①   9 April 08

Straight line regression

$$h_{jj} = \frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum (x_k - \bar{x})^2}$$

"responses with large values of $|x_j - \bar{x}|$ will strongly affect the slope of the fitted line"

Several x's

Cook's distance

$$C_j = \frac{1}{p s^2} (\hat{y} - \hat{y}_{-j})^T (\hat{y} - \hat{y}_{-j})$$

$$\hat{y}_{-j} = X \hat{\beta}_{-j}$$

"measures the overall change in the fitted values when the j-th case is deleted"

"large values of C_j arise if a case has high leverage or a large standardized residual"

$$C_j = r_j^2 h_{jj} / p(1-h_{jj})$$

Table 8.3

Table 8.3 Data from bicycle experiment, together with fitted values $\widehat{y}$, raw residuals e, standardized residuals, r, deletion residuals r', leverages h and Cook distances C.

Setup	Seat height	Dynamo	Tyre pressure	Time y	$\widehat{y}$	e	r	r'	h	C
1	−1	−1	−1	51	52.62	−1.625	−0.99	−0.99	0.25	0.08
2	−1	−1	−1	54	52.62	1.375	−0.84	0.83	0.25	0.06
3	1	−1	−1	41	41.75	−0.750	−0.46	−0.44	0.25	0.02
4	1	−1	−1	43	41.75	1.250	0.76	0.75	0.25	0.05
5	−1	1	−1	54	55.75	−1.750	−1.06	−1.07	0.25	0.09
6	−1	1	−1	60	55.75	4.250	2.59	3.72	0.25	0.56
7	1	1	−1	44	44.87	−0.875	−0.53	−0.52	0.25	0.02
8	1	1	−1	43	44.87	−1.875	−1.14	−1.16	0.25	0.11
9	−1	−1	1	50	49.50	0.500	0.30	0.29	0.25	0.01
10	−1	−1	1	48	49.50	−1.500	−0.91	−0.91	0.25	0.07
11	1	−1	1	39	38.62	0.375	0.23	0.22	0.25	0.00
12	1	−1	1	39	38.62	0.375	0.23	0.22	0.25	0.00
13	−1	1	1	53	52.62	0.375	0.23	0.22	0.25	0.00
14	−1	1	1	51	52.62	−1.625	−0.99	−0.99	0.25	0.08
15	1	1	1	41	41.75	−0.750	−0.46	−0.44	0.25	0.02
16	1	1	1	44	41.75	2.250	1.37	1.43	0.25	0.16

Sometimes e_j is called a *raw residual*.

we obtain $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$. The fitted value $\widehat{y} = X\widehat{\beta} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = Hy$ is the orthogonal projection of y onto the plane spanned by the columns of X, and the matrix representing that projection is H. Notice that $\widehat{y}$ is unique whether or not $X^{\mathrm{T}}X$ is invertible.

Figure 8.2 shows that the vector of residuals, $e = y - \widehat{y} = (I_n - H)y$, and the vector of fitted values, $\widehat{y} = Hy$, are orthogonal. To see this algebraically, note that

$$\widehat{y}^{\mathrm{T}}e = y^{\mathrm{T}}H^{\mathrm{T}}(I_n - H)y = y^{\mathrm{T}}(H - H)y = 0, \tag{8.6}$$

because $H^{\mathrm{T}} = H$ and $HH = H$, that is, the projection matrix H is symmetric and idempotent (Exercise 8.2.5). The close link between orthogonality and independence for normally distributed vectors means that (8.6) has important consequences, as we shall see in Section 8.3. For now, notice that (8.6) implies that

$$y^{\mathrm{T}}y = (y - \widehat{y} + \widehat{y})^{\mathrm{T}}(y - \widehat{y} + \widehat{y}) = (e + \widehat{y})^{\mathrm{T}}(e + \widehat{y}) = e^{\mathrm{T}}e + \widehat{y}^{\mathrm{T}}\widehat{y}, \tag{8.7}$$

as is clear from Figure 8.2 by Pythagoras' theorem. That is, the overall sum of squares of the data, $\sum y_j^2 = y^{\mathrm{T}}y$, equals the sum of the residual sum of squares, $SS(\widehat{\beta}) = \sum(y_j - \widehat{y}_j)^2 = e^{\mathrm{T}}e$, and the sum of squares for the fitted model, $\sum \widehat{y}_j^2 = \widehat{y}^{\mathrm{T}}\widehat{y}$.
Such decompositions are central to analysis of variance, discussed below.

8.2.3 Likelihood quantities

Chapter 4 shows how the observed and expected information matrices play a central role in likelihood inference, by providing approximate variances for maximum likelihood estimates. To obtain these matrices for the normal linear model, note that the

Collinearity & poorly conditioned X

$\overset{n \times p}{X}$

For $\hat{\beta}$ to be defined, need $(X^T X)^{-1}$ to exist

need $X^T X$ of full rank

Spectral decomposition svd( )

$\overset{p \times p}{X^T X} = U D U^T$

D: diagonal, U: orthogonal

$D = \text{diag}\{d_1, \ldots, d_p\}$ ordered

d's ≥ 0

rank $r(X^T X) = \#\{d_j \neq 0\}$

numerical trouble can arise

condition number $\sqrt{\max\{d_j\} / \min\{d_j\}}$

For the cement $p = 5$

eigens 44676, 5965.4, 810.0, 105.4, .00012

condition number 6056

Ill-conditioned, Hard to interpret.

Parameter	Estimate	S.e.	Dropping x_5
β_0	62.41	70.07	71.64 (14.14)
β_1	1.55	0.74	1.45 (.12)
β_2	0.51	0.72	0.42 (.19)
β_3	0.10	0.75	
β_4	-0.14	0.71	-0.24 (.17)

Ridge regression

$$\hat{\beta}_\lambda = (X^T X + \lambda I)^{-1} X^T y \ , \ \lambda \geq 0$$

conditions $X^T X$

shrinks estimates towards 0

now biased

May need to rescale x's

Might pick λ by cross-validation

$$CV(\lambda) = \sum (y_j - \hat{y}_{-j})^2$$

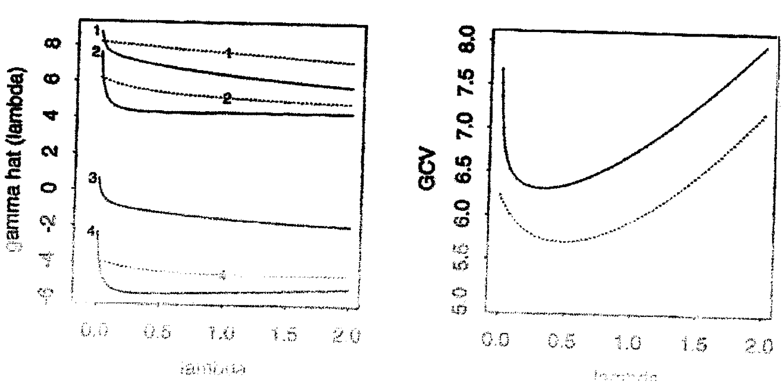

Figure 8.7 Ridge regression analysis of cement data. Left: variation of elements of $\hat{\gamma}_\lambda$ as a function of λ, for models with all four covariates (solid) and with x_1, x_2, and x_4 only (do... Right: generalized cross-validation criterion GCV(λ) for these mod...

(6)

10.3 Generalized linear models

$y = X\beta + \varepsilon$

$-\infty < y_j < \infty$, continuous valued, constant variance

Normal density $\quad -\infty < y < \infty$

Poisson $\quad 0, 1, 2, \ldots$

Binomial $\quad \frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m}{m}$

Jacamar data

	Aphrissa boisduvalli N/S/E	*Phoebis argante* N/S/E	*Dryas iulia* N/S/E	*Pierella luna* N/S/E	*Consul fabius* N/S/E	*Siproeta stelenes*† N/S/E
Unpainted	0/0/14	6/1/0	1/0/2	4/1/5	0/0/0	0/0/1
Brown	7/1/2	2/1/0	1/0/1	2/2/4	0/0/3	0/0/1
Yellow	7/2/1	4/0/2	5/0/1	2/0/5	0/0/1	0/0/3
Blue	6/0/0	0/0/0	0/0/1	4/0/3	0/0/1	0/1/1
Green	3/0/1	1/1/0	5/0/0	6/0/2	0/0/1	0/0/3
Red	4/0/0	0/0/0	6/0/0	4/0/2	0/0/1	3/0/1
Orange	4/2/0	6/0/0	4/1/1	7/0/1	0/0/2	1/1/1
Black	4/0/0	0/0/0	1/0/1	4/2/2	7/1/0	0/1/0

† includes *Philaethria dido* also.

Table 10.2 Response of a rufous-tailed jacamar to individuals of seven species of palatable butterflies with artifically coloured wing undersides. (N=not sampled, S = sampled and rejected, E = eaten)

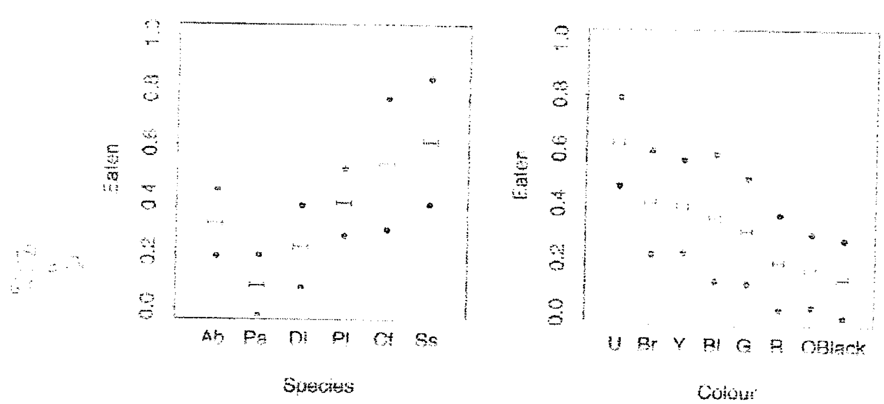

Figure 10.2 Proportion of butterflies eaten [illegible] for different species and wing colour.

density / mass function

$$f(y;\theta,\varphi) = \exp\left\{\frac{y\theta - b(\theta)}{\varphi} + c(y;\varphi)\right\}$$

exponential family (φ known)

$E(Y) = b'(\theta) = \mu$

$\mathrm{var}(Y) = \phi\, b''(\theta) \quad = \varphi V(\mu)$

linear predictor $\eta = x^T\beta$

$\eta = g(\mu) \qquad g$: link

φ: dispersion parameter

normal

$\theta = \mu$

$\varphi = \sigma^2$

$b(\theta) = \frac{1}{2}\theta^2$

$c(y;\varphi) = -\frac{1}{2\varphi}y^2 - \frac{1}{2}\log(2\pi\varphi)$

$V(\mu) = 1$