3-24-15

Classical (non-Bayesian) inference for a simple logistic regression model.

Recall the logistic regression likelihood function:

$$L(\alpha,\beta) \propto e^{\sum_{i=1}^{n}[y_i(\alpha+\beta x_i) - \ln(1+e^{\alpha+\beta x_i})]}$$

Goal: Choose α and β that make our (x,y) data as likely as possible. That is, choose α and β that maximize $L(\alpha,\beta)$

Note: The values of α and β that maximize $L(\alpha,\beta)$ are the <u>same</u> values that maximize $\ln(L(\alpha,\beta))$.

<u>Aside</u> Consider the task of finding the value x^* that maximizes the function $f(x)$. We know x^* is the value that satisfies $\frac{d}{dx}f(x)\big|_{x=x^*}=0$ (Assuming a max, Not min)

Now, we wish to find x^* that maximizes $\ln(f(x))$. We know x^* is the value that satisfies $\frac{d}{dx}\ln(f(x))\big|_{x=x^*}=0$

However, $\frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)}$ $\quad x^*$ s.t. $f'(x)=0$ assume $f(x)\neq 0$

To find the value of α and β that maximize $\ln(L(\alpha,\beta))$, we will use multivariate Newton-Raphson.

General multivariate Newton Raphson method →

General Multivariate Newton-Raphson Method:

Let $\vec{\theta}$ be a vector of parameters, and let $l(\vec{\theta})$ denote the natural log of the likelihood function. Iterate through the $\vec{\theta}$ vectors using the following recursive relation:

$$\vec{\theta}_{(i+1)} = \vec{\theta}_{(i)} + \Bigg(\underbrace{-\frac{\partial^2 l}{\partial \vec{\theta}^2}}_{\text{matrix}}\Bigg|_{\vec{\theta}_{(i)}}\Bigg)^{-1} \cdot \Bigg(\underbrace{\frac{\partial l}{\partial \vec{\theta}}}_{\text{vector}}\Bigg|_{\vec{\theta}_{(i)}}\Bigg)$$

This will find the root of l'

If l were univariate, we would find the root of l' by iterating through:

$$\theta_{i+1} = \theta_i - \frac{l'(\theta_i)}{l''(\theta_i)}$$

$$\frac{\partial^2 l}{\partial \vec{\theta}^2}\Bigg|_{\vec{\theta}_{(i)}} = \begin{bmatrix} \frac{\partial^2 l}{\partial \vec{\theta}_{[1]}^2} & \frac{\partial^2 l}{\partial \vec{\theta}_{[1]}\vec{\theta}_{[2]}} & \cdots & \frac{\partial^2 l}{\partial \vec{\theta}_{[1]}\vec{\theta}_{[K]}} \\ \vdots & \ddots & \frac{\partial^2 l}{\partial \vec{\theta}_{[2]}^2} & \vdots \\ \frac{\partial^2 l}{\partial \vec{\theta}_{[K]}\vec{\theta}_{[1]}} & & \cdots & \frac{\partial^2 l}{\partial \vec{\theta}_{[K]}^2} \end{bmatrix}$$

$$\frac{\partial l}{\partial \vec{\theta}}\Bigg|_{\vec{\theta}_{(i)}} = \left\{\frac{\partial l}{\partial \theta_{[1]}}, \frac{\partial l}{\partial \theta_{[2]}}, \ldots, \frac{\partial l}{\partial \theta_{[K]}}\right\}\Bigg|_{\vec{\theta}_{(i)}}$$

For our example, $\vec{\theta} = \{\alpha, \beta\}$. Therefore,

$$\vec{\theta}_{(i+1)} = \vec{\theta}_{(i)} - \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} \end{bmatrix}_{\vec{\theta}_{(i)}}^{-1} \cdot \begin{bmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \end{bmatrix}_{\vec{\theta}_{(i)}}$$

∂ = partial derivative

Notice $\ln(L(\alpha,\beta)) = \Sigma\left[y_i(\alpha+\beta x_i) - \ln(1+e^{\alpha+\beta x_i})\right]$

$$\frac{\partial l}{\partial \alpha} = \Sigma\left[y_i - \frac{1}{1+e^{\alpha+\beta x_i}} \cdot e^{\alpha+\beta x_i}\right]$$

$$\frac{\partial l}{\partial \beta} = \Sigma\left[y_i x_i - \frac{1}{1+e^{\alpha+\beta x_i}} \cdot e^{\alpha+\beta x_i} \cdot x_i\right]$$

$$= \Sigma x_i\left[y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right]$$

derive $\frac{\partial l}{\partial \alpha}$ w/ respect to α

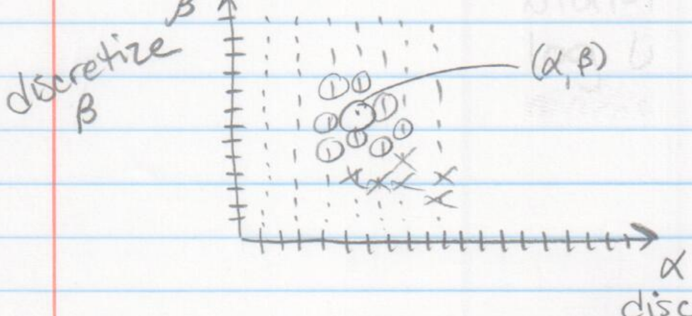
$$\frac{\partial^2 l}{\partial \alpha^2} = -\Sigma\left[\frac{\underbrace{(1+e^{\alpha+\beta x_i})}_{\text{bottom}}\underbrace{(e^{\alpha+\beta x_i})}_{\text{d of top}} - \underbrace{(e^{\alpha+\beta x_i})}_{\text{top}}\underbrace{(e^{\alpha+\beta x_i})}_{\text{d of bottom}}}{\underbrace{(1+e^{\alpha+\beta x_i})(1+e^{\alpha+\beta x_i})}_{(\text{bottom})^2}}\right]$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta} = -\Sigma\left[\frac{(1+e^{\alpha+\beta x_i})(e^{\alpha+\beta x_i})x_i - (e^{\alpha+\beta x_i})(e^{\alpha+\beta x_i})x_i}{(1+e^{\alpha+\beta x_i})(1+e^{\alpha+\beta x_i})}\right]$$

$$\frac{\partial^2 l}{\partial \beta^2} = -\Sigma\left[\frac{(1+e^{\alpha+\beta x_i})(e^{\alpha+\beta x_i})x_i^2 - (e^{\alpha+\beta x_i})(e^{\alpha+\beta x_i})x_i^2}{(1+e^{\alpha+\beta x_i})(1+e^{\alpha+\beta x_i})}\right.$$

HW10 (1) Use the Newton-Raphson method to maximize the log likelihood from last week's logistic regression model. Use the HW9 data.

(2) Find a radius of convergence for starting α and β values.



Let $p_i = \dfrac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$

$\dfrac{\partial l}{\partial \alpha} = \sum (y_i - p_i)$ $\quad\quad$ $\dfrac{\partial l}{\partial \beta} = \sum x_i (y_i - p_i)$

$\dfrac{\partial^2 l}{\partial \alpha^2} = -\sum p_i (1-p_i)$ $\quad\quad$ $\dfrac{\partial^2 l}{\partial \beta^2} = -\sum x_i^2 p_i (1-p_i)$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta} = -\sum x_i p_i (1-p_i) = \frac{\partial^2 l}{\partial \beta \partial \alpha}$$

R : invert matrix "Solve" $\quad$ $m = c\left[\frac{\partial^2 l}{\partial \alpha^2}, \frac{\partial^2 l}{\partial \alpha \partial \beta}, \frac{\partial^2 l}{\partial \beta \partial \alpha}, \frac{\partial^2 l}{\partial \beta^2}\right]$

solve(m)

M= matrix(c(a11, a12, a12, a22), ncol=2)

col1 row1 $\quad$ col2 row1 $\quad$ col1 row2 $\quad$ col2 row2

Solve(M) %*% V