Linear regression

Monday, May 27, 2024   21:53

① $y = x\beta + \varepsilon$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $x = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & & & \\ 1 & \vdots & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times p}$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{p\times 1}$ $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$\varepsilon \sim MVN(0, I\sigma^2)$

$Y \sim MVN(\underbrace{X^T\beta}_{n\times 1}, \underbrace{I\sigma^2}_{n\times n})$

求 MLE

likelihood: Y is a $n\times 1$ vector n-dimensional normal dist.

$$\frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left\{-\frac{1}{2}(y-u)^T |\Sigma|^{-1} (y-u)\right\}$$

这里错了, 是 inverse, not det ☆

Σ^{-1} $\begin{bmatrix} \frac{1}{\sigma^2} & & \\ & \frac{1}{\sigma^2} & \\ & & \ddots \frac{1}{\sigma^2} \end{bmatrix}$ n×n diagonal matrix

$$\frac{1}{\sqrt{(2\pi)^n (\sigma^2)^n}} \exp\left\{-\frac{1}{2}(y - X\beta)\cancel{(\sigma^2)^n}(y - X\beta)\right\}$$

$[y_1 \cdots y_i \cdots y_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

log-likelihood:

$$l(\beta, \Sigma) = -\log(2\pi)^{\frac{n}{2}} - n\log\sigma^2 - \frac{1}{\cancel{2(\sigma^2)^n}}\left\{(y - X\beta)^T \Sigma (y - X\beta)\right\}$$

n 是 sum over.

相对于 y^T 里面的每个元素.

$\begin{vmatrix} \frac{1}{\sigma^2} & & \\ & \frac{1}{\sigma^2} & \\ & & \ddots \frac{1}{\sigma^2} \end{vmatrix}$

$$-\frac{1}{\cancel{2(\sigma^2)^n}}\sum_{i=1}^n (y_i - x_i\beta)^2 \cdot -\frac{1}{2}\cdot\frac{1}{\sigma^2}$$

$x_i^T = [x_{i1}, x_{i2}, x_{i3} \cdots x_{ip}]$ the covariate values of i^{th} subject

the whole term is a scalar

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta} = -\frac{1}{2\sigma^2}\sum_{i=1}^n 2(y_i - x_i\beta)\cdot x_i$$

$$\frac{\partial l(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2}\Sigma(y_i - x_i\beta)^2 \cdot \frac{1}{\sigma^3} = 0$$

The term Σ=1nXTX\Sigma = \frac{1}{n} X^T XΣ=n1XTXis used in the context of Fisher information to simplify calculations and because it converges to the true covariance matrix of the predictors as nnnbecomes large. For exact covariance calculations, the predictors should be centered by subtracting their means, and the sample covariance matrix is then 1nX~TX~\frac{1}{n} \tilde{X}^T \tilde{X}n1X~TX~.

再试一遍!!!

$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$ $x = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ \vdots & & & & \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}_{n\times p} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$ $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$ $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ iid

$y = x\beta + \varepsilon \rightarrow E(y) = x\beta$

求 β_{mle}, σ^2_{mle}

① $y \sim MVN(x\beta, \sigma^2 I)$ $\sigma^2 I \to \Sigma$

∴ likelihood, n-dimensional normal.

$$L(\beta, \sigma^2) = \frac{1}{\sqrt{(2\pi)^n \cdot |\Sigma|}} \cdot \exp\left\{-\frac{1}{2}\underbrace{(y - x\beta)^T}_{1\times n}\Sigma^{-1}\underbrace{(y - x\beta)}_{n\times 1}\right\}$$

$|\Sigma| \to (\sigma^2)^n$; $\Sigma^{-1} = \begin{vmatrix} \frac{1}{\sigma^2} & & \\ & \frac{1}{\sigma^2} & \\ & & \ddots \frac{1}{\sigma^2} \end{vmatrix}$

② ∴ log-likelihood:

$$l_n(\beta, \sigma^2) = -n\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i^T\beta)^2$$

错误, 有个 $\sqrt{\ }$, $\sigma^{2n/2} = \sigma^n$

$x_i^T = [x_{i1}\ x_{i2} \cdots x_{ip}] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip}$

这里 x_i 是个 1×p 的行向量, 所以我们写为 x_i^T

③ ∴ score function

$$\frac{\partial l_n(\beta, \sigma^2)}{\partial \beta} = -\frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^n (y_i - x_i^T\underbrace{\beta}_{p\times 1})x_i^T = 0$$

错误: 这里应该是 x_i, 因为 β 是个 p×1 vector, 所以对 $x_i\beta$ 求导后是一个 p×1

$$\frac{\partial l_n(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^3}\sum_{i=1}^n (y_i - x_i^T\beta)^2 = 0$$

错误: 这里 LHS 是对 σ^2 求导, ↳ RHS 却对 σ 进行了求导! 统一为对 σ 求导.

$(\sigma^2)' = -2\frac{1}{\sigma^3}$ 求导错误

$$\frac{\partial l_n(\beta, \sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{2\sigma^3}\Sigma(y_i - x_i\beta)^2 = 0$$

From ①, we have $\sum_{i=1}^n (y_i - \underbrace{x_i^T}_{1\times p}\underbrace{\beta}_{p\times 1})\underbrace{x_i}_{p\times 1} = 0$

$$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \underbrace{x_i^T}_{1\times p}\underbrace{\beta}_{p\times 1}\underbrace{x_i}_{p\times 1} = 0$$

$$\underbrace{x^T}_{p\times n}\underbrace{y}_{n\times 1} - \underbrace{X^T}_{p\times n}\underbrace{X}_{n\times p}\underbrace{\beta}_{p\times 1} = 0$$

$\sum_{i=1}^n y_i x_i$ y_i scalar; x_i: 1×p vector

$X^T y$

$\underbrace{[x_1\ x_2 \cdots x_n]}_{p\times n}\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1}$

$X^T Y = p\times 1$

Tricks: $\Sigma y_i x_i$ 是个 p×1 的 vector, 则应该是 p×sth, sth×1 的两个 stuff 相乘

$\sum_{i=1}^n x_i^T\beta\cdot x_i = p\times 1$

$x_i = p\times 1$ p·sth. sth·p·p×1

$[x_1\ x_2 \cdots x_n] \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} \cdot \beta$

$p\times n = X^T$ $n\times p = X$ tricks: 凑出 X

新的一天, 元气满满出发

$$x^T y = X^T x\beta \Rightarrow \hat\beta = (X^T X)^{-1} X^T y$$ ▨

OLS $\hat\beta = \arg\min_\beta \frac{1}{2}\sum_{i=1}^n (y_i - \langle x_i, \beta\rangle)^2$. In this setting the least squares estimator can be written in closed form as: $\hat\beta = \hat\Sigma^{-1}\left[\frac{1}{n}\sum_{i=1}^n x_i y_i\right]$

$\Sigma = \frac{X^T X}{n}$

$$\frac{\partial l_n(\beta, \sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\Sigma(y_i - x_i\beta)^2 = 0$$

$$n\sigma^2 = \Sigma(y_i - x_i^T\beta)^2 = 0$$

$$\hat\sigma^2 = \frac{1}{n}\Sigma(y_i - x_i^T\hat\beta)^2$$ ▨

一些性质: assume $(X^TX)^{-1}$ exists;

① $\hat\beta \sim N_{p+1}(\beta, \sigma^2(X^TX)^{-1})$

↳ 这里求出来的是方差-协方差矩阵, 所以才能得到每个 parameter 的方差

② $n\hat\sigma^2/\sigma^2 \sim \chi^2(n-p-1)$

③ $\hat\beta$ and $\hat\sigma^2$ are independent

→ in practice $\sigma^2 \Rightarrow \hat\sigma^2 = \frac{1}{n-k}\Sigma(y_i - \hat y_i)^2$

$\boxed{\text{quadratic form thm. } y \sim MVN(\mu_y, \Sigma_y), \text{ then } MVN(A\mu_y, A\cdot\Sigma_y A^T)}$

① proof. $\underbrace{\hat\beta}_{p\times 1} = \underbrace{(X^TX)^{-1}x^T}_{A,\ p\times n}\underbrace{y}_{n\times 1}$ $y \sim MVN(X\beta, \sigma^2 I)$

$\hat\beta$ 是 p×1 vector, 那么 Σ 是 p×p 的 ∴ $A\cdot A^T$ (p×n · n×p)

∴ $\hat\beta \sim N(\underbrace{(X^TX)^{-1}}_{整体}\underbrace{X^TX}_{I}\beta,\ (X^TX)^{-1}X^T\cdot[(X^TX)^{-1}X^T]^T\cdot\sigma^2)$

$(X^TX)^{-1}\underbrace{X^TX(X^TX)^{-1}}_{I}\sigma^2)$

$(AB)^T = B^TA^T$

$(X^T)^T[(X^TX)^{-1}]^T$, use the fact that for inversible matrix, inverse of transpose = transpose of inverse.

$= X[(X^TX)^T]^{-1}$

$= X[X^TX]^{-1}$

$N(\beta,$

$N(\beta, (X^TX)^{-1}\cdot\sigma^2)$

projection of x on y $\frac{x^Ty}{L_x L_y}\cdot\frac{1}{L_y}\cdot y$

② $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (y_i - x_i^T\hat\beta)^2 = \frac{1}{n}\|Y - X\hat\beta\|^2 = \frac{1}{n}\|Y - X\cdot(X^TX)^{-1}X^TY\|$

$= \frac{1}{n}\|(I - X(X^TX)^{-1}X^T)Y\|^2$

Let $\boxed{H = X(X^TX)^{-1}X^T}$

$\hat Y = X\hat\beta$; $\hat Y = \underbrace{\underbrace{X}_{n\times p}\cdot(X^TX)^{-1}\underbrace{X^T}_{p\times n}}_{H,\ n\times n}\cdot\underbrace{Y}_{n\times 1}$.

symmetrix idempotent

① $H^T = H$

② $HH = H$

③ dimension H: $n\times n$

$\boxed{\text{quadratic form } y^Ty = \sum_{i=1}^n y_i^2}$

则最后是一个 scalar. 那就凑 1×sth. · sth·1

$(Y - X\hat\beta)^T(Y - X\hat\beta)$ plug in $\hat\beta$

$= [Y - \underbrace{X(X^TX)^{-1}X^T}_{H}Y]^T(Y - X(X^TX)^{-1}X^TY)$

$= (Y - HY)^T(Y - HY)$ $Y \to \varepsilon$ 为了用 quadratic thm.

$= [X\beta + \varepsilon - H(X\beta + \varepsilon)]^T(X\beta + \varepsilon - H(X\beta + \varepsilon))$

$X(X^TX)^{-1}X^T\cdot X\beta = X\beta$ (i.e., $HX\beta = X\beta$)

$= (\varepsilon - H\varepsilon)^T(\varepsilon - H\varepsilon)$

$= [(I - H)\varepsilon]^T[(I - H)\varepsilon]$

$= \varepsilon^T(I - H)^T(I - H)\cdot\varepsilon$

$= \varepsilon^T(I - H)\varepsilon$

H is symmetric, idempotent
$(I-H)(I-H) = I\cdot I - H - H + H\cdot H$
$= I - 2H + H$
$= I - H$ idempotent
∴ $(I-H)^T = I - H$ symmetric

∴ $\hat\sigma^2 = \frac{1}{n}\varepsilon^T(I - H)\varepsilon$

quadratic theorem 3.
if $y \sim N(0, \sigma^2 I)$, M is symmetric, idempotent,
then $\frac{y^TMy}{\sigma^2} \sim \chi^2(\mathrm{tr}M)$

idempotent matrix: eigenvalues 1 or 0.

$$\frac{n\hat\sigma^2}{\sigma^2} = \frac{n}{\sigma^2}\cdot\frac{1}{n}\cdot\varepsilon^T(I - H)\varepsilon$$

$$= \frac{\varepsilon^T(I - H)\varepsilon}{\sigma^2}$$

• $\varepsilon \sim N(0, \sigma^2 I)$

• $I - H$: symmetric, idempotent

$\sim \chi^2(\mathrm{tr}(I - H))$ trace = Rank X = p/p+1

trace: sum of diagonal elements

sum of eigenvalues.

n×n Sum eigenvalue (A−B) = eigenvalue A (sum) − eigenvalue B (sum)

trace = n   Sum eigenvalue I = n

Sum eigenvalue H (p) = trace H (p) = rank H (p) = rank X (p) ⇐ X = n×p

H is a fun of X

H idempotent, trace = rank

trace(I−H) = n−p ⇐ Sum eigenvalues (I−H) = n−p.

$$\therefore \frac{n\hat\sigma^2}{\sigma^2} \sim \chi^2(n - (p+1))$$

$E(\frac{n\hat\sigma^2}{\sigma^2}) = n - (p+1) \Rightarrow E(\hat\sigma^2) = \frac{n-(p+1)}{n}\cdot\sigma^2$ ∴ $\hat\sigma^2_{mle}$ is biased

∴ $E(\hat\sigma^2)\cdot\frac{n}{n-(p+1)} = \sigma^2 \Rightarrow \frac{n}{n-(p+1)}\cdot\frac{1}{n}\Sigma(y_i - x\hat\beta)^2 = \frac{1}{n-(p+1)}\Sigma(y_i - x\hat\beta)^2$ unbiased.

③

$$\hat\sigma^2 = \frac{1}{n}\Sigma(y_i - x_i^T\beta)^2$$

$$= \frac{1}{n}(y - X\hat\beta)^T(y - X\hat\beta)$$

$Y = HY$

Show $\hat\beta$ and $\hat\sigma^2$ indep ⇒ show $\hat\beta \perp\!\!\!\perp (y - X\hat\beta)$

$\hat\beta = \underbrace{(X^TX)^{-1}X^T}_{A,\ p\times n}\underbrace{y}_{n\times 1}$

转为求 $A\cdot B =$

$Y - X\hat\beta = Y - \underbrace{X\cdot(X^TX)^{-1}X^T}_{H}Y = \underbrace{(I - H)}_{B,\ n\times n}\underbrace{Y}_{n\times 1}$

$A\cdot B = (X^TX)^{-1}X^T\cdot(I - H)$

$= (X^TX)^{-1}X^T - (X^TX)^{-1}X^T\cdot X\cdot(X^TX)^{-1}X^T$

$= 0$ ▨