Lecture Notes 5: Non-Linear Regression

In this lecture notes, first we will discuss about different non-linear regression model. A non-linear regression model has the following form:

$$y = f(x, \theta) + \epsilon, \tag{1}$$

here $x = (x_1, \ldots, x_k)$ is a vector valued independent variables, they are observed, and $\theta = (\theta_1, \ldots, \theta_p)$ is a vector valued unknown parameter, and ϵ is a error random variable with mean zero and finite variance. It is called a non-linear regression model, when the function $f(x, \theta)$ is a non-linear function of the parameter vector θ. We do not treat it as a non-linear regression model if it is a non-linear function of x only. For example if

$$f(x; \theta) = \theta_1 e^{x\theta_2},$$

is a non-linear regression model, where as

$$f(x, \theta) = \theta_1 \sin(x_1) + \theta_2 \sin(x_2)$$

is not treated as a non-linear regression model.

Non-linear models are used either when they are suggested by theoretical consideration or to build known non-linear behavior into a model. For example

Example 1: Theoretical chemistry predicts that for a given sample of gas kept at a constant temperature, the volume and pressure p of gas satisfy the following relation: $pv^\gamma = c$, or $p = cv^{-\gamma}$. Here γ is a constant and it depends on the gas, and it needs to be estimated from the data. For example, a sample of a given gas is kept at a certain temperature, and if we observe the pressure of the gas at different volumes, say $(v_1, \ldots, v_n)$ and we observe the pressures as $(p_1, \ldots, p_n)$. Then based on the theory it is expected we can have the following assumptions:

$$p_i = cv_i^{-\gamma} + \epsilon_i, \tag{2}$$

because it is expected that one cannot measure the pressure accurately for a given volume, and there will be some error, and that is ϵ_i. Now based on the data namely $(p_1, v_1), \ldots, (p_n, v_n)$, for a given data, one needs to estimate c and γ. It may be observed that the parameter c in the model is a linear parameter, whereas γ is a non-linear parameter. It is clear that the model (2) is a non-linear regression model.

Now on the other hand from the relation $p = cv^{-\gamma}$, one can obtain, by taking logarithm on both sides;

$$\ln p = \ln c - \gamma \ln v.$$

Now if we write $y_i = \ln p_i$, and $\delta = \ln c$, then one can think of the following model

$$y_i = \delta - \gamma \ln v_i + \epsilon_i. \tag{3}$$

It is interesting to observe that the model (3) is a linear regression model. Therefore, it is important to observe that in this case it depends on the model assumption whether we can assume it to be a linear model or a non-linear model. If it is assumed that the error is multiplicative then we can transform the model as a linear regression model, otherwise, if the error is assumed to be additive, then it will be a non-linear regression model.

Example 2: The relationship between number of migrants people and the size of the two urban areas the following 'gravity'-type non-linear model has been proposed in the literature:

$$I_{AB} = \frac{\alpha P_A P_B}{D^\beta}. \tag{4}$$

Here I_{AB} denotes the amount of migration from A to B, P_A, P_B denote the population size of A and B, respectively, D denotes the distance between A and B. Here α and β are two unknown parameters. The problem is to estimate α and β from the amount of migration, population size and the associated distance data. Therefore, if $y_1, \ldots, y_n$ denote the number of migrants people from $A_1, \ldots, A_n$ to B, and $P_1, \ldots, P_n$ denote the population size at $A_1, \ldots, A_n$, respectively, further P_B denote the population size at B, and if $d_1, \ldots, d_n$ denote the distance between $A_1, \ldots, A_n$ to B, then the non-linear regression model can be written as

$$y_i = \frac{\alpha P_i P_B}{d_i^\beta} + \epsilon_i; \quad i = 1, \ldots, n.$$

Hence, based on the observed data, one is interested to estimate α and β. It may be noted this may be used quite effectively in predicting the number of immigrants in the future.

Example 3:

The following Figure 1 provides the ECG plot of a normal male adult. It is clearly a non-linear function, and it repeats (periodic) the pattern after certain times. It has been observed that this kind of data (signal) can be explained using a model of the following form:

$$y_t = \sum_{k=1}^{p} \{A_k \cos(\omega_k t) + B_k \sin(\omega_k t)\} + \epsilon_t. \tag{5}$$

Here $A_k^2 + B_k^2$ is known as amplitude of the signal and ω_k is known as frequencies. Further p is known as the number of components. One is interested in estimating p, $A_1, \ldots, A_p$, $B_1, \ldots, B_p$ and $\omega_1, \ldots, \omega_p$.

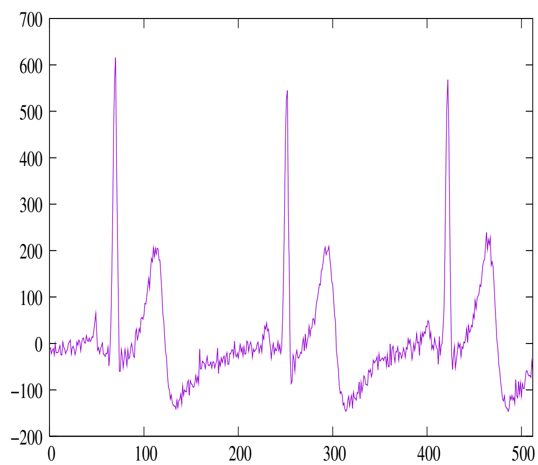

Figure 1: ECG plot of a normal male adult

The estimation of the unknown parameters of an ECG signal can be used for compression purposes. For example, it is not needed to store the entire data set, instead we can store only few parameters, from which the whole signal can be easily generated. It can be used for future prediction, and hence if there is any abnormalities, it may be detected.

Before going into the details about non-linear regression model, first we would like to give a brief review about the linear regression model. As it has been mentioned that the linear regression model is a special case of the non-linear regression model. Therefore, it is expected that whatever theories we develop for the non-linear regression model, it should hold for the linear regression model also.

Let us write a linear regression model in the following matrix form. It will help us in many respects. A linear regression model can be written in the following form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{6}$$

Here $\mathbf{Y}$ is a $n \times 1$ observation vector, $\mathbf{X}$ is a $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ is a $n \times 1$ random vector. It is assumed for the time that

$$E(\epsilon_i) = 0 \quad \text{and} \quad \text{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j. \end{cases}$$

The main issue in a linear regression model, is to obtain the estimate of $\boldsymbol{\beta}$ and σ^2, from the given observations namely $(\mathbf{Y}, \mathbf{X})$, and develop their properties. I believe all of you have gone through a course in linear regression model, hence I am not going to describe many things in details. The main points I am going to mention which most likely you may not have seen before.

First of all the least squares estimator of $\boldsymbol{\beta}$ can be obtained by minimizing the residual sums of square, and that is

$$Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Now $\widehat{\boldsymbol{\beta}}$, the LSE of $\boldsymbol{\beta}$ can be obtained by minimizing $Q(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. It can be done in different ways. We can differentiate $Q(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and equate it to 0, otherwise, we can use the projection method also, to obtain the LSE of $\boldsymbol{\beta}$. It becomes

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

There are several issues associated with it. We will discuss it in the lecture notes.