Chapter 4 : Resampling

Newton's three sisters

March 2, 2025

Department of Statistics
Sungshin Women's University

Outline

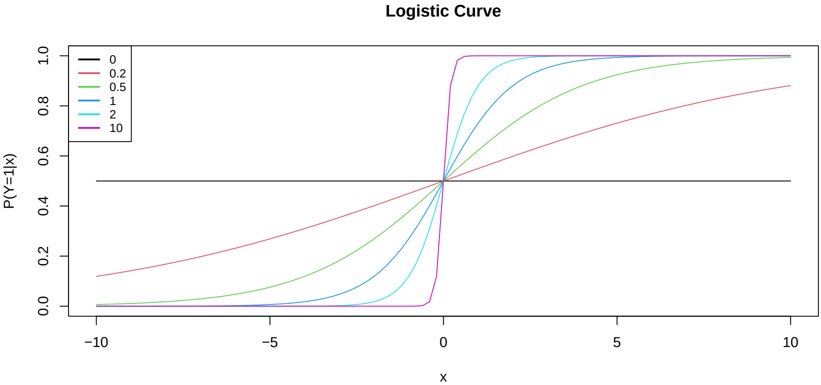
1 Logistic Regression

2 Netwon Raphson Method

3 Linear and Quadratic Discrimination

- The probabilities of $y = 1$ and $y = -1$ are expressed by $\frac{e^{\beta_0 + x\beta}}{1+e^{\beta_0 + x\beta}}$ and $\frac{1}{1+e^{\beta_0 + x\beta}}$

$$\frac{1}{1 + e^{-y(\beta_0 + x\beta)}}$$



increasing monotonically and convex and concave

$$f'(x) = \beta \frac{e^{-(\beta_0 + x\beta)}}{(1 + e^{-(\beta_0 + x\beta)})^2} \geq 0$$
$$f''(x) = -\beta^2 \frac{e^{-(\beta_0 + x\beta)}[1 - e^{-(\beta_0 + x\beta)}]}{(1 + e^{-(\beta_0 + x\beta)})^3}$$

- We see that $f(x)$ is increasing monotonically and is convex and concave when $x < -\beta_0/\beta$ and $x > -\beta_0/\beta$, they chage at $x = 0$, when $\beta = 0$

Outline

1 Logistic Regression

2 Netwon Raphson Method

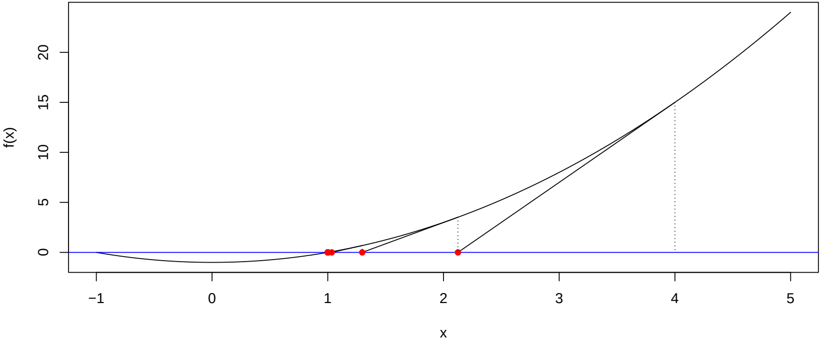
3 Linear and Quadratic Discrimination

Netwon Raphson

- Tangent line is $y - f(x_i) = f'(x_i)(x - x_i)$ the intersection with y = 0

$$x_{i+1} \triangleq x_i - \frac{f(x_i)}{f'(x_i)}$$

- Example $f(x) = x^2 - 1$ and $x_0 = 4$



Newton Raphson Method for two variables

- We can even be applied to two variables and two equations

$$\begin{cases} f(x,y) = 0, \\ g(x,y) = 0 \end{cases}$$

- We can see Newton Raphson is extended to

$$\begin{bmatrix} x \\ y \end{bmatrix} \leftarrow \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} & \frac{\partial f(x,y)}{\partial y} \\ \frac{\partial g(x,y)}{\partial x} & \frac{\partial g(x,y)}{\partial y} \end{bmatrix}^{-1} \begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}$$

- We apply the same method to the problem of finding $\beta_0 \in \mathbb{R}$ and $\beta \in \mathbb{R}^p$ such that $\bigtriangledown l(\beta_0, \beta) = 0$

$$(\beta_0, \beta) \leftarrow (\beta_0, \beta) - \{\bigtriangledown^2 l(\beta_o, \beta)\}^{-1} \bigtriangledown l(\beta_0, \beta)$$

- If we differentiate the negative lon-likelihood $l(\beta_0, \beta) \in \mathbb{R} \times \mathbb{R}^{p+1}$ and we let $v_i = e^{-y_i(\beta_0 + x_i \beta)}, i = 1, \cdots, N$
- The vector $\bigtriangledown(\beta_0, \beta) \in \mathbb{R}^{p+1}$ such that the jth element is $\frac{\partial l(\beta_0, \beta)}{\partial \beta_j}$ can be expressed by $\bigtriangledown l(\beta_0, \beta) = -X^T u$ with

$$u = \begin{bmatrix} \frac{y_1 v_1}{1+v_1} \\ \vdots \\ \frac{y_N v_N}{1+v_N} \end{bmatrix}$$

- Note $y_i = \pm 1$, i.e., $y^2 = 1$, the matrix $\bigtriangledown^2 l(\beta_0, \beta)$ such that the (i, k)th element is $\frac{\partial^2 l(\beta_0, \beta)}{\partial \beta_j \beta_k}, j, k = 0, 1, \cdots, p$ can be expressed by $\bigtriangledown^2 l(\beta_0, \beta) = X^T W X$

$$W = \begin{bmatrix} \frac{v_1}{(1+v_1)^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{v_N}{(1+v_N)^2} \end{bmatrix}$$

Outline

1 Logistic Regression

2 Netwon Raphson Method

3 Linear and Quadratic Discrimination

a

-a