Lecture Notes 2: Non-Linear Regression

In the last lecture I had given a brief description of the course and about the grading scheme. I had also mentioned about the linear regression model. It should be mentioned that the non-linear regression model is a generalization of the linear regression model. Hence, whatever methods we develop for the non-linear regression model it should work for the linear regression model also as a special case.

As I had mentioned before that at the beginning in the couple of lectures we would like discuss different aspects of the linear regression models. Many of you had already taken a course on the linear regression model, but I would like to discuss some of the issues which most probably you have not been exposed to. Let us remember that the simple linear regression model can be described as follows: Suppose I have a set of n data points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. I want to a fit a *best* straight line through these points. Mathematically speaking suppose it is assumed that $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ have the following relation

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i; \qquad i = 1, \ldots, n, \tag{1}$$

where ϵ_i's are independent identically distributed random variables with mean zero and finite variance, then what are the *best* values (estimators) of β_0 and β_1. It may be mentioned that the mathematical model (1) may be one of the most basic model which has been used to understand the relation between the two variables. There are several reasons to use this model to understand the relation between the two variables. First of all it is very simple and easy to understand and even if it may not be exactly true it may be used as a first order approximation.

Most probably you have seen this model even in your high school. The most common estimators of β_0 and β_1 can be obtained as the argument minimum of $Q(\beta_0, \beta_1)$, where

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{2}$$

Note that the minimization of $Q(\beta_0, \beta_1)$ can be obtained in a very simple form. Since $Q(\beta_0, \beta_1)$ is a nice differentiable function in both the arguments we differentiate it with respect to β_0 and β_1 and put them equal to zero and solve them. We obtain

$$\begin{aligned}
\frac{\partial}{\partial \beta_0} Q(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \quad \text{and} && (3) \\
\frac{\partial}{\partial \beta_1} Q(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} x_i \, (y_i - \beta_0 - \beta_1 x_i) = 0 && (4)
\end{aligned}$$

From (3) and (4) we obtain

$$\begin{aligned}
\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i &= 0 \quad \text{and} && (5) \\
\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i y_i &= 0. && (6)
\end{aligned}$$

If we denote

$$A = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad B = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad C = \frac{1}{n} \sum_{i=1}^{n} x_i y_i,$$

then (5) and (6) can be written as

$$\begin{bmatrix} 1 & A \\ A & C \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} B \\ C \end{bmatrix}. \tag{7}$$

Then if $C \neq A^2$, then

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 1 & A \\ A & C \end{bmatrix}^{-1} \begin{bmatrix} B \\ C \end{bmatrix}. \tag{8}$$

Hence,

$$\widehat{\beta}_0 = \frac{BC - AC}{C - A^2} \quad \text{and} \quad \widehat{\beta}_1 = \frac{C - AB}{C - A^2}. \tag{9}$$

Therefore, we have observed that in the simple linear regression model if $C \neq A^2$, then the least squares estimators of β_0 and β_1 can be obtained in explicit forms. This is one of the reasons that the least squares estimators are the most popular estimators of the unknown parameters.

Answer the following questions:

Question 1: What will happen if $C = A^2$?

Question 2: Find the least squares estimator of β_0, if it is known that $\beta_1 = 0$.

Question 3: Find the least squares estimator of β_0, if it is known that $\beta_1 = 1$.

Question 4: Find the least squares estimator of β_0, if it is known that $\beta_1 = \beta_0$.

Question 5: Suppose we denote

$$A_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad B_n = \frac{1}{n}\sum_{i=1}^{n} y_i \quad C_n = \frac{1}{n}\sum_{i=1}^{n} x_i y_i,$$

$$\widehat{\beta}_{0n} = \frac{B_n C_n - A_n C_n}{C_n - A_n^2} \quad \text{and} \quad \widehat{\beta}_{1n} = \frac{C_n - A_n B_n}{C_n - A_n^2}.$$

Express A_{n+1}, B_{n+1} and C_{n+1} in terms of A_n, B_n and C_n, respectively. Hence express $\widehat{\beta}_{0(n+1)}$ and $\widehat{\beta}_{1(n+1)}$ in terms of $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$, respectively.

It is important to discuss the properties of the least squares estimators of β_0 and β_1. Note that it is assumed that A_n's are real numbers but B_n's and C_n's are random variables. Hence, $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$ are also random variables. So the properties of $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$ mean how these random variables behave? For example $E(\widehat{\beta}_{0n})$, $E(\widehat{\beta}_{1n})$ or $V(\widehat{\beta}_{0n})$ and $V(\widehat{\beta}_{1n})$? What will happen to $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$ and $n \to \infty$. Is it possible to construct confidence intervals of β_0 and β_1 based on $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$ etc. We will discuss about these properties later.

Although the least squares estimators are very easy to calculate and we will see later they have several desirable properties, but one major problem about the least squares estimators is that they are not robust. They are very sensitive if there are some 'odd' observations, which are known as 'outliers' in the data set. In presence of very few outliers can affect the performance of the leas squares significantly. Due to this reason there are several robust estimators have been proposed in the literature. The most famous and very intuitive robust estimators are the least absolute deviation estimators, and they can be defined as follows. The least absolute deviation estimators of β_0 and β_1 can be obtained as the argument minimum of $Q_{LAD}(\beta_0, \beta_1)$, where

$$Q_{LAD}(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \,. \tag{10}$$

From the expression (10) it is clear that the above criterion puts less weight to an 'odd' observation compared to a least squares criterion. Due to this reason, the estimators obtained based on the criterion (10) is more robust compared to the least squares criterion. It can be shown empirically as well as theoretically also that the least absolute deviation estimators are more robust than the least squares estimators.

One major problem for the least absolute deviation estimators is the computational issue. Unlike the least squares estimators, the least absolute deviation estimators cannot be obtained in closed forms. The least absolute deviation estimators have to be obtained numerically unless for some specific cases. One simple method can be by grid search method in two dimensions. It can be implemented as follows. Assume that the true values of β_0 and β_1 lie within the range $[-M, M] \times [-M, M]$. Take grid points in $[-M, M] \times [-M, M]$ and compute the functional values at each grid point and choose that point for which it has the minimum value. The main advantage of this method is that if we choose M properly the method is going to work and depending our need we can choose the grid size, but the main the disadvantage of this method is that the computational time might be quite large. Answer the following questions.

Question 6: Find the least absolute deviation estimator of β_0, if $\beta_1 = 0$.

Question 7: Find the least absolute deviation estimator of β_0, if $\beta_1 = 1$.

Question 8: Find the least absolute deviation estimator of β_1, if $\beta_0 = 0$.

Question 9: Find the least absolute deviation estimator of β_1, if $\beta_0 = 2$.