

Linear Regression

Data $(X_1, Y_1), \dots, (X_n, Y_n)$

If $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ where

$\varepsilon_i | X_i \sim N(0, \sigma^2)$, then

MLE estimates for β_0, β_1 are

solution

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (\beta_0 + \beta_1 X_i - Y_i)^2$$

Sol:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$$

Multi-variable linear regression

$$(\vec{X}_1, y_1) \dots (\vec{X}_n, y_n)$$

$$\vec{X}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{bmatrix}$$

$$L(\hat{y}, y) = (\hat{y} - y)^2$$

$$r_{\beta}(\vec{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

$$R_n(\beta) = \frac{1}{n} \sum_{i=1}^n L(r_{\beta}(X_i), y_i)$$

$$= \frac{1}{n} \sum_{i=1}^n (\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} - y_i)^2$$

recall, $\|\vec{z}\| = \sqrt{z_1^2 + \dots + z_n^2}$

write $R_n(\beta)$ in terms of norm
of some vector

$$= \frac{1}{n} \left\| \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \dots + \beta_k X_{1k} - y_1 \\ \beta_0 + \beta_1 X_{21} + \dots + \beta_k X_{2k} - y_2 \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \dots + \beta_k X_{nk} - y_n \end{bmatrix} \right\|^2$$

$$= \frac{1}{n} \left\| \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \dots + \beta_k X_{1k} \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \dots + \beta_k X_{nk} \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|^2$$

$$= \frac{1}{n} \left\| \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|^2$$

$\underbrace{\hspace{15em}}_X \quad \underbrace{\hspace{5em}}_\beta \quad \underbrace{\hspace{5em}}_y$

$$= \frac{1}{n} \| X\beta - y \|^2$$

What about non-linearities?

$$L(\hat{y}, y) = (\hat{y} - y)^2$$

$$r_{\beta}(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Write down $R_n(\beta) = \frac{1}{n} \sum_{i=1}^n L(r_{\beta}(x_i), y_i)$
as a matrix problem

Logistic regression

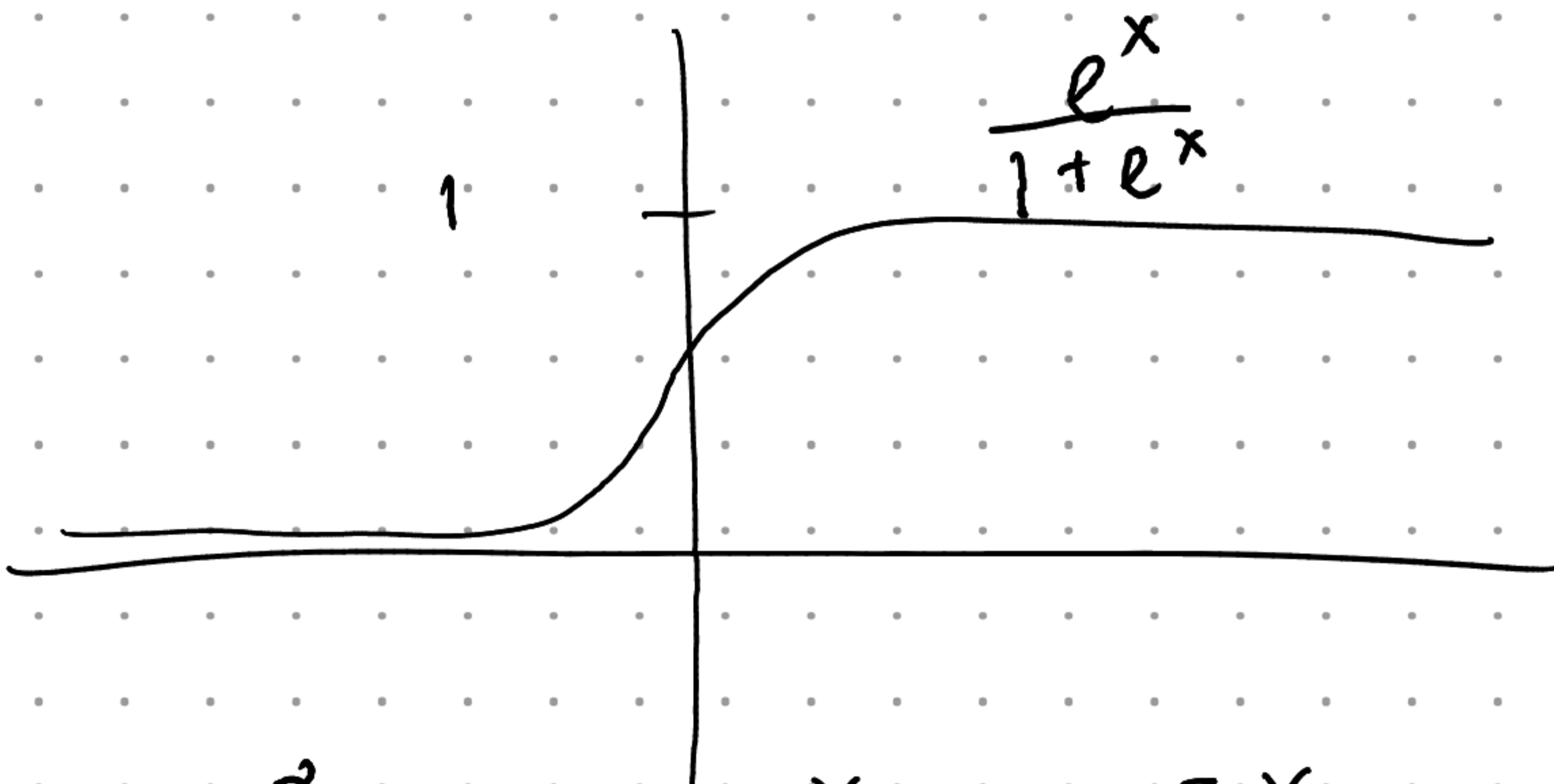
Y_i are 0 or 1

Given \vec{X}_i , we want to output

$P_i \in (0,1)$ describing the "probability" $Y_i = 1$

Assume statistical model

$$P_i(\vec{\beta}) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})}$$



Find $\vec{\beta}$ st $\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}$ big if $Y_i = 1$
small if $Y_i = 0$

