Linear Regression

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

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

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

MLE estimate for β_0, β_1 are solution to

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

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-variate linear regression

$(\vec{X}_1, Y_1) \quad \cdots \quad (\vec{X}_n, Y_n)$

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

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

$$r_\beta(\vec{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

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

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

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

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

$$= \frac{1}{n} \left\| \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \cdots + \beta_k X_{1k} - Y_1 \\ \beta_0 + \beta_1 X_{21} + \cdots + \beta_k X_{2k} - Y_2 \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \cdots + \beta_k X_{nk} - Y_n \end{bmatrix} \right\|^2$$

$$= \frac{1}{n} \left\| \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \cdots + \beta_k X_{1k} \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \cdots + \beta_k X_{nk} \end{bmatrix} - \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \right\|^2$$

$$= \frac{1}{n} \left\| \underbrace{\begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}}_{\mathbb{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{\beta} - \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{Y} \right\|^2$$

$$= \frac{1}{n} \left\| \mathbb{X}\beta - Y \right\|^2$$

What about non-linearities?

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

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

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

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

as a matrix problem

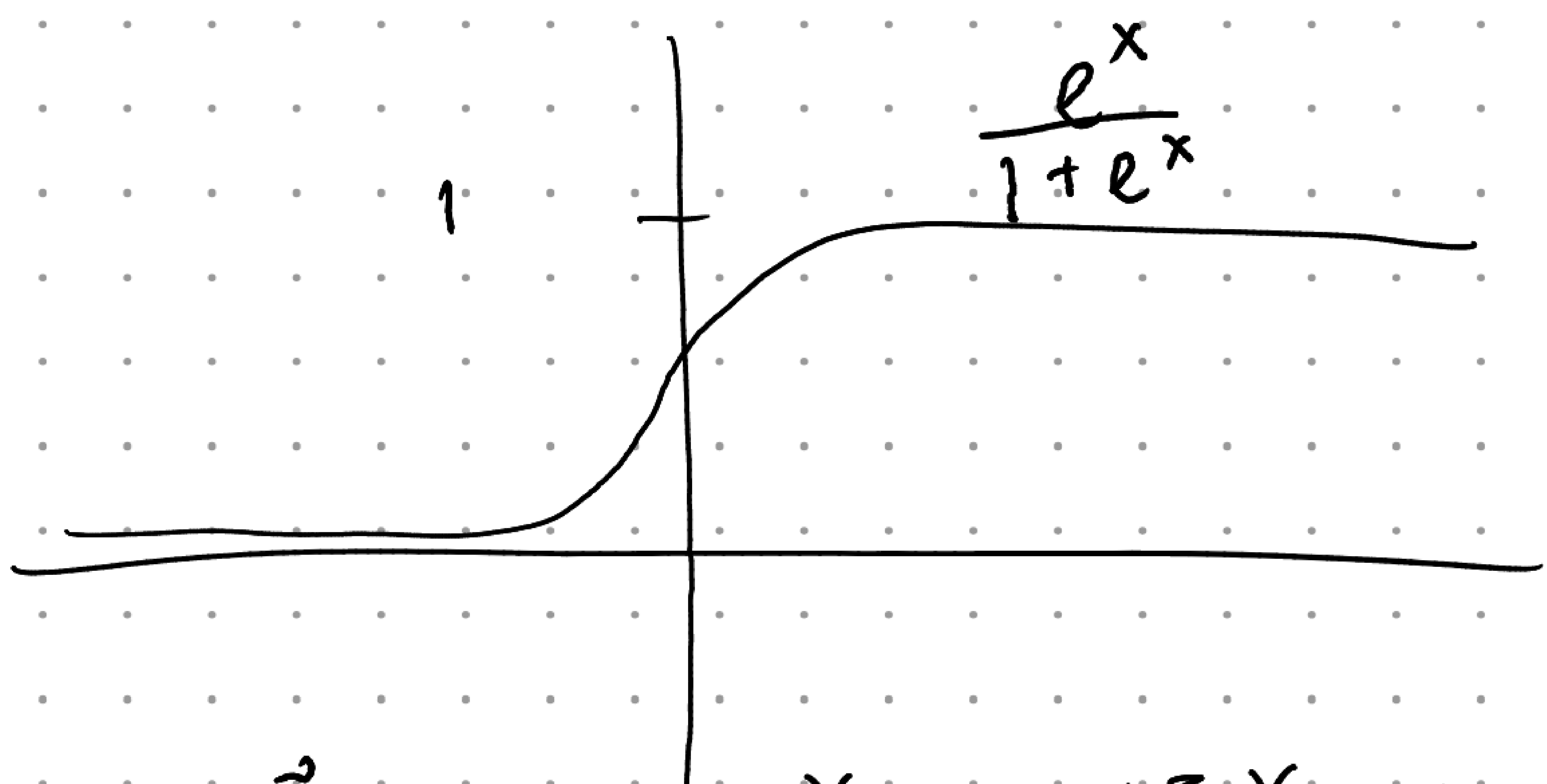
Logistic regression

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

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



Find $\vec{\beta}$ s.t. $\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$ big if $Y_i = 1$
small if $Y_i = 0$

Since $Y_i = 0$ or $Y_i = 1$ and

$$\mathbb{P}[Y_i = 1 | \vec{X}_i] = p_i,$$

$$Y_i | \vec{X}_i \sim \text{Ber}(p_i)$$

$$f_i(y) = p_i^y (1-p_i)^{1-y} = \begin{cases} p_i & y = 1 \\ 1-p_i & y = 0 \end{cases}$$

$$\mathcal{L}_n(\beta) = \prod_{i=1}^{n} f_i(Y_i)$$

$$= \prod_{i=1}^{n} p_i^{Y_i} (1-p_i)^{1-Y_i}$$

$$\ell_n(\beta) = \sum_{i=1}^{n} Y_i \ln(p_i) + (1-Y_i) \ln(1-p_i)$$

$$= \sum_{i: Y_i = 1} \ln(p_i) + \sum_{i: Y_i = 0} \ln(1-p_i)$$

Now, maximize $\ell_n(\beta)$ by numerical method