Lecture Notes 4: Non-Linear Regression

So far we have discussed about the simple linear regression model and also different method of estimations. We have further discussed about a simple numerical technique which can be used to find an approximate solution of a non-linear equation if the function is sufficiently smooth. Although, we have illustrated for a function of one variable, the method can be easily extended for more than one variable also. Now before moving to multiple linear regression model and non-linear regression model, we will be discussing another specific regression model which plays a very important role in different applications.

Let us consider the example which has been given in Lecture Note 1 on Rumford cooling experiment, where you observe temperature versus time. Suppose we denote the temperature at the time point t as y_t, then we can think of a more general model than a simple regression model $y_t = \beta_0 + \beta_1 t + \epsilon_t$, for example

$$y_t = \beta_0 + \beta_1 t + \ldots + \beta_p t^p + \epsilon_t; \quad t = t_1, \ldots, t_n. \tag{1}$$

Here, ϵ_t has the same assumptions as before, i.e. it is assumed that ϵ_t's are independent and identically distributed random variables with mean zero and finite variance. The above model (1) is known as the polynomial regression model. In practice it has lots of significance as the simple linear regression model can take only straight line, the polynomial regression model can take curve lines also. Now all the methods which we have proposed for the simple linear regression model, can be used directly. For illustrative purposes, we have taken $p = 2$, and $t_i = i$, but all the methods can be used for any general p and for general t_i's also. Therefore, we consider the following quadratic model:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t; \quad t = 1, 2, \ldots, n. \tag{2}$$

Therfore, the least squares estimators of β_0, β_1 and β_2 can be obtained as the argument minimum of $Q_{LSE}(\beta_0, \beta_1, \beta_2)$, where

$$Q_{LSE}(\beta_0, \beta_1, \beta_2) = \sum_{t=1}^{n} \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2 \right)^2. \tag{3}$$

Note that $Q_{LSE}(\beta_0, \beta_1, \beta_2)$ is a nice differentiable function and the least squares estimators of β_0, β_1 and β_2 can be obtained as the solutions of the following three linear equations:

$$\frac{\partial}{\partial \beta_0} Q_{LSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2 \right) = 0 \qquad (4)$$

$$\frac{\partial}{\partial \beta_1} Q_{LSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} t \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2 \right) = 0 \qquad (5)$$

$$\frac{\partial}{\partial \beta_2} Q_{LSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} t^2 \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2 \right) = 0. \qquad (6)$$

Let us use the following notations:

$$A_1 = \sum_{t=1}^{n} y_t, \qquad A_2 = \sum_{t=1}^{n} t y_t, \qquad A_3 = \sum_{t=1}^{n} t^2 y_t,$$

and

$$C_1 = \sum_{t=1}^{n} t, \quad C_2 = \sum_{t=1}^{n} t^2, \quad C_3 = \sum_{t=1}^{n} t^3, C_4 = \sum_{t=1}^{n} t^4.$$

Then (13), (14) and (6) can be written as

$$A_1 - n\beta_0 - C_1\beta_1 - C_2\beta_2 = 0 \qquad (7)$$
$$A_2 - C_1\beta_0 - C_2\beta_1 - C_3\beta_2 = 0 \qquad (8)$$
$$A_3 - C_2\beta_0 - C_3\beta_1 - C_4\beta_2 = 0. \qquad (9)$$

The equations (7), (8) and (9) can be expressed in a matrix form as

$$\begin{bmatrix} n & C_1 & C_2 \\ C_1 & C_2 & C_3 \\ C_2 & C_3 & C_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}. \qquad (10)$$

Hence, the solutions of (7), (8) and (9) can be obtained as

$$\begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \widehat{\beta_2} \end{bmatrix} = \begin{bmatrix} n & C_1 & C_2 \\ C_1 & C_2 & C_3 \\ C_2 & C_3 & C_4 \end{bmatrix}^{-1} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}, \qquad (11)$$

provided the above matrix is invertable. In this case becasue of the specific structure of the matrix, it can be shown that the above matrix is invertable. From (11), it can be seen that the least squares estimators of β_0, β_1 and β_2 can be obtained in explicit forms when the exist.

It has been shown before that although the least squares estimators can be obtained in explicit forms, and they have some nice desirable properties, they are not robust. In presence of few outliers the performace of the least squares estimators affect significantly. All the robust estimators like least absolute deviation estimators or Huber-M estimators can be used even in this case also to produce more robust estimators. In this case we are going to introduce another robust estimators which can be obtained quite easily similar to the least squares estimators.

Suppose $w(t)$ is a positive valued continuous function defined on $[0, 1]$, such that $w(t) \geq \gamma > 0$, for all $t \in [0, 1]$. Let us define

$$Q_{WLSE}(\beta_0, \beta_1, \beta_2) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2\right)^2. \qquad (12)$$

It is assumed that $w(t)$ is known is advance. We obtain the estimators of β_0, β_1 and β_2 by the argument minimum of $Q_{WLSE}(\beta_0, \beta_1, \beta_2)$. We call these estimators as the weighted least squares estimators of β_0, β_1 and β_2. Now the motivation of using the weighted least squares estimators as more robust estimators than the least squares estimators is the following. Suppose the outliers are in the middle portion of the time scale, i.e. near the time point $\frac{t}{2}$. In that case if $w(t) = 1 + (t - 0.5)^2$, then $Q_{WLSE}(\beta_0, \beta_1, \beta_2)$ puts less weight in the middle portion of the data than towards the end point. Similarly, if it is known that the outliers are at the beginning of the data sequence, then one can choose a weight function which is an increasing function, hence it puts less weight at the beginning and more weight afterwards.

From (12) we obtain three normal equations as follows:

$$\frac{\partial}{\partial \beta_0} Q_{WLSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2\right) = 0 \quad (13)$$

$$\frac{\partial}{\partial \beta_1} Q_{LSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} t w\left(\frac{t}{n}\right) \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2\right) = 0 \quad (14)$$

$$\frac{\partial}{\partial \beta_2} Q_{LSE}(\beta_0, \beta_1, \beta_2) = -2 \sum_{t=1}^{n} t^2 w\left(\frac{t}{n}\right) \left(y_t - \beta_0 - \beta_1 t - \beta_2 t^2\right) = 0. \quad (15)$$

If we use the notation

$$\widetilde{A}_1 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) y_t, \quad \widetilde{A}_2 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t y_t, \quad \widetilde{A}_3 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t^2 y_t,$$

and

$$\widetilde{C}_1 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t, \quad \widetilde{C}_2 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t^2, \quad \widetilde{C}_3 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t^3, \widetilde{C}_4 = \sum_{t=1}^{n} w\left(\frac{t}{n}\right) t^4.$$

Then (13), (14) and (6) can be written as

$$\widetilde{A}_1 - n\beta_0 - \widetilde{C}_1 \beta_1 - \widetilde{C}_2 \beta_2 = 0 \quad (16)$$
$$\widetilde{A}_2 - \widetilde{C}_1 \beta_0 - \widetilde{C}_2 \beta_1 - \widetilde{C}_3 \beta_2 = 0 \quad (17)$$
$$\widetilde{A}_3 - \widetilde{C}_2 \beta_0 - \widetilde{C}_3 \beta_1 - \widetilde{C}_4 \beta_2 = 0. \quad (18)$$

The equations (16), (17) and (18) can be expressed in a matrix form as

$$\begin{bmatrix} n & \widetilde{C}_1 & \widetilde{C}_2 \\ \widetilde{C}_1 & \widetilde{C}_2 & \widetilde{C}_3 \\ \widetilde{C}_2 & \widetilde{C}_3 & \widetilde{C}_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \widetilde{A}_1 \\ \widetilde{A}_2 \\ \widetilde{A}_3 \end{bmatrix}. \quad (19)$$

Hence, the solutions of (16), (17) and (18) can be obtained as

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} n & \widetilde{C}_1 & \widetilde{C}_2 \\ \widetilde{C}_1 & \widetilde{C}_2 & \widetilde{C}_3 \\ \widetilde{C}_2 & \widetilde{C}_3 & \widetilde{C}_4 \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{A}_1 \\ \widetilde{A}_2 \\ \widetilde{A}_3 \end{bmatrix}, \quad (20)$$

provided the above matrix is invertable.

There are two important questions which need to be answered. First one is how to choose the weight function if it is not known from where the outliers can be present, and the second question is how to choose p in the general

polynomial regression model (1). The second question we will address it later when we will discuss the general multiple regression model. For the first question the following method can be used.

Weight Function Bank

We choose different weight functions say $w_1(t), \ldots, w_K(t)$, where each of the weight function satisfies the properties as defined before. We make it normalized, i.e. $\int_0^1 w_j(t)dt = 1$, for $j = 1, \ldots, K$. Now find the WLSEs for each of the weight functions, and choose that particular weight function for which it provides the minimum weighted residual sums of squares, i.e. the minimum $Q_{WLSE}(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$. It is recommended that one of the weight function can be taken as $w(t) = 1$, hence the LSEs also become one of the members.