Lecture Notes 8: Non-Linear Regression

In the previous lecture notes I had described about two different methods to compute the least squares or maximum likelihood estimators under the assumptions of Gaussian errors. The basic ideas of the two different methods are the following. In the first one, which is known as the Gauss-Newton method, where the Taylor series approximation is made on the non-linear function $f(\boldsymbol{\theta})$, and in the second case, which is known as the Newton's method, where the approximation is made on the residual sum of squares directly. In this note I would like to give some geometric interpretations of both the methods.

Gauss-Newton Method

Let us recall that we would like to find $\widehat{\boldsymbol{\theta}}$, the least squares estimator of $\boldsymbol{\theta}^*$, and $\boldsymbol{\theta}^{(0)}$ is an initial guess, i.e. it is expected to close to $\widehat{\boldsymbol{\theta}}$. In this method, we expand the function $f(\boldsymbol{\theta})$, around the point $\boldsymbol{\theta}^{(0)}$. I would like to explain this with some specific example. Suppose we have the following non-linear regression model

$$y_t = e^{-t\theta} + \epsilon_t; \qquad t = 1, \ldots, n.$$

In this case $f_t(\theta) = e^{-t\theta}$. Now to understand it better let us assume a simple case when $n = 2$, i.e. $t = 1, 2$, and let us assume we have the observation say $(y_1, y_2) = (0.5, 0.5)$. The function

$$f(\theta) = \begin{pmatrix} f_1(\theta) \\ f_2(\theta) \end{pmatrix} = \begin{pmatrix} e^{-\theta} \\ e^{-2\theta} \end{pmatrix}$$

is known as the expectation surface. We want to obtain the least squares estimate of θ, i.e. the value of θ say $\widehat{\theta}$ which minimizes

$$Q(\theta) = (0.5 - e^{-\theta})^2 + (0.5 - e^{-2\theta})^2.$$

In this case first we will show how we can obtain an initial guess of $\widehat{\theta}$. First we plot $Q(\theta)$ in Figure 1. It is clear from the residual sum of squares that $\widehat{\theta}$ is close to 0.5, hence we can take $\theta^{(0)} = 0.5$. Now let us loot at the expectation surface, it has been plotted in Figure 2. Here x axis denotes $f_1(\theta)$ and the y axis denotes $f_2(\theta)$. Now with the initial guess $\theta^{(0)} = 0.5$, it is point on the

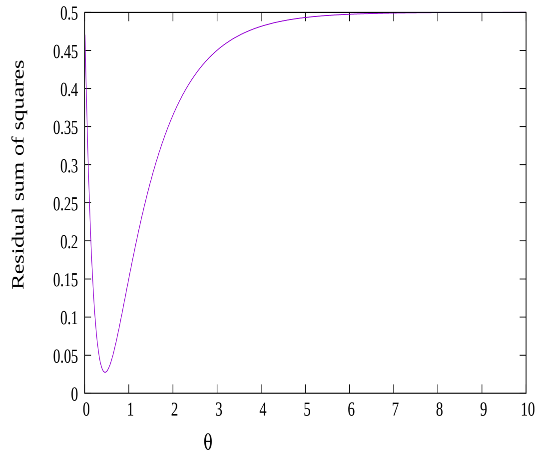

Figure 1: Residual sum of squares

expectation surface namely $(e^{-0.5}, e^{-1.0}) \approx$ (0.6065,0.3679). Let us expand the function $f(\theta)$ near the point $\theta^{(0)} = 0.5$ with the first order Taylor series expansion, and that will be

$$f(\theta) = \begin{pmatrix} f_1(\theta) \\ f_2(\theta) \end{pmatrix} = \begin{pmatrix} e^{-\theta} \\ e^{-2\theta} \end{pmatrix} \approx \begin{pmatrix} e^{-\theta^{(0)}} \\ e^{-2\theta^{(0)}} \end{pmatrix} + \begin{pmatrix} -e^{-\theta^{(0)}}(\theta - \theta^{(0)}) \\ -2e^{-2\theta^{(0)}}(\theta - \theta^{(0)}) \end{pmatrix}.$$

The non-linear expectation surface $f(\theta)$ is approximated by a linear plane namely

$$\begin{pmatrix} e^{-\theta^{(0)}} \\ e^{-2\theta^{(0)}} \end{pmatrix} + \begin{pmatrix} -e^{-\theta^{(0)}}(\theta - \theta^{(0)}) \\ -2e^{-2\theta^{(0)}}(\theta - \theta^{(0)}) \end{pmatrix}. \tag{1}$$

It is the tangent plane of the expectation surface at the point $(e^{-0.5}, e^{-1.0})$. The tangent plane (1) is a straight line passing through the point $(e^{-\theta^{(0)}}, e^{-2\theta^{(0)}})$, and it has the slope 2. Based on the above approximation, the residual sum of squares $Q(\theta)$ can be approximated as

$$\begin{aligned} Q(\theta) &= (0.5 - e^{-\theta})^2 + (0.5 - e^{-2\theta})^2 \\ &\approx (0.5 - e^{-\theta^{(0)}} + e^{-\theta^{(0)}}(\theta - 0.5))^2 + (0.5 - 2e^{-2\theta^{(0)}} + 2e^{-2\theta^{(0)}}(\theta - 0.5))^2 \\ &= (r_1 + e^{-\theta^{(0)}}(\theta - 0.5))^2 + (r_2 + 2e^{-2\theta^{(0)}}(\theta - 0.5))^2. \end{aligned} \tag{2}$$

Here $(r_1, r_2) = (0.5 - e^{-\theta^{(0)}}, 0.5 - 2e^{-2\theta^{(0)}}) \approx (0.5 - 0.6065, 0.5 - 0.3679)$. Therefore, $\theta^{(1)}$ which minimizes the last expression of (2) can be obtained by simple projection method, and that is

$$\theta^{(1)} \approx 0.5 + (e^{-2\theta^{(0)}} + 4e^{-4\theta^{(0)}})^{-1}(r_1 e^{-\theta^{(0)}} + r_2 e^{-2\theta^{(0)}}) = 0.5 - 0.0176 = 0.4824$$

The increment $\delta^1 = -0.0176$. In this case $Q(\theta^{(0)}) \approx 0.0288$ and $Q(\theta^{(1)}) \approx 0.0279$. Therefore, the residual sum of squares has decreased. Hence, $\theta^{(1)}$ is closer to $\widehat{\theta}$ than $\theta^{(0)}$.

Sometime it may happen that $Q(\theta^{(0)}) < Q(\theta^{(1)})$. In this case it may not be wise to move to $\theta^{(1)}$ from $\theta^{(0)}$. In this case we suggest to choose $0 < \lambda < 1$, such that $Q(\theta^{(0)}) > Q(\lambda\theta^{(1)})$, and in that case move to $\lambda\theta^{(1)}$ from $\theta^{(0)}$ rather that to $\theta^{(1)}$. It is known as the step factor modification.

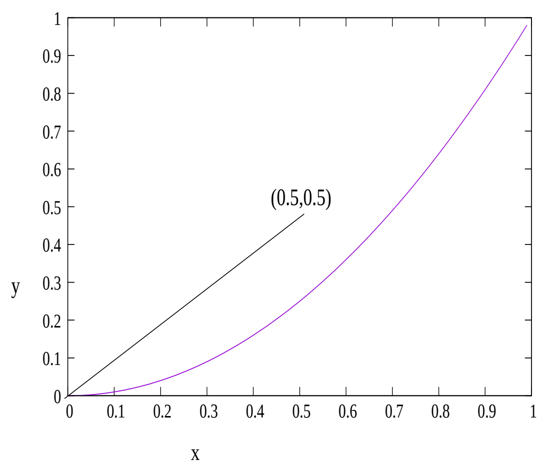

Figure 2: Expectation surface

Newton Method

In this method we expand the function $Q(\theta)$ by two terms Taylor series expansion directly near the point $\theta^{(0)}$, and that is

$$Q(\theta) = Q(\theta^{(0)}) + (\theta - \theta^{(0)})\frac{d}{d\theta}Q(\theta^{(0)}) + \frac{1}{2}(\theta - \theta^{(0)})^2\frac{d^2}{d\theta^2}Q(\theta^{(0)}),$$

here

$$\frac{d}{d\theta}Q(\theta) = 2(0.5 - e^{-\theta})e^{-\theta} + 2(0.5 - e^{-2\theta})e^{-2\theta} = e^{-\theta} - e^{-2\theta} - 2e^{-4\theta}$$

$$\frac{d^2}{d\theta^2}Q(\theta) = -e^{-\theta} + 2e^{-2\theta} + 8e^{-4\theta}.$$

Hence

$$\frac{d}{d\theta}Q(0.5) = e^{-0.5} - e^{-1.0} - 2e^{-2.0} = -0.0320$$

$$\frac{d^2}{d\theta^2}Q(0.5) = -e^{-0.5} + 2e^{-1.0} + 8e^{-2.0} = 1.2119.$$

Therefore,

$$Q(\theta) = 0.0288 - 0.0320(\theta - 0.5) + 0.6060(\theta - 0.5)^2 \tag{3}$$

The approximation surface of $Q(\theta)$ near $\theta^{(0)}$ has been presented in Figure 3). It is a quadratic surface (parabola in case of dimension one), and the minimum occurs at $\theta^{(1)}$, where $\theta^{(1)}$ can be obtained as the unique minimum of (3), and

$$\theta^{(1)} = 0.5 + \frac{0.0320}{2 \times 0.6060} = 0.5264.$$

Here also similarly, the step factor modification can be implemented.

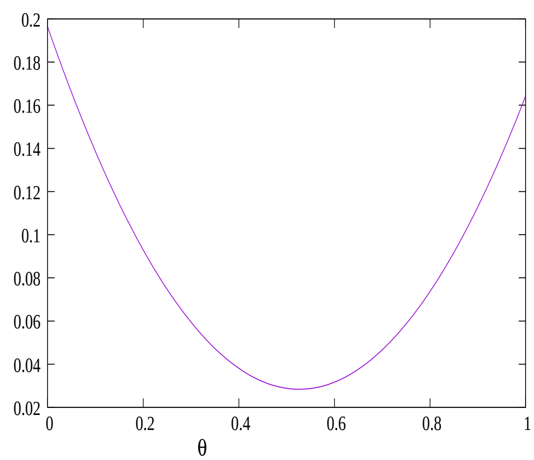

Figure 3: Approximation of $Q(\theta)$ surface