(logistic) regression

social office hour: Lúpulo 6-7 pm

AMS 206 14 Mar 19

$y = (y_1, \dots, y_n)$ ①

M_1: $\ell\ell_1(\underset{\sim}{\theta}_1 \mid y \dots M_1 \mathcal{B})$ $(\underset{\sim}{\theta}_1)_{k_1}$

(constant) $k_1 = 2$

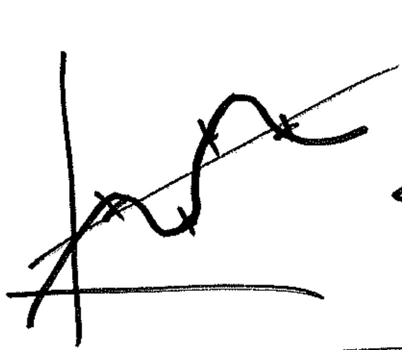
M_2: $\ell\ell_2(\underset{\sim}{\theta}_2 \mid y \dots M_2 \mathcal{B})$ $(\underset{\sim}{\theta}_2)_{k_2}$

(linear) $k_2 = 3$

one idea: prefer M_2 if its $\ell\ell_2 > \ell\ell_1$

goodness of fit

$$\ell\ell(\hat{\underset{\sim}{\theta}}_{2,MLE} \mid y\, M_2 \mathcal{B}) > \ell\ell(\hat{\underset{\sim}{\theta}}_{1,MLE} \mid y\, M_1 \mathcal{B})$$



model complexity

$k_2 > k_1$?

more complex

want large $\left[\ell\ell_2(\hat{\underset{\sim}{\theta}}_{2,MLE} \mid y\, M_2 \mathcal{B}) - f(k_2) \right]$ K ?

$- f(k_2, n)$

②

one good answer

$$BIC(M_j \mid \underset{\sim}{y}\, M_j\, \mathcal{B})$$

Schwarz (1982)

$$= -2\,\ell\ell_j(\hat{\underset{\sim}{\theta}}_{j,\,MLE} \mid \underset{\sim}{y}\, M_j\, \mathcal{B}) + k_j \cdot \log(n)$$

Bayesian information criterion

AIC (Akaike) (1975?)

$$AIC(M_j \mid \underset{\sim}{y}\, M_j\, \mathcal{B})$$

$$= -2\,\ell\ell_j(\hat{\underset{\sim}{\theta}}_{j,\,MLE} \mid \underset{\sim}{y}\, M_j\, \mathcal{B}) + 2k_j$$

$$\begin{array}{cc} \underset{\sim}{y} & \hat{\underset{\sim}{y}} \end{array}$$

$$\left[\begin{array}{c|c} y_1 & \hat{y}_1 \\ \vdots & \vdots \\ y_n & \hat{y}_n \end{array}\right]$$

$$\begin{array}{c|c} \in \mathbb{R} & \in \mathbb{R} \end{array}$$

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \text{root mean squared error (RMSE)}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| = \text{mean absolute deviation (MAD)}$$

$$Y_i = \sum_{j=1}^{k} \beta_j X_{ij} + e_i \quad (i = 1, \dots, n)$$

e_i IID $N(0, \sigma_e^2)$

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \underset{n\times 1}{\mathbf{Y}}, \quad \underset{k\times 1}{\underset{\sim}{\beta}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\underset{n\times k}{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \qquad \underset{n\times 1}{\underset{\sim}{e}} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times k}{X} \cdot \underset{k\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{e}} \qquad \underset{(k+1)\times 1}{\underset{\sim}{\theta}} = (\underset{\sim}{\beta}, \sigma_e^2)$$

fact: $$\hat{\underset{\sim}{\beta}}_{MLE} = (X^T X)^{-1} X^T \mathbf{Y} \qquad E_{RS}(\hat{\underset{\sim}{\beta}}_{MLE}) = \underset{\sim}{\beta}$$

$$\hat{\mathbf{Y}} = X \cdot \hat{\underset{\sim}{\beta}}_{MLE} \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$ (biased on low side)

$$\hat{\sigma}^2_{\text{unbiased}} = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$C(\hat{\beta}_{MLE}) = \hat{\Sigma}_{\hat{\beta}_{MLE}} = \sigma_e^2 (X^T X)^{-1}$$

---

simple linear regression

$$V(\hat{Y}_i) = V(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= V(\hat{\beta}_0) + V(\hat{\beta}_1 x_i) + 2C(\hat{\beta}_0, \hat{\beta}_1 x_i)$$

$$= \boxed{V(\hat{\beta}_0)} + x_i^2 \boxed{V(\hat{\beta}_1)} + 2x_i \boxed{C(\hat{\beta}_0, \hat{\beta}_1)}$$

5

$$\underset{n \times 1}{\underset{\sim}{y}} = \begin{pmatrix} 1s \\ \& \\ 0s \end{pmatrix} \qquad \underset{n \times k}{X} = \begin{bmatrix} 1 & \text{preg.} & & \\ \vert & & & \cdots \\ 1 & & & \end{bmatrix}$$

$n = 768$, $k = 9$

1s & 0s

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{e}$$

logistic regression (special case of **generalized linear models**)

$$(Y_i \mid p_i) \stackrel{I}{\sim} \text{Bernoulli}(p_i)$$

$$(i = 1, \ldots, n)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \left(\sum_{j=1}^{k} \underbrace{\beta_j}_{\in \mathbb{R}} \, \underbrace{x_{ij}}_{\in \mathbb{R}}\right) \in \mathbb{R}$$

$$(\underset{\sim}{\beta} \mid \mathcal{D}\,\mathcal{B}) \sim p(\underset{\sim}{\beta} \mid \mathcal{D}\,\mathcal{B})$$