統計解析 II　21〜33

(21)

(a)

$$\frac{P(Y=1\mid x)}{P(Y=0\mid x)} = \exp(\beta_0 + x\beta)$$

ここで $P(Y=1\mid x) = 1 - P(Y=0\mid x)$ より

$$P(Y=1\mid x) = \exp(\beta_0 + x\beta)\{1 - P(Y=1\mid x)\}$$

$$P(Y=1\mid x) = \frac{\exp(\beta_0+x\beta)}{1+\exp(\beta_0+x\beta)}$$

(b) コード

(c) $y_i \in \{-1, 1\}$ の時

$$P(Y=1\mid x) = \frac{1}{1+\exp\{-(\beta_0+x\beta)\}}$$

$$P(Y=-1\mid x) = \frac{1}{1+\exp(\beta_0+x\beta)}$$

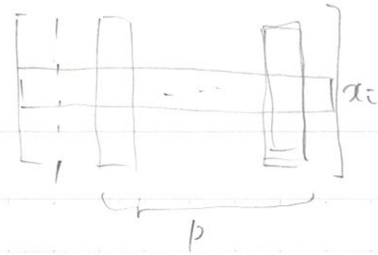
従って $$P(Y=y\mid x) = \frac{1}{1+\exp\{-y(\beta_0+x\beta)\}}$$

尤度

$$\prod_{i=1}^{N} \frac{1}{1+\exp\{-y_i(\beta_0+x\beta)\}}$$

よってLassoの目的関数は

$$\mathcal{L} = -\log L = \frac{1}{N}\sum_{i=1}^{N} \log\left[1+\exp\{-y_i(\beta_0+x\beta)\}\right] + \lambda\|\beta\|_1$$ Ⓠ



(22)

$\mathbb{X} \in \mathbb{R}^{N\times(p+1)}$ デザイン行列

(a) $L(\beta_0, \beta) = \sum_{i=1}^{N} \log\{1+\exp(-y_i(\beta_0+x_i\beta))\}$

$j=1,2,\cdots,p$ の時.

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{N} - \frac{\exp(-y_i(\beta_0+x_i\beta))}{1+\underbrace{\exp(-y_i(\beta_0+x_i\beta))}_{v_i}} y_i x_{ij}$$

$$= -\sum_{i=1}^{N} \frac{v_i y_i}{1+v_i} x_{ij}$$

$$= -\begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{bmatrix}^T \underbrace{\begin{bmatrix} \frac{v_1 y_1}{1+v_1} \\ \vdots \\ \frac{v_N y_N}{1+v_N} \end{bmatrix}}_{u}$$

$j=0$ の時は $x_{0j}=1$ となるので

$$\nabla L = -\mathbb{X}^T u$$

$$z_i = \exp\{-y_i(\beta_0 + x_i\beta)\}$$

(b)

$$\frac{\partial}{\partial\beta_k}\frac{\partial L}{\partial\beta_j} = -\frac{\partial}{\partial\beta_k}\sum_{i=1}^{N} x_{ij}\, y_i \frac{z_i}{1+z_i}$$

$$= -\sum_{i=1}^{N} x_{ij}\, y_i \frac{z_i'(1+z_i) - z_i z_i'}{(1+z_i)^2}$$

ここで $\frac{\partial}{\partial\beta_k} z_i = -z_i\, y_i\, x_{ik}$

$$\frac{\partial^2 L}{\partial\beta_k\partial\beta_j} = \sum_{i=1}^{N} x_{ij}\, x_{ik}\, y_i^2 \frac{z_i}{(1+z_i)^2}$$

$$= \mathbb{X}^T \mathbb{W} \mathbb{X}$$

□

(23)

(a) $N < p$ であり、$\mathrm{rank}(X) = N$ の時.

$$X = \underbrace{\boxed{\quad\quad\quad}}_{p}\}N \in \mathbb{R}^{N\times(p+1)}$$

22より $\nabla L = -\mathbb{X}^T \mathbb{u}$ 局所最適解において $\nabla L = 0$ となるので

find β s.t. $-\mathbb{X}^T \mathbb{u} = 0$ を求めればよい

$$-\mathbb{X}\mathbb{X}^T \mathbb{u} = 0$$

$\mathbb{X}\mathbb{X}^T$ は正則より $\mathbb{u} = \mathbb{0}$ → 有限の β では $\mathbb{u}$ を 0 にできない.

(b) $y_i \exp\{-y_i(\beta_0 + x_i\beta)\} = 0$ $(i=1,2,\dots p)$

この時 $N<p$ より $y_i(\beta_0 + x_i\beta) > 0$ $(i=1,2,\dots p)$ とするような β_0, β の組が存在してそれを何倍もすることでいくらでも $\mathbb{u}$ を 0 に近づけることができるので β_0, β は発散する □

(25) $$P(Y=k, x) = \frac{e^{\beta_{0k} + x\beta^{(k)}}}{\sum_{l=1}^{K} e^{\beta_{0l} + x\beta^{(l)}}} \quad (k=1,\cdots,K)$$

(a)

右辺の指数部に $\gamma_0 + x\gamma$ 引いたものは

$$\frac{e^{\beta_{0k} + x\beta^{(k)} - \gamma_0 - x\gamma}}{\sum_{l=1}^{K} e^{\beta_{0l} + x\beta^{(l)} - \gamma_0 - x\gamma}}$$

$$= \frac{\cancel{e^{-\gamma_0 - x\gamma}}\, e^{\beta_{0k} + x\beta^{(k)}}}{\cancel{e^{-\gamma_0 - x\gamma}} \sum_{l=1}^{K} e^{\beta_{0l} + x\beta^{(l)}}} = P(Y=k, x)$$

(b)

$$\mathbb{B} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pK} \end{bmatrix} \in \mathbb{R}^{p\times K}$$

γ_1が決まる.

$\beta^{(k)}$ クラスkに対する重み.

$\mathbb{B}$が決定した際に

$\beta_{j,1},\cdots,\beta_{j,K}$ に対して中央値を計算し,

$$\gamma_j = \mathrm{median}(\beta_{j1} \cdots \beta_{jK}) \quad (j=1,\cdots,p)$$

として $k=1,2\cdots,K$ に対して

$$\beta_{jk} \leftarrow \beta_{jk} - \gamma_j$$ と更新すれば

目的関数を最小にすることができる

→ L1正則化を入れることで最適値が一意に決まる.

(c)

$\beta_{01} \cdots \beta_{0K}$ が求まってから

$$\beta_{0i} \leftarrow \beta_{0i} - \frac{1}{K}\sum_{k=1}^{K} \beta_{0k}$$ と更新を行う.

(27)

$$L := -\frac{1}{N}\sum_{i=1}^{N}\sum_{h=1}^{K} I(y_i = h) \log \frac{\exp}{\sum_{\ell=1}^{K} \exp\{\beta_{0\ell} + x_i \beta^{(\ell)}\}}$$

↗ 示指関数

$$\frac{\partial^2 L}{\partial \beta_{jk} \partial \beta_{j'k'}} = \begin{cases} \sum_{i=1}^{N} x_{ij} x_{ij'} \pi_{ik}(1-\pi_{ik}) & k = k' \\ \sum_{i=1}^{N} x_{ij} x_{ij'} \pi_{ik} \pi_{ik'} & k \neq k' \end{cases}$$

$$= \sum_{i=1}^{N} x_{ij} x_{ij'} \underline{w_{ikk'}}$$

$W_i \in \mathbb{R}^{K \times K}$ が非負定値であることを示す.

$$\pi_{ik} = \frac{\exp(\beta_{0k} + x_i \beta^{(k)})}{\sum_{\ell=1}^{K} \exp(\beta_{0\ell} + x_i \beta^{(\ell)})}$$

$$w_{ikk} = \pi_{ik}(\underline{1-\pi_{ik}})$$

k 以外の確率.

$$= \pi_{ik} \sum_{k' \neq k} \pi_{ik'}$$

$$= \sum_{k' \neq k} |w_{ikk'}|$$

Gershgorin の定理より

W_i は非負定値となる.

(28) $P(Y=k) = \frac{\mu^k}{k!} e^{-\mu} \quad (k=0,1,2\cdots)$

$\mu = E[Y]$ が $x \in \mathbb{R}^p$ に依存して

$$\mu(x) = E[Y \mid X=x] = e^{\beta_0 + x\beta} \quad (x \in \mathbb{R}^p)$$

(a)

尤度は $\prod_{i=1}^{N} \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i}$ 　　　y_iは定数より無視.

negative log likelihood

$$\min \quad L \equiv -\sum_{i=1}^{N} \{ y_i(\beta_0 + x_i\beta) - e^{\beta_0 + x_i\beta} \}$$

L1正則化項

$$\frac{1}{N} L(\beta_0, \beta) + \lambda \|\beta\|_1$$

(b)

$$\nabla L = \frac{\partial L}{\partial \beta_j} = -\left\{ \sum_{i=1}^{N} y_i x_{ij} - x_{ij} e^{\beta_0 + x_i\beta} \right\}$$

$$= -\sum_{i=1}^{N} x_{ij} \left(y_i - e^{\beta_0 + x_i\beta} \right)$$

デザイン行列 $\tilde{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Np} \end{pmatrix}$ 　$u = \begin{pmatrix} y_1 - e^{\beta_0 + x_1\beta} \\ \vdots \\ y_N - e^{\beta_0 + x_N\beta} \end{pmatrix}$ として

$\nabla L = \tilde{X}^T u$ 　と表せる　$\boxed{//}$

(c)

$$\nabla L_j = -\sum_{i=1}^{N} x_{ij}\left(y_i - e^{\beta_0 + x_i\beta}\right)$$ より

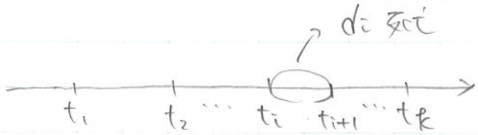
$$\frac{\partial}{\partial \beta_k}\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{N} x_{ij}\, x_{ik}\, e^{\beta_0 + x_i\beta}$$

従って

$$\mathbb{W} = \begin{pmatrix} e^{\beta_0 + x_1\beta} & & 0 \\ & \ddots & \\ 0 & & e^{\beta_0 + x_N\beta} \end{pmatrix}$$ を用いて

$$\nabla^2 L = \tilde{\mathbb{X}}^T \mathbb{W} \mathbb{X}$$ と表せる.

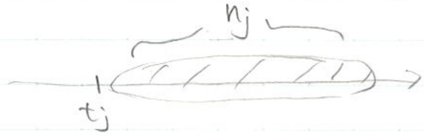
30



$[t_i, t_{i+1})$ で死亡者数 d_i

$$D = \sum_{j=1}^{k} d_j$$ --- 総死亡者数

$[t_j, t_{j+1})$ での打ち切り m_j と書く.



t_j までの生存者は

$$n_j = \sum_{i=j}^{k} (d_i + m_i) \quad (j = 1, 2, \dots, k)$$

生存時間 T が t より大きい確率 $S(t)$ は、 $t_\ell \leq t < t_{\ell+1}$ の時

$$\hat{S}(t) = \begin{cases} 1 & (t < t_1) \\ \prod_{i=1}^{\ell} \frac{n_i - d_i}{n_i} & (t \geq t_1) \end{cases}$$

$$n_i = \sum_{j=i}^{k} (d_j + m_j)$$

$[t_\ell, t_{\ell+1})$ で打ち切りがない時 $n_\ell = d_\ell$ より

$$n_i - d_i = \sum_{j=i}^{k} d_j - d_i = n_{i+1}$$

$t \geq t_1$ の時

$$\frac{\cancel{n_2}}{n_1} \cdot \frac{n_3}{\cancel{n_2}} \cdots \frac{n_{\ell+1}}{\cancel{n_\ell}} = \frac{n_{\ell+1}}{N}$$

$$\hat{S}(t) = \begin{cases} 1 & (t < t_1) \\ \frac{n_{\ell+1}}{N} & (t \geq t_1) \end{cases}$$

(31) 部分尤度関数

$$\max \prod_{i:\delta_i=1} \frac{e^{x_i\beta}}{\sum_{j\in R_i} e^{x_j\beta}}$$

$\delta_i=1$ ··· 死亡

$\delta_i=0$ ··· 調査打ち切り

R_i ··· y_i 以上の y_j の添字集合

↓ L1正則化 NNL

$$L := -\frac{1}{N}\sum_{i:\delta_i=1} \log \frac{e^{x_i\beta}}{\sum_{j\in R_i} e^{x_j\beta}} + \lambda\|\beta\|_1$$

(a)

$j\in R_i \quad \delta_i=1 \iff i\in C_j$ とするとき

$S_i = \sum_{h\in R_i} e^{x_h\beta}$ とする

$$\frac{\partial L}{\partial \beta_k} = -\sum_{i:\delta_i=1} \frac{\cancel{\sum_{j\in R_i} e^{x_j\beta}}}{e^{x_i\beta}} \frac{x_{jk}e^{x_j\beta} - \sum_{j\in R_i} x_{jk}e^{x_j\beta}}{\{\sum e^{x_j\beta}\}^2}$$

$$= -\sum_{i:\delta_i=1}\left\{x_{jk} - \frac{\sum_{j\in R_i} x_{jk}e^{x_j\beta}}{\sum_{j\in R_i} e^{x_j\beta}}\right\}$$

$$= -\sum_{i:\delta_i=1}\left\{x_{jk} - \frac{\sum_{j\in R_i} x_{jk}e^{x_j\beta}}{S_i}\right\}$$

$$= -\sum_{i:\delta_i=1} x_{jk} + \sum_{j=1}^{N}\sum_{i\in C_j} \frac{x_{jk}e^{x_j\beta}}{S_i}$$

$$= -\sum_{i=1}^{N} x_{ik}\left\{\delta_i - \sum_{j\in C_i}\frac{e^{x_i\beta}}{\sum_{h\in R_j} e^{x_h\beta}}\right\} \quad \square$$

(b)

$$\frac{\partial^2 L}{\partial \beta_k \partial \beta_l} = \sum_{i=1}^{N} x_{ik} \sum_{j \in C_i} \frac{\partial}{\partial \beta_l} \left(\frac{e^{x_i \beta}}{\sum_{h \in R_j} e^{x_h \beta}} \right)$$

$$= \sum_{i=1}^{N} x_{ik} \sum_{j \in C_i} \frac{1}{\left(\sum_{h \in R_j} e^{x_h \beta}\right)^2} \left\{ x_{il} e^{x_i \beta} \sum_{s \in R_j} e^{x_s \beta} - e^{x_i \beta} \sum_{h \in R_j} x_{hl} e^{x_h \beta} \right\}$$

$$= \sum_{i=1}^{N} \sum_{h=1}^{N} x_{ik} x_{hl} \sum_{j \in C_i} \frac{e^{x_i \beta}}{\left(\sum_{r \in R_j} e^{x_r \beta}\right)^2} \left\{ I(i=h) \sum_{s \in R_j} e^{x_s \beta} - I(h \in R_j) e^{x_h \beta} \right\}$$

$$\frac{\partial^2 L}{\partial \beta_k \partial \beta_l} = \sum_{i=1}^{N} \sum_{h=1}^{N} x_{ik} x_{hl} w_{ih}$$

$$w_{ii} = \sum_{j \in C_i} \frac{e^{x_i \beta}}{\left(\sum_{r \in R_j} e^{x_r \beta}\right)^2} \left\{ \sum_{s \in R_j} e^{x_s \beta} - I(i \in R_j) e^{x_h \beta} \right\}$$

∎