สมการ Linear Regression

จากพื้นฐานถึง Generalized Model

คู่มือสมการคณิตศาสตร์แบบครบถ้วน

เอกสารอ้างอิงสำหรับนักศึกษาและนักวิจัย

27 มิถุนายน พ.ศ. 2568

เมธานนท์ แก้วกระจ่าง

Contents

เมธานนท์ แก้วกระจ่าง

1 สมการพื้นฐาน (Basic Linear Regression)

1.1 Simple Linear Regression (Feature เดียว)

สมการทั่วไป:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

โดยที่:

- y = ตัวแปรตาม (dependent variable)
- x = ตัวแปรอิสระ (independent variable)
- β_0 = intercept (จุดตัดแกน y)
- β_1 = slope (ความชัน)
- ε = error term (ค่าผิดพลาด)

สมการสำหรับการทำนาย:

$$\hat{y} = \beta_0 + \beta_1 x \tag{2}$$

1.2 กรณีจุดข้อมูลไม่กี่จุด

สำหรับ 2 จุด: $(x_1, y_1), (x_2, y_2)$

$$\beta_1 = \frac{y_2 - y_1}{x_2 - x_1} \tag{3}$$

$$\beta_0 = y_1 - \beta_1 x_1 \tag{4}$$

สำหรับ 3 จุด: $(x_1, y_1), (x_2, y_2), (x_3, y_3)$

ใช้ Least Squares Method:

$$\beta_1 = \frac{\sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{3}(x_i - \bar{x})^2} \tag{5}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{6}$$

2 สมการสำหรับ Dataset ขนาดใหญ่

2.1 Simple Linear Regression (Feature เดียว)

รูปแบบทั่วไป:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{7}$$

สมการ Normal Equations:

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{8}$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \beta_1\frac{1}{n}\sum_{i=1}^{n} x_i \tag{9}$$

2.2 รูปแบบเมทริกซ์ (Simple Linear Regression)

สมการเมทริกซ์:

$$\mathsf{y} = \mathsf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{10}$$

โดยที่:

$$\mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathsf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{11}$$

สมการหาค่าพารามิเตอร์:

$$\hat{\boldsymbol{\beta}} = \left(\mathsf{X}^T\mathsf{X}\right)^{-1}\mathsf{X}^T\mathsf{y} \tag{12}$$

3 Multiple Linear Regression

3.1 สมการทั่วไป

รูปแบบพื้นฐาน:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{13}$$

สำหรับ n จุดข้อมูล:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{14}$$

3.2 รูปแบบเมทริกซ์ (Multiple Linear Regression)

สมการเมทริกซ์:

$$\mathsf{y} = \mathsf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{15}$$

โดยที่:

$$\mathsf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathsf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{16}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{17}$$

สมการหาค่าพารามิเตอร์:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{18}$$

3.3 Normal Equations สำหรับ Multiple Regression

สมการระบบ:

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y} \tag{19}$$

รูปแบบกระจาย:

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{i1} + \beta_2 \sum_{i=1}^{n} x_{i2} + \cdots + \beta_p \sum_{i=1}^{n} x_{ip} \tag{20}$$

$$\sum_{i=1}^{n} x_{i1}y_i = \beta_0 \sum_{i=1}^{n} x_{i1} + \beta_1 \sum_{i=1}^{n} x_{i1}^2 + \beta_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + \beta_p \sum_{i=1}^{n} x_{i1}x_{ip} \tag{21}$$

$$\vdots \tag{22}$$

$$\sum_{i=1}^{n} x_{ip}y_i = \beta_0 \sum_{i=1}^{n} x_{ip} + \beta_1 \sum_{i=1}^{n} x_{ip}x_{i1} + \beta_2 \sum_{i=1}^{n} x_{ip}x_{i2} + \cdots + \beta_p \sum_{i=1}^{n} x_{ip}^2 \tag{23}$$

4 Generalized Linear Model (GLM)

4.1 รูปแบบทั่วไป

สมการพื้นฐาน:

$$g(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta} \tag{24}$$

โดยที่:

- $g(\cdot)$ = link function
- $\mu_i = E[y_i]$ = expected value
- $\mathbf{x}_i$ = vector ของ covariates สำหรับ observation ที่ i

4.2 องค์ประกอบของ GLM

1. **Random Component:**

$$y_i \sim f(y_i|\theta_i, \phi) \tag{25}$$

2. **Systematic Component:**

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} \tag{26}$$

3. **Link Function:**

$$g(\mu_i) = \eta_i \tag{27}$$

4.3 ตัวอย่าง GLM สำหรับกรณีต่างๆ

Linear Regression:

- Distribution: Normal
- Link function: Identity, $g(\mu) = \mu$
- $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$

Logistic Regression:

- Distribution: Binomial
- Link function: Logit, $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
- $\log\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta}$

Poisson Regression:

- Distribution: Poisson
- Link function: Log, $g(\mu) = \log(\mu)$
- $\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$

4.4 การประมาณค่าพารามิเตอร์ใน GLM

Maximum Likelihood Estimation:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \log L(\boldsymbol{\beta}) \tag{28}$$

Iteratively Reweighted Least Squares (IRLS):

$$\boldsymbol{\beta}^{(t+1)} = \left(\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)} \tag{29}$$

โดยที่:

- $\mathsf{W}^{(t)}$ = diagonal weight matrix
- $\mathsf{z}^{(t)}$ = working response vector

5 Regularized Regression

5.1 Ridge Regression (L2 Regularization)

Cost Function:

$$J(\boldsymbol{\beta}) = \frac{1}{2n}\|\mathsf{y} - \mathsf{X}\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_2^2 \tag{30}$$

รูปแบบกระจาย:

$$J(\boldsymbol{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha\sum_{j=1}^{p}\beta_j^2 \tag{31}$$

สมการหาค่าพารามิเตอร์:

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = \left(\mathsf{X}^T\mathsf{X} + \alpha\mathsf{I}\right)^{-1}\mathsf{X}^T\mathsf{y} \tag{32}$$

โดยที่ $\alpha \geq 0$ คือ regularization parameter และ **I** คือ identity matrix

5.2 LASSO Regression (L1 Regularization)

Cost Function:

$$J(\boldsymbol{\beta}) = \frac{1}{2n}\|\mathsf{y} - \mathsf{X}\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1 \tag{33}$$

รูปแบบกระจาย:

$$J(\boldsymbol{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha\sum_{j=1}^{p}|\beta_j| \tag{34}$$

Optimization Problem:

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \arg\min_{\boldsymbol{\beta}}\left\{\frac{1}{2n}\|\mathsf{y} - \mathsf{X}\boldsymbol{\beta}\|_2^2 + \alpha\|\boldsymbol{\beta}\|_1\right\} \tag{35}$$

5.3 Elastic Net Regression

Cost Function:

$$J(\boldsymbol{\beta}) = \frac{1}{2n}\|\mathsf{y} - \mathsf{X}\boldsymbol{\beta}\|_2^2 + \alpha\left(\rho\|\boldsymbol{\beta}\|_1 + \frac{1-\rho}{2}\|\boldsymbol{\beta}\|_2^2\right) \tag{36}$$

โดยที่:

- $\alpha \geq 0$ = regularization strength
- $0 \leq \rho \leq 1$ = mixing parameter
- เมื่อ $\rho = 1$ จะเป็น LASSO
- เมื่อ $\rho = 0$ จะเป็น Ridge

6 Polynomial Regression

6.1 สมการพื้นฐาน

Polynomial ระดับ d:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d + \varepsilon \tag{37}$$

รูปแบบเมทริกซ์:

$$\mathsf{y} = \mathsf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{38}$$

โดยที่:

$$\mathsf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix} \tag{39}$$

6.2 Multivariate Polynomial Regression

สำหรับ 2 ตัวแปร (degree 2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon \tag{40}$$

รูปแบบทั่วไป:

$$y = \sum_{i=0}^{d} \sum_{j=0}^{d-i} \beta_{ij} x_1^i x_2^j + \varepsilon \tag{41}$$

7 Logistic Regression

7.1 Binary Logistic Regression

Logit Function:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{42}$$

Probability Function:

$$p(y = 1|\mathsf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}} = \frac{1}{1 + e^{-\mathsf{x}^T\boldsymbol{\beta}}} \tag{43}$$

Sigmoid Function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad \text{โดยที่} \quad z = \mathsf{x}^T\boldsymbol{\beta} \tag{44}$$

7.2 Maximum Likelihood Estimation

Log-Likelihood:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{45}$$

Cost Function (Cross-Entropy):

$$J(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{46}$$

7.3 Multinomial Logistic Regression

สำหรับ K classes:

$$p(y = k|\mathbf{x}) = \frac{e^{\mathbf{x}^T\boldsymbol{\beta}_k}}{\sum_{j=1}^{K} e^{\mathbf{x}^T\boldsymbol{\beta}_j}}, \quad k = 1, 2, \dots, K \tag{47}$$

Softmax Function:

$$\text{softmax}(z_k) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \tag{48}$$

8 Robust Regression

8.1 Huber Regression

Huber Loss Function:

$$L_\delta(r) = \begin{cases} \frac{1}{2}r^2 & \text{if } |r| \le \delta \\ \delta|r| - \frac{1}{2}\delta^2 & \text{if } |r| > \delta \end{cases} \tag{49}$$

โดยที่ $r = y - \mathbf{x}^T\boldsymbol{\beta}$ และ δ คือ threshold parameter

8.2 Quantile Regression

สำหรับ quantile ที่ τ (0 < τ < 1):

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\boldsymbol{\beta}) \tag{50}$$

โดยที่:

$$\rho_\tau(u) = u(\tau - \mathbf{1}_{u<0}) = \begin{cases} \tau u & \text{if } u \ge 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \tag{51}$$

9 สรุปและเปรียบเทียบ

9.1 ตารางสรุปสมการหลัก

รูปแบบ	สมการพื้นฐาน	Cost Function	การหาพารามิเตอร์
Linear	$y = \beta_0 + \beta_1 x + \varepsilon$	$\frac{1}{2n} \lVert \mathsf{y} - \mathsf{X}\boldsymbol{\beta} \rVert_2^2$	$\hat{\boldsymbol{\beta}} = (\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\mathsf{y}$
Ridge	$y = \mathsf{x}^T\boldsymbol{\beta} + \varepsilon$	$\frac{1}{2n} \lVert \mathsf{y} - \mathsf{X}\boldsymbol{\beta} \rVert_2^2 + \alpha \lVert \boldsymbol{\beta} \rVert_2^2$	$\hat{\boldsymbol{\beta}} = (\mathsf{X}^T\mathsf{X} + \alpha\mathsf{I})^{-1}\mathsf{X}^T\mathsf{y}$
LASSO	$y = \mathsf{x}^T\boldsymbol{\beta} + \varepsilon$	$\frac{1}{2n} \lVert \mathsf{y} - \mathsf{X}\boldsymbol{\beta} \rVert_2^2 + \alpha \lVert \boldsymbol{\beta} \rVert_1$	Coordinate Descent
Logistic	$\log \frac{p}{1-p} = \mathsf{x}^T\boldsymbol{\beta}$	$-\frac{1}{n} \sum[y \log p + (1 - y) \log(1 - p)]$	Newton-Raphson
Polynomial	$y = \sum_{i=0}^{d} \beta_i x^i + \varepsilon$	$\frac{1}{2n} \lVert \mathsf{y} - \mathsf{X}\boldsymbol{\beta} \rVert_2^2$	$\hat{\boldsymbol{\beta}} = (\mathsf{X}^T\mathsf{X})^{-1}\mathsf{X}^T\mathsf{y}$

Table 1: สรุปสมการ Regression ทุกประเภท

9.2 ตารางเปรียบเทียบ Regularization

Method	Penalty	คุณสมบัติหลัก	ข้อดี/ข้อเสีย
Ridge	L2: $\lVert \boldsymbol{\beta} \rVert_2^2$	Shrink coefficients	ดี: Stable, เสีย: ไม่ feature selection
LASSO	L1: $\lVert \boldsymbol{\beta} \rVert_1$	Sparse solutions	ดี: Feature selection, เสีย: Unstable
Elastic Net	L1 + L2	ทั้งสองอย่าง	ดี: Balanced, เสีย: Extra parameter

Table 2: เปรียบเทียบ Regularization Methods

9.3 การวัดประสิทธิภาพ

สำหรับ Regression:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{52}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{53}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{54}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{55}$$

สำหรับ Classification:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{56}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{57}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{58}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{59}$$

10 Cross-Validation และ Model Selection

10.1 K-Fold Cross-Validation

CV Error:

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k}\text{MSE}_i \tag{60}$$

โดยที่ MSE_i คือ Mean Squared Error ของ fold ที่ i

10.2 Information Criteria

Akaike Information Criterion (AIC):

$$\text{AIC} = 2k - 2\ln(\hat{L}) \tag{61}$$

Bayesian Information Criterion (BIC):

$$\text{BIC} = k\ln(n) - 2\ln(\hat{L}) \tag{62}$$

โดยที่:

- k = จำนวนพารามิเตอร์
- n = จำนวนข้อมูล
- $\hat{L}$ = maximum likelihood

10.3 Regularization Path

สำหรับ LASSO และ Ridge:

$$\alpha_{\text{optimal}} = \arg\min_{\alpha} CV(\alpha) \tag{63}$$

Grid Search:

$$\alpha \in \{10^{-4}, 10^{-3}, \ldots, 10^{2}\} \tag{64}$$

11 ข้อสมมุติและข้อจำกัด

11.1 ข้อสมมุติสำคัญ

1. **Linearity:** ความสัมพันธ์เชิงเส้นระหว่าง predictors และ response
2. **Independence:** ความเป็นอิสระของ observations
3. **Homoscedasticity:** ความคงที่ของ variance
4. **Normality:** การแจกแจงปกติของ residuals (สำหรับ Linear Regression)
5. **No Multicollinearity:** ตัวแปรอิสระไม่มีความสัมพันธ์กันสูง

11.2 การตรวจสอบข้อสมมุติ

Residual Analysis:

$$e_i = y_i - \hat{y}_i \tag{65}$$

$$\text{Standardized Residual} = \frac{e_i}{\sqrt{\text{MSE}(1 - h_{ii})}} \tag{66}$$

โดยที่ h_{ii} คือ leverage values

Durbin-Watson Test สำหรับ Autocorrelation:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{67}$$

12 บทสรุป

เอกสารนี้ได้นำเสนอสมการ Linear Regression ตั้งแต่ระดับพื้นฐานจนถึงขั้นสูง ครอบคลุม:

- **Linear Regression:** Simple และ Multiple
- **Regularized Methods:** Ridge, LASSO, Elastic Net
- **Generalized Linear Models:** GLM framework
- **Polynomial Regression:** สำหรับ non-linear relationships
- **Logistic Regression:** สำหรับ classification
- **Robust Methods:** Huber และ Quantile regression
- **Model Selection:** Cross-validation และ Information criteria

สมการเหล่านี้เป็นพื้นฐานสำคัญในการวิเคราะห์ข้อมูลและการเรียนรู้ของเครื่อง (Machine Learning) ที่นักศึกษาและนักวิจัยควรเข้าใจอย่างถ่องแท้

จบเอกสาร