^{\mathrm{T}} = V^{\mathrm{T}}V = I_n$. Then we can write

$$Q^{-1} = VD^{-1}V^{\mathrm{T}} = W, \quad W^{1/2} = VD^{-1/2}V^{\mathrm{T}},$$

say, where $W^{1/2}$ is symmetric, and hence

$$y_* = W^{1/2}y \sim (W^{1/2}X\beta, \sigma^2 W^{1/2}Q(W^{1/2})^{\mathrm{T}} \sim (X_*\beta, \sigma^2 I_n),$$

say, because $W^{1/2}Q(W^{1/2})^{\mathrm{T}} = W^{1/2}W^{-1}W^{1/2} = I_n$. Hence

$$\begin{aligned} \widehat{\beta} &= (X_*^{\mathrm{T}}X_*)^{-1}X_*^{\mathrm{T}}y_* \\ &= \{X^{\mathrm{T}}(W^{1/2})^{\mathrm{T}}W^{1/2}X\}^{-1}X^{\mathrm{T}}(W^{1/2})^{\mathrm{T}}W^{1/2}y \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wy, \end{aligned}$$

the hat matrix is

$$H = X_*(X_*^{\mathrm{T}}X_*)^{-1}X_*^{\mathrm{T}} = W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2},$$

and the residual sum of squares is

$$\begin{aligned} y_*^{\mathrm{T}}(I_n - H)y_* &= y^{\mathrm{T}}\{W - WX(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W\}y \\ &= y^{\mathrm{T}}W^{1/2}\{I_n - W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2}\}W^{1/2}y. \end{aligned}$$

(b) When Q is diagonal we have $\mathrm{var}(y_j) \propto q_{jj} = 1/w_j$, and then we can write

$$\widehat{\beta} = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wy = \left(\sum_{j=1}^{n} w_j x_j x_j^{\mathrm{T}}\right)^{-1} \sum_{j=1}^{n} w_j x_j^{\mathrm{T}} y_j,$$

so we see that the contribution to $\widehat{\beta}$ from the jth case, (x_j, y_j), is given weight w_j, where the weight w_j is large if the corresponding variance q_{jj} is small.