# Linear regression

Monday, May 27, 2024 21:53

① $y = X\beta + \varepsilon$

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

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

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

求 MLE

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

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

→ 这里错了，是 inverse，not det ☆ $\Sigma^{-1}$

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

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

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

log-likelihood:

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

→ n 次 sum over. 相当于 $y^T$ 里面每一项都乘上 $\frac{1}{\sigma^2}$

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

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

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

→ the whole term is a scalar

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

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

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 $\beta_{mle}$, $\sigma^2_{mle}$

① $y \sim MVN(X\beta, \sigma^2 I)$ → $\Sigma$

∴ likelihood, n-dimensional normal.

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

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

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 √，$\sigma^{2n/2} = \sigma^n$

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

→ 这里 $x_i$ 是个 $1\times p$ 的行向量，所以我们写为 $x_i^T$

③ ∴ score function

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

→ 错误：这里应该是 $x_i$，因为 $\beta$ 是个 $p\times 1$ vector，所以对 $x_i\beta$ 求导后也是一个 $p\times 1$

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

错误：这里 LHS 写对 $\sigma^2$ 求导，→ RHS 却对 $\sigma$ 进行了求导！统一为对 $\sigma$ 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 $\sum_{i=1}^n (y_i - x_i^T\beta)\, x_i = 0$ ($1\times p$, $p\times 1$; $p\times 1$)

$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i^T\beta\, x_i = 0$ ($p\times 1$; $1\times p$ $p\times 1$ $p\times 1$)

$X^Ty - X^TX\beta = 0$ ($p\times n$ $n\times 1$; $p\times n$ $n\times p$ $p\times 1$)

> $\sum_{i=1}^n y_i x_i$ — $y_i$ scalar; $x_i$: $1\times p$ vector
>
> $X^Ty$ ($p\times n$, $n\times 1$)
>
> $[x_1\ x_2 \cdots x_n]_{p\times n} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$
>
> $X^TY = p\times 1$
>
> Tricks: $\Sigma y_i x_i$ 是一个 $p\times 1$ 的 vector，则应该是 $p\times$sth. sth$\times 1$ 的两个 stuff 相乘

> $\sum_{i=1}^n x_i^T\beta\cdot x_i = p\times 1$ ($1\times p$ $p\times 1$ $p\times 1$)
>
> $x_i = p\times 1$ → $p\cdot$sth. sth$\cdot p\cdot p\times 1$
>
> $[x_1\ x_2 \cdots x_n]_{p\times 1} \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}_{1\times p} \cdot \beta$
>
> $p\times n: X^T$ $n\times p = X$ → tricks: 对应 X

新的一天，元气满满出发

$$X^Ty = X^TX\beta \Rightarrow \hat\beta = (X^TX)^{-1}X^Ty \quad ▨$$

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

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

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

---

一些性质：assume $(X^TX)^{-1}$ exists;

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

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

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

> quadratic form thm.
>
> $y \sim MVN(\mu_y, \Sigma_y)$, then $MVN(A\mu_y\ A\cdot\Sigma_y A^T)$

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

$\hat\beta$ 是 $p\times 1$ vector，那么 $\Sigma$ 是 $p\times p$ 的 ∴ $A\cdot A^T$ ($p\times n\cdot n\times p$)

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

整体

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

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

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

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

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

symmetric idempotent

① $H^T = H$

② $HH = H$

③ dimension $H: n\times n$

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

因为最后是一个 scalar，那就是 $1\times$sth. $\cdot$ sth$\cdot 1$

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

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

$= (Y - HY)^T(Y - HY)$ → $Y \rightarrow \varepsilon$ 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., $HX\beta = X\beta$)

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

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

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

> 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

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

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

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

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

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

trace: sum of diagonal elements

sum of eigenvalues.

$n\times n$ Sum eigenvalue (A−B) = eigenvalue A sum − eigenvalue B sum.

trace = n Sum eigenvalue I = n

H is a fun of X

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

H idempotent, trace = rank

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

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

③

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

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

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

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

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

转为求 $A\cdot B = 0$

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

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

$= 0$ ▨