Информация фишера

$$I_x(\theta) = D_\theta \, U_x(\theta) \,, \quad \text{где} \quad U_x(\theta) = \frac{\partial L_x(\theta)}{\partial \theta}$$

Многомерный случай

$\theta \in \mathbb{R}^d$

$$I_x(\theta)_{ij} = E_\theta \left(\frac{\partial L_x(\theta)}{\partial \theta_i} \; \frac{\partial L_x(\theta)}{\partial \theta_j} \right) = - E_\theta \, \frac{\partial^2 L_x(\theta)}{\partial \vartheta_i \, \partial \vartheta_j}$$

Пример $\quad X_1, \dots, X_n \sim \mathcal{N}(a, \sigma^2) \qquad I_X(a, \sigma^2) - ?$

Замечание [L1-L9] $\hat\theta$ – ОМП для θ с ас. матр. ков. $i_X^{-1}(\theta) = I_{X_1}^{-1}(\theta)$

$$L_X(\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\sum -\frac{1}{2}\frac{(x_i-a)^2}{\sigma^2}} \qquad l_X(\theta) = \frac{n}{2}\log 2\pi - n\log\sigma - \sum\frac{(x_i-a)^2}{2\sigma^2}$$

$$\frac{\partial l_X(\theta)}{\partial a} = \sum\frac{x_i-a}{\sigma^2} \qquad \frac{\partial^2 l_X(\theta)}{\partial a\,\partial a} = -\frac{n}{\sigma^2}$$

$$\frac{\partial l_X(\theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \sum\frac{(x_i-a)^2}{2\sigma^4} \qquad \frac{\partial^2 l_X(\theta)}{\partial a\,\partial\sigma^2} = \sum\frac{a-x_i}{\sigma^4} = \frac{\partial^2 l_X(\theta)}{\partial\sigma^2\,\partial a}$$

$$\frac{\partial^2 l_X(\theta)}{\partial\sigma^2\,\partial\sigma^2} = \frac{n}{2\sigma^4} - \sum\frac{(x_i-a)^2}{\sigma^6}$$

$$I_x(\theta) = -E_\theta\left[\frac{\partial^2 l_x(\theta)}{\partial\theta_i\,\partial\theta_j}\right]_{ij} = -\begin{pmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} - \frac{n}{\sigma^4} \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

$$\Rightarrow \quad i(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

$$i^{-1}(\theta) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$$

$(\overline{X},\ \overline{X^2} - \overline{X}^2)$ — а.н.о. (a, σ^2) с ас. дисп. $\begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$

Обобщенная линейная модель

$y = \mu_\theta(x)$ — ожидаемый отклик огр на g: $\exists g^{-1}$

$g(\mu_\theta(x)) = x^T\theta$ — линеаризация

$y_i \sim P_{\mu_\theta(x_i)}$ — наблюдаемый отклик

$\hat{\theta}$ — по гаусс. модели для $g(y)$

предсказания: $\hat{y} = g^{-1}(x^T\hat{\theta})$

Логистическая регрессия

$y_i \sim \mathrm{Bern}(\sigma(x^T\theta))$ $\qquad x \in \mathbb{R}^{d \times 1}$

g - линеаризация - логит-функция (σ^{-1})

Считаем $I_x(\theta)$

$$l_x(\theta) = \sum_{i=1}^{n} \left[y_i \log \sigma(x_i^T\theta) + (1-y_i) \log (1-\sigma(x_i^T\theta)) \right]$$

$$\frac{\partial l_x(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left(y_i - \sigma(x_i^T\theta) \right) x_i$$

$$\frac{\partial^2 l_x(\theta)}{\partial \theta^2} = \sum_{i=1}^{n} x_i \left(\sigma(x_i^T\theta)(1-\sigma(x_i^T\theta)) \right) x_i^T \in \mathbb{R}^{d \times d}$$

$$= -X^T V(\theta) X = -I_x(\theta)$$

$$V(\theta) = \mathrm{diag}\left(\sigma(x_i^T\theta)(1-\sigma(x_i^T\theta)) \right)$$

Пуассоновская регрессия

$$y_i \sim \mathrm{Pois}(\lambda_\theta(x_i))$$

$$\lambda_\theta(x_i) = e^{x_i^T\theta}$$

g – линеаризация $=\log$

$$L_x(\theta) = e^{-\sum \lambda_\theta(x_i)} \prod \frac{\lambda_\theta(x_i)^{y_i}}{y_i!}$$

$$l_x(\theta) = -\sum_{i=1}^{n} \lambda_\theta(x_i) + \sum_{i=1}^{n} \left[y_i \log \lambda_\theta(x_i) - \log y_i! \right] =$$

$$= \sum_{i=1}^{n} \left[y_i (x_i^T\theta) - \log y_i! - e^{x_i^T\theta} \right]$$

$$\frac{\partial L_x(\theta)}{\partial \theta} = \sum_{i=1}^{n} y_i x_i^T - x_i^T e^{x_i^T \theta}$$

$$\frac{\partial L_x^2(\theta)}{\partial \theta^2} = -\sum_{i=1}^{n} x_i e^{x_i^T \theta} x_i^T = -X^T E(\theta) X$$

$$E(\theta) = \operatorname{diag}\left(e^{x_i^T \theta} \right)$$

Итого:

Логистическая $I_x(\theta) = X^T V(\theta) X$

Пуассоновская $I_x(\theta) = X^T E(\theta) X$

- Ас. дов. инт. для θ_j $\qquad \left(\hat{\theta}_j \pm z_{1-\alpha/2}\sqrt{I_x^{-1}(\hat{\theta})_{jj}}\right)$

- $H_0: \theta_j = 0$ — незначимость j-ого пр-ка

$$T_j^\circ(X,Y) = \frac{\hat{\theta}_j}{\sqrt{I_x^{-1}(\hat{\theta})_{jj}}} \xrightarrow{H_0} N(0,1)$$

- Предсказательный дов. интервал

$$\sigma(x_{new}^T\theta) \in \left(\sigma(x_{new}^T\hat{\theta} - \hat{\delta}),\ \sigma(x_{new}^T\hat{\theta} + \hat{\delta})\right)$$

$$\hat{\delta} = z_{1-\alpha/2}\sqrt{x_{new}^T I_x^{-1}(\hat{\theta})\, x_{new}}$$

← это для логрег. можно заменить σ на μ для других

Калибровка

- Гистограмная

$$B = [0,1] = \bigsqcup_{i=1}^{k} B_i$$

$P(x_i)$ – предсказанные вероятности

q_i – выдаваемые вероятности

не можем поставить значения дробей, т.к. надо для кросс-валидации что-то придумывать

$$\sum_{j=1}^{k} \left| \frac{\sum_{i=1}^{n} \overset{\text{valid}}{y_i}\, I\{\overset{\text{train}}{p}(x_i) \in B_j\}}{\sum_{i=1}^{n} I\{p(x_i) \in B_j\}} - q_j \right| \to \min_{q_1 \dots q_k}$$

← можно взять $(\dots)^2$ вместо $|\dots|$

- Изотоническая

обычно её используют

1) хотим $q_1 \leq \ldots \leq q_n$

2) оптимиз. кол-во бинов

$$\sum_{i=1}^{n} \left(y_i - g(p(x_i))\right)^2 \to \min_{g}$$

g - кусочно лин

есть какой-то алгос, подбирающий эту штуку

• Калибровка Платта

$$P_{x_i}(y_i = 1) = \sigma(a\, p(x_i) + b)$$

Хотим $\hat{a}, \hat{b}$ => ищем ОМП по валидации

$$\sum_{i=1}^{n} y_i \log \sigma(a\, p(x_i) + b) + (1-y_i) \log(1-\sigma(a\, p(x_i)+b)) \to \max_{a,b}$$

Поправка (для не переобучения на валидации):

← про поправку можно умолчать на экзамене

$$y_i \to t_0 := \frac{1}{\#\{i : y_i = 0\} + 2}$$

$$1 - y_i \to t_1 := \frac{\#\{i \mid y_i = 1\}}{\#\{i \mid y_i = 0\} + 2}$$