Linear Regression

2024年9月15日 23:59

Regression model:

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

We have

$$EY = X\beta$$
$$Var\, Y = \sigma^2 I$$

① (Estimate β)

(OLS) $\hat{\beta} = \underset{\beta}{\arg\max}\ \|Y - X\beta\|_2^2$

(MLE) $\hat{\beta} = \underset{\beta}{\arg\max}\ L(\beta) = \underset{\beta}{\arg\max}\ \log(p(Y|X))$

From both:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Then

$$E(\hat{\beta}) = (X^T X)^{-1} X^T X\beta = \beta$$
$$Var(\hat{\beta}) = (X^T X)^{-1} X^T \cdot X (X^T X)^{-1}\ Var(Y) = \sigma^2 (X^T X)^{-1}$$

② (prediction interval)

Difference: prediction of a future value v.s. predict of the mean response

Given value of new data

X_{new}

Predict new response:

$$Y_{new} = X_{new}\beta + \varepsilon \quad \sim N(X_{new}\beta, \varepsilon)$$
$$\hat{Y}_{new} = X_{new}\hat{\beta} \quad \sim N(X_{new}\beta, \sigma^2 X_{new}^T (X^T X)^{-1} X_{new})$$

mean square error to estimate σ^2

Prediction error:

$$Y_{new} - \hat{Y}_{new} \sim N(0, \sigma^2(1 + X_{new}^T (X^T X)^{-1} X_{new}))$$

Then

$$s^2\{pred\} \triangleq MSE(1 + X_{new}^T (X^T X)^{-1} X_{new})$$
$$\frac{Y_{new} - \hat{Y}_{new}}{s^2\{pred\}} \sim t(n-p)$$

$\Rightarrow$ $(1-\alpha)$ percent prediction interval of $Y_{new} = X_{new}\beta + \varepsilon$

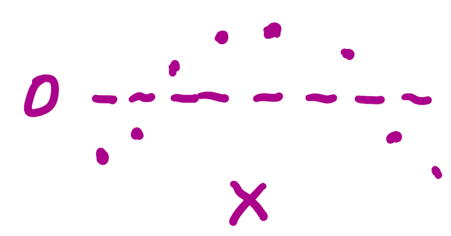
$$\hat{Y}_{new} \pm t(1-(\alpha/2); n-p)\, s\{pred\}$$

In confidence interval for mean of y, not $[1+]$ term

③ (Diagnostics)

- Diagnostic
  - $\to$ predict Y: plot
  - $\to$ base on residual $\{e_i\}$
    - $\to$ L.I.N.E.
      - $\to$ Linearity: e_i v.s. x_i plot
      - $\to$ independent: same
      - $\to$ Normality of ε: QQ plot
      - $\to$ equal variance of ε: Scale location
    - $\to$ Outlier: $e_i^* >$ constant

e_i



× Linear

(Optional) Inference:

$$\begin{cases} e_i = Y_i - \hat{Y}_i \sim N(0, (1-h_{ii})\sigma^2) \\ h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{SS_{xx}} \\ e_i : \text{approximately independent} \end{cases}$$

studentized:

$$\frac{e_i}{s\{e_i\}} = \frac{e_i}{\sqrt{MSE(1-h_{ii})}}$$

semi: $e_i^* = \frac{e_i}{MSE}$

(Remedical)

Non-linear: transform $\begin{cases} \sqrt{\ } \\ \log \\ \text{Box cox} \end{cases}$