读书报告

第三章 线性模型

刘精昌

September 26, 2016

引例

工资与教育水平关系

考查工人工资水平与其受教育关系：

a 工资水平（每小时美元数）：用 Y 表示

b 受教育程度（受教育年数）：用 X 表示

c 非可观测因素，如工作经验、天生素质、工作时间等其他因素

引例

工资与教育水平关系

- 观测的数据：$(x_i, y_i), i = 1, \cdots, n$ （受访人数），

i	x_i	y_i	i	x_i	y_i
1	5.3	1.4	9	8.5	3.2
2	11.0	3.9	10	7.1	8.6
3	9	6.3	11	15	4
4	8.7	8.6	12	12.0	9.0
5	10	12	13	29	12
6	15.5	12	14	19.7	13.1
7	21	16	15	15.1	10
8	19	14.4	16	15.7	16

定义

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1} + e_i,\ i = 1, 2, \ldots, n$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{11} & \cdots & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{11} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_{p-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{p-1} \end{pmatrix}$$

$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1} + e_i,\ i = 1, 2, \ldots, n$$

(a) 拟合值：$\hat{y}_i = \hat{\beta}_0 + x_{i1}\hat{\beta}_1 \cdots + x_{i,p-1}\hat{\beta}_{p-1}$

(b) 残差（residual）:$\varepsilon_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - x_{i1}\hat{\beta}_1 \cdots + x_{i,p-1}\hat{\beta}_{p-1}$

(c) 残差平方和（residual sum of squares RSS）

:$RSS(\beta) = \sum_{i=1}^{n}\varepsilon_i = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) = \left(\hat{y} - X\hat{\beta}\right)^T\left(\hat{y} - X\hat{\beta}\right)$

参数求解

最小二乘法：RSS 最小

1. $E(\hat{\beta}) = \left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right) = y^T y - 2y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta}$
2. $\dfrac{\partial E\left(\hat{\beta}\right)}{\partial \hat{\beta}} = 0$
3. $X^T X\hat{\beta} = X^T y$
4. $\hat{\beta} = (X^T X)^{-1} X^T y$

MLE

1. $y_i = \beta^T x_i + e_i, e_i \sim N\left(0, \sigma^2\right)$
2. $y_i \sim N\left(\beta^T x_i, \sigma^2\right)$
3. $\overset{\wedge}{\beta} \triangleq \arg\max_{\beta} \log p\left(D \,|\beta\right)$
4. $l\left(\beta\right) \triangleq \log p\left(D \,|\beta\right) = \sum_{i=1}^{n} \log p\left(y_i \,|\beta\right) = \sum_{i=1}^{n} \log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \beta^T x_i\right)^2\right)\right] = -\frac{1}{2\sigma^2} RSS - \frac{N}{2}\log\left(2\pi\sigma^2\right)$

平方和的定义

a、 总平方和 (Total Sum of Squares TSS):

$$TSS \triangleq \sum_{i=1}^{n} (y_i - \overline{y})^2, \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

b、 解释平方和（Explained Sum of Squares ESS）:

$$ESS \triangleq \sum_{i=1}^{n} \left(\hat{y}_i - \hat{y} \right)^2, \overline{\hat{y}} = \frac{1}{n} \sum_{i=1}^{n} \hat{y}_i$$

拟合优度

判定系数的定义

$R^2 = \frac{ESS}{TSS}$

判定系数的性质

- $0 \leq R^2 \leq 1$
- R^2 越接近 1，拟合效果越好
- R^2 越接近 0，拟合效果越差

几何解释

增加惩罚项

ridge regression

- $J(\beta) = \frac{1}{n}RSS + \lambda \|\beta\|_2^2$
- $\underset{ridge}{\overset{\wedge}{\beta}} = \min_{\beta} \arg J(\beta)$
- $\underset{ridge}{\overset{\wedge}{\beta}} = \left(\lambda I + X^T X\right)^{-1} X^T y$

LASSO

- $J(\beta) = \frac{1}{n}RSS + \lambda \|\beta\|_2^1$