Lecture September 23

Intercept and OLS

$$C(\beta) = \frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j\right)^2$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \beta_0 - \sum_{j=1}^{p-1} \underline{x_{ij}}\beta_j\right)^2$$

Assume

$$\tilde{y}_i = \beta_0 + \beta_1 x_i$$

$$C(\beta) = \frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

$$\frac{\partial C}{\partial \beta_0} = 0 = -\frac{2}{n}\sum_{i=0}^{n-1}\left(y_i - \beta_0 - \beta_1 x_i\right)$$

$$\sum_{i=0}^{n-1}\beta_0 = n\beta_0$$

$$= \sum_{i=0}^{n-1} y_i - \beta_1\sum_{i=0}^{n-1} x_i$$

$\Rightarrow$

$$\beta_0 = \frac{1}{n}\sum_{i=0} y_i - \frac{\beta_1}{n}\sum_{i=0} x_i$$

$$\boxed{\mu_y - \frac{\beta_1}{n}\sum_{i=0}^{n-1} x_i}$$

$$X \to \bar{X} = X - \mathbb{E}[X]$$

$$\bar{X}_{ij} = X_{ij} - \mu_j$$

mean of column j

$$\mu_j = \frac{1}{n}\sum_{i=0}^{n-1} X_{ij}$$

Simpler case with only x_i

$$X = \begin{bmatrix} 1 & x_0^1 \\ 1 & x_1^1 \\ \vdots & \vdots \\ 1 & x_{n-1}^1 \end{bmatrix}$$

$$\mu_x = \frac{1}{n}\sum_{i=0}^{n-1} x_i$$

$$\sum_{i=0}^{n-1} x_i \Rightarrow \sum_{i=0}^{n-1} (x_i - \mu_x)$$

$$\beta_0 = \mu_y - \frac{\beta_1}{n} \left(\sum_{i=0}^{n-1} (x_i - \mu_x) \right)$$

$$\boxed{\beta_0 = \mu_y}$$

Classification

$$y_i = \begin{cases} 1 & \text{TRUE} \\ 0 & \text{FALSE} \end{cases}$$

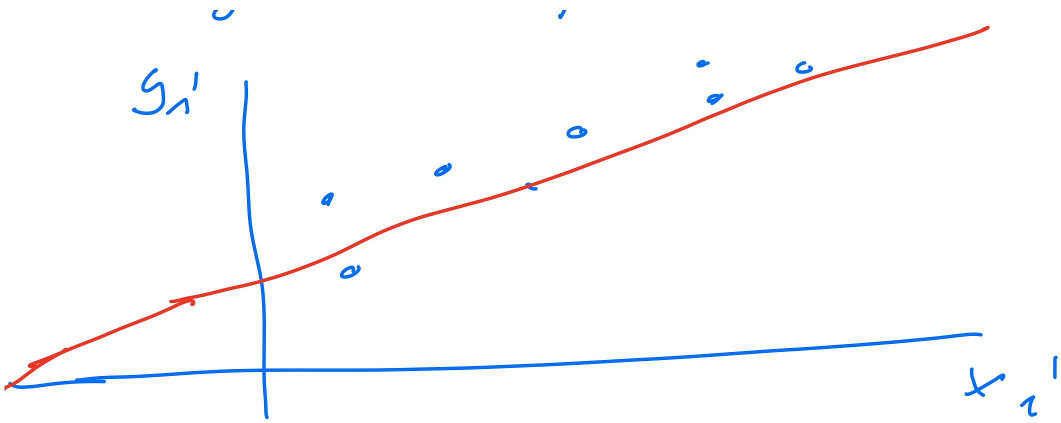
Binary output/target

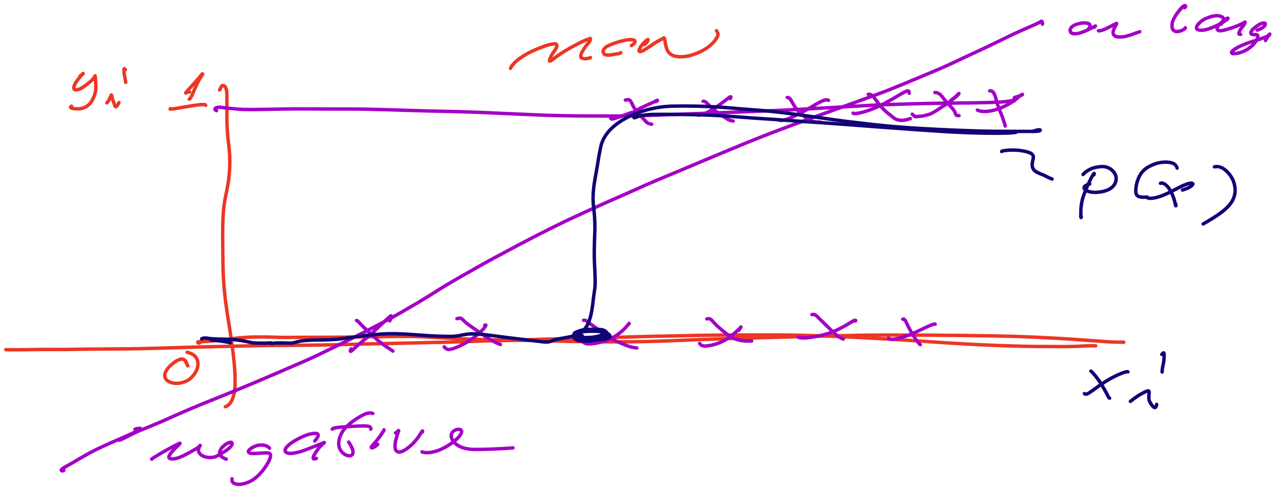
$x_i \Rightarrow y_i$ Discrete variables

Till now

$$y_i = f(x_i) + \varepsilon_i$$

$$f(x_i) \simeq \beta_0 + \beta_1 x_i$$





$\boxed{f(x)} \longrightarrow p(x)$

if $p(x) \leq 0.5$ then $\tilde{y}_i = 0$

if $p(x) \geq 0.5$ then $\tilde{y}_i = 1$

$0 \leq p(x) \leq 1$

$p(x)$ can be treated as a probability.

$\sum p(x_i) = 1$

$$\sum_{i \in D}$$

$$p(y_i = \underline{1} \mid x_i) = p$$

$$p(y_i = 0 \mid x_i) = 1 - p$$

$$p(y_i = 0 \mid x_i) + p(y_i = 1 \mid x_i) = \underline{1}$$

$$y_i = f(x_i) + \varepsilon_i$$

now $\quad y_i = p(x_i) + \varepsilon_i$

$$y_i = \underline{1} = \cancel{p(x_i)} + \varepsilon_i \quad \Rightarrow$$

$\varepsilon_i = \underline{1 - p}(x_i)$ with probability p

$$y_i = 0 = p + \varepsilon_i \Rightarrow$$

$\varepsilon_i = -P(x_i)$ with probability $1 - P$

$$\mathbb{E}[\varepsilon] = \sum_i p_i \varepsilon_i$$

$$= \underbrace{(1-p)p}_{y_i = \underline{1}} + \underbrace{p(1-p)}_{y_i = 0}$$

$$= \dots$$

$= 0$

$$\text{var}[\varepsilon] = (1-p)^2 p + (-p)(1-p)$$
$$= p(1-p)$$

$\varepsilon \sim$ Binomial distribution

$$D = \{ [x_0, y_0], [x_1, y_1] \ldots, [x_{n-1}, y_{n-1}] \}$$

$$P(D|\beta) = \prod_{i=0}^{n-1} \underline{p(x_i|\beta)^{y_i} (1-p(x_i|\beta))^{1-y_i}}$$

$y_i = 1$ then we have $p(x_i|\beta)$ is $y_i = 1$

$y_i = 0$ then we have $1 - p(x_i|\beta)$ $y_i = 0$

Linear Regression

$$p(D|\beta) = \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y_i - x_{i*}\beta)^2}{2\sigma^2}\right)}$$

$$C(\beta) = -\log P(D|\beta)$$

we want β which maximizes the likelihood.

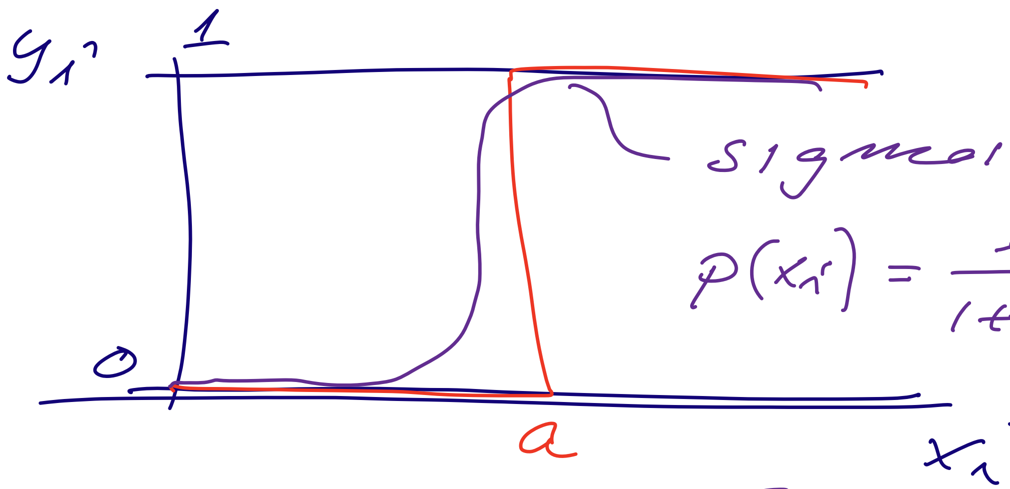
$$p(D|\beta) = \prod_{i=0}^{n-1} \underbrace{p(x_i|\beta)}_{p_i}{}^{y_i} \left(\underbrace{1 - p(x_i|\beta)}_{1-p_i}\right)^{1-y_i}$$

$$y_i = [0, 1]$$

$$C(\beta) = -\log P(D|\beta)$$

$$= -\sum_{i=0}^{n-1} \left\{ y_i \log p_i + (1-y_i)\log(1-p_i) \right\}$$

$$p_i = p(x_i \mid \beta) =:$$



$$p(x_i) = \frac{e^{x_i}}{1+e^{x_i}}$$

$$p(x_i \mid \beta) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

$$1 - p(x_i \mid \beta) = \frac{1}{1+e^{\beta_0+x_i}}$$