Logistic Regression

$y_1, \dots, y_n$ binary responses

$x_1, \dots, x_n$ continuous explanatory variable

$$Pr(Y = y_i \mid X = x_i) = \theta(x_i)$$

$$\text{logit}(\theta(x_i)) = \log\left(\frac{\theta(x_i)}{1-\theta(x_i)}\right) = \underbrace{\alpha + \beta x_i}$$

or

$$\underbrace{\frac{\theta(x_i)}{1-\theta(x_i)}}_{\text{odds}} = e^{\alpha + \beta x_i}$$

odds : exponential function of x_i

Interpretation :

$$\frac{\dfrac{\theta(x+1)}{1-\theta(x+1)}}{\dfrac{\theta(x)}{1-\theta(x)}} = \frac{e^{\alpha+\beta(x+1)}}{e^{\alpha+\beta x}} = e^{\beta}$$

we can have multiple explanatory variables :

$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \sum_{j=1}^{p} \beta_j x_{i,j}$$

- what if we have a binary explanatory variable ?

$$e^{\beta} = \frac{\dfrac{Pr(Y=1 \mid X=1)}{1 - Pr(Y=1 \mid X=1)}}{\dfrac{Pr(Y=1 \mid X=0)}{1 - Pr(Y=1 \mid X=0)}}$$

Once you fit the model, you can look at:

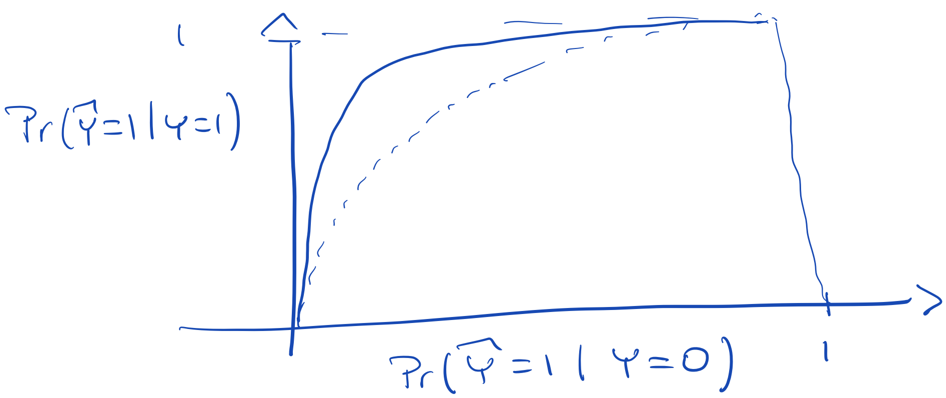
Sensitivity:

$$Pr(\hat{Y} = 1 \mid Y=1)$$

Specificity:

$$Pr(\hat{Y} = 0 \mid Y=0)$$

ROC : Receiving Operating Characteristic Curve



Estimation : Maximum Likelihood

$y_i \overset{ind}{\sim} \text{Bernoulli}(\theta_i)$

Likelihood:

$$L(\alpha, \beta) = \prod_{i=1}^{n} \theta_i^{y_i} (1-\theta_i)^{1-y_i}$$

$$\text{Logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \beta x_i$$

$$\frac{\theta_i}{1-\theta_i} = e^{\alpha+\beta x_i}$$

$$\theta_i = e^{\alpha+\beta x_i} - \theta_i e^{\alpha+\beta x_i}$$

$$\theta_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

$$\log L(\alpha,\beta) = \sum_{i=1}^{n} y_i \log(\theta_i) + \sum_{i=1}^{n} (1-y_i)\log(1-\theta_i) = \cdots$$

$$= \sum_{i=1}^{n} y_i(\alpha+\beta x_i) + \sum_{i=1}^{n} \frac{1}{e^{\alpha+\beta x_i}}$$

To maximize $\log L(\alpha,\beta)$:

$$\frac{\partial \log L(\alpha,\beta)}{\partial \alpha} = 0 \qquad \frac{\partial \log L(\alpha,\beta)}{\partial \beta} = 0$$

<u>Deviance</u>

$$D = 2\left[\log L(\hat{\theta}_{Max}; y) - \log L(\hat{\theta}; y)\right]$$

- Sampling distribution of D:
  - If the model is a "good fit" $D \approx \chi^2_{N-p}$

    or $D \cong N-p$

- For group data:

$$X^2 = \sum \frac{(\text{observed} - \text{fitted})^2}{\text{fitted}}$$

The deviance D can be used to compare nested models:

M_0 : smaller model with q parameters

M_1 : larger model with p parameters

$q < p$ <u>and</u> M_0 nested in M_1

D_0, D_1 : deviances

$$\Delta D = D_0 - D_1$$

$H_0 : \beta = \beta_0$ $\quad$ $H_1 : \beta = \beta_1$

Under H_0 $\quad \Delta D \simeq \chi^2_{p-q}$