multiple linear regression

STAT 206
16 Mar 21

DD extra OH

$$y_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\beta_j + e_i \quad (i = 1, \dots, n)$$

e_i IID $N(0, \sigma^2)$ (1)

$$D = \left[\, y \,\Bigg\|\, \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \,\Bigg|\, \underset{\sim}{x}_1 \;\; \underset{\sim}{x}_2 \; \cdots \; \underset{\sim}{x}_k \,\right] \updownarrow n$$

choosing the optimal subset of the $\underset{\sim}{x}_j$: the variable selection problem

with k predictors, there are 2^k different subsets

TRT 3.2(A)

here $k = 21$; $2^k = 2.1$ million distinct subsets

BIC = special case of Bayes factors with a specific UI prior (unit-information (UI) prior)

parameter vector $\underset{\sim}{\theta}_j$: $|\underset{\sim}{\theta}_j| = k_j$

$$BIC(M_j \mid D\,B) = -2\,\ell\ell(\hat{\theta}_{MLE} \mid D\,M_j\,B) + k_j \log(n)$$

②

$$BIC(M_j \mid \mathbf{y}) = -2\,\ell\ell\big[(\hat{\theta}_j)_{MLE} \mid D\,M_j\,\mathbf{y}\big]$$

goodness of fit

$$+\; k_j \ln(n)$$

(complexity penalty)

small BIC values indicate better models

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

(residual sum of squares)

(good models have small RSS)

smallest possible RSS = 0

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