Линейная регрессия

Что мы уже знаем?

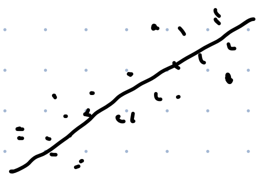
$$y = X \cdot w$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \begin{bmatrix} 1 & x_{11} & \dots & x_{1d} \\ 1 & x_{21} & \dots & x_{2d} \\ 1 & x_{31} & \dots & x_{3d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nd} \end{bmatrix} \quad \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i w)^2$$

$$\frac{1}{n} (y - Xw)^T (y - Xw) \to \min_w$$

$$\hat{w} = (X^T X)^{-1} X^T y$$



$$MSE + \lambda \cdot R(w)$$

$$\|w\|_2 = \sum_{j=1}^{d} w_j^2 \quad \text{Ridge}$$

$$\|w\|_1 = \sum |w_j| \quad \text{Lasso}$$

модель
Что для меня важно?

я хочу знать, что будет завтра

прогнозы

y

несм.
сост.
эф.

я хочу понять, что за фигня произошла сегодня

w

нужны чтобы w_1 был классным

$$y = w_0 + w_1 x + w^T z$$

переменная воздействия

контрольные переменные

→ ЦБ г П

→ оффлайн АБ-тест

магазины
ПЕРЕКРЁСТОК

→ treatment (акция)

→ control

$$y = w_0 + w_1 \cdot d_i + w_2 x_1 + w_3 x_2 + \ldots + w_d x_d$$

$$d_i = \begin{cases} 1, & \text{treatment} \\ 0, & \text{control} \end{cases}$$

$$\boxed{\begin{aligned} H_0 &: w_1 = 0 \\ H_A &: w_1 \neq 0 \end{aligned}}$$

MSE – ф. потерь

Т. Гаусса-Маркова

→ X – детерминированные

→ $y = Xw + \varepsilon$

→ $\mathbb{E}(\varepsilon) = 0$

$Var(\varepsilon) = \begin{bmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \sigma^2 & \\ 0 & & & \ddots\ \sigma^2 \end{bmatrix}$

$Var(\varepsilon_i) = \sigma^2$ гомоскедастичность

гетероскедастичность

$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$

$\varepsilon_1, \dots, \varepsilon_n \sim iid\ (0, \sigma^2)$

→ X – случайные

→ $y = Xw + \varepsilon$

→ $(x_i, y_i) \sim iid$

$(x_i, \varepsilon_i) \sim iid$ → σ^2, $cov(\varepsilon_i, \varepsilon_j) = 0$

→ $\mathbb{E}(x_i^4) < \infty$, $\mathbb{E}(y_i^4) < \infty$ – нет выбросов

→ $Cov(x_i, \varepsilon_i) = 0$

$\mathbb{E}(\varepsilon_i | x_i) = 0$

Экзогенность

Эндогенность.

ошибка

$y_i = w x_i + \underset{\sim}{\varepsilon_i}$

ЗП

„талант" → $\underset{\sim}{z_i} + u_i$

$Cov(x_i, z_i)$

$\Rightarrow Cov(x_i, \varepsilon_i)$

$\hat{w}$ – не состоятельная

$\Rightarrow \hat{w} = (X^TX)^{-1}X^Ty = \sum_{i=1}^{n} y_i \cdot \gamma_i$

BLUE

- сост.
- эфф.
- несмещ.

не нужна

+ $\varepsilon_i \sim N(0, \sigma^2)$ BUE

$\hat{w} \overset{asy}{\sim} N(w;\ \sigma^2 \cdot (X^TX)^{-1})$

→ д.и.

→ гипотезы

Big Data

$\dfrac{\hat{w} - w_0}{se(\hat{w})} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$

Если Small Data, $\varepsilon_i \sim N(0, \sigma^2)$ $\quad \dfrac{\hat{w} - w_0}{se(\hat{w})} \underset{H_0}{\sim} t(n-d)$

BVD:

$$MSE = \underset{\overset{\|}{0}}{bias^2} + Var + \sigma^2$$

самый низкий для bias = 0 — т.г.м.

Регуляризатор: $bias^2 \uparrow \quad Var \downarrow$

- Lasso
- Ridge

$\hat{w}$ – смещён

$\Rightarrow$ все предпосылки рушатся.!!

Ослабляем предпосылки:

гетероскед.

$$\frac{\hat{w}_j - w_j}{se_{HC}(\hat{w}_j)} \xrightarrow[n/k \to \infty]{d} N(0, 1)$$

$$\sigma^2 \cdot I_n \downarrow \hat{\Omega}$$

$$\hat{Var}(\hat{w}|X) = (X^T X)^{-1} X^T \hat{\Omega} X (X^T X)^{-1}$$

Heteroskedasticity Consistent

$$Var(\varepsilon | X) = \Omega = diag(\sigma_1, \dots, \sigma_n) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \; \sigma_n^2 \end{bmatrix}$$

$$\mathbb{E}(\varepsilon_i | X) = 0$$

$$Var(\varepsilon_i | X) = \mathbb{E}(\varepsilon_i^2 | X) = \sigma_i^2$$

$$\hat{\varepsilon}_1 = y_1 - \underbrace{x_1 \cdot \hat{w}}_{\hat{y}_1}$$

$\hat{\sigma}_1^2 = \hat{\varepsilon}_1^2$ наивная оценка

HC0 (White):

$$\hat{\Omega}_{HC0} = \begin{bmatrix} \varepsilon_1^2 & & 0 \\ & \varepsilon_2^2 & \\ 0 & & \ddots \; \varepsilon_n^2 \end{bmatrix}$$

НС 3:

y_{42} ε_{42}

$y = wX$ по всей выборке без 42-го.

$\hat{y}_{42} = \hat{w}X$

$\hat{\varepsilon}^{cv}_{42} = y_{42} - \hat{w}X$

Такая коррекция быстрее сх. к $N(0,1)$

$$\hat{\Omega}_{HC2} = \begin{bmatrix} (\hat{\varepsilon}^{cv}_1)^2 & & 0 \\ & \ddots & \\ 0 & & (\hat{\varepsilon}^{cv}_n)^2 \end{bmatrix}$$

Автокорреляция

$Cov(\varepsilon_i, \varepsilon_j)$ $\Rightarrow$ Поправочка

$\parallel$

$\mathbb{E}(\varepsilon_i \cdot \varepsilon_j) \Rightarrow \hat{Cov}(\varepsilon_i, \varepsilon_j) = \hat{\varepsilon}_i \cdot \hat{\varepsilon}_j$

АБ:

$y_i = w_0 + w_1 d_i$

$y_i = \hat{w}_0$ $\quad d_i = 0$ $\quad \hat{w}_0 = \bar{y}_c$

$y_i = \hat{w}_0 + \hat{w}_1$ $\quad d_i = 1$ $\quad \hat{w}_0 + \hat{w}_1 = \bar{y}_T$

$H_0: w_1 = 0$ $\Leftrightarrow$ $H_0: \mu_T = \mu_c$

$H_A: w_1 \neq 0$ $\quad H_A: \mu_T \neq \mu_c$