Lecture Notes 3: Non-Linear Regression

In the last lecture I had discussed about the least absolute deviation (LAD) estimators, and that can be obtained by the argument minimum of $Q_{LAD}(\beta_0, \beta_1)$, where

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

We have already seen that the minimization of $Q_{LAD}(\beta_0, \beta_1)$ cannot be obtained in explicit forms unlike least squares estimators. In this case the the LAD estimators has to be obtained numerically. Several methods may be used to minimize $Q_{LAD}(\beta_0, \beta_1)$, but since it is not a differentiable function one has to use a technique which does not require derivative information. For example the brute force grid search method may be used, or random search method as we have discussed also can be used. Some of the other methods like genetic algorithm (GE) or Nelder-Mead method may be used for this purpose. But if β_0 or β_1 is known, then the minimization can be done without using any numerical method.

Question 1: If $\{z_1, \ldots, z_n\}$ are n real numbers prove that $a = \text{median}\{z_1, \ldots, z_n\}$ minimizes $\sum_{i=1} |z_i - a|$.

Question 2: If Z is a discrete random variable with the probability mass function $P(Z = a_i) = p_i$, for $i = 1, 2, \ldots$, where $a_1, a_2, \ldots$ are real numbers and $p_i > 0$ with $\sum_{i=1}^{\infty} p_i = 1$. Prove that $E|Z - a|$ is minimum for $a = \text{median}(Z)$.

Other than the LAD estimators there are some other robust estimators which are available in the literature. Another important one is known as the Huber-M estimators. It can be written in a very general form. Several well known estimators can be obtained as special cases. The Huber-M estimators of β_0 and β_1 can be obtained as the argument minimum of $Q_H(\beta_0, \beta_1)$, where $Q_H(\beta_0, \beta_1) = \sum_{i=1}^{n} \rho(y_i - \beta_0 - \beta_1 x_i)$, here the function ρ can be of different forms. For example, if $\rho(x) = |x|^2$, then Huber-M estimators become the least squares estimators, if $\rho(x) = |x|$, then Huber-M estimators become the least absolute deviation estimators. We can use the following bounded function $\rho(x)$ as follows:

$$\rho(x) = \begin{cases} x^2 & \text{if} \quad x^2 \le c \\ c & \text{if} \quad x^2 > c. \end{cases} \tag{1}$$

Here, c is some known constant. Since $\rho(x)$ is bounded, it can also produce robust estimators of the unknown parameters. Unfortunately, in this case also the estimators cannot be obtained in closed explicit forms. Once has to use some numerical algorithm to solve this problem. Further, since $\rho(x)$ is not a differentiable function one has to use some numerical algorithm which does not require derivative information.

Problem 3: Prove that if $\rho(x)$ satisfies (1), and $\beta_1 = 1$, then Huber-M estimator of β_0 can be obtained in explicit form.

Problem 4: Prove that if $\rho(x)$ satisfies (1), then Huber-M estimators can be obtained by solving one dimensional optimization problem.

Now we will be discussing another method of estimation of the unknown parameters β_0 and β_1 in a simple linear regression problem. In this case we minimize the sum of squares of perpendicular distances from the point (x_i, y_i) to the straight line $y = \beta_0 + \beta_1 x$. Let us recall that the perpendicular distance d_i from the point (x_i, y_i) to the straight line $y = \beta_0 + \beta_1 x$ is

$$d_i = \frac{|y_i - \beta_0 - \beta_1 x_i|}{\sqrt{1 + \beta_1^2}}.$$

Hence, the estimators of β_0 and β_1 can be obtained by the argument minimum of $Q_{TLS}(\beta_0, \beta_1)$, where

$$Q_{TLS}(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{1 + \beta_1^2} \right) = \frac{1}{1 + \beta_1^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{2}$$

The estimators which are obtained by minimizing (2) is known as the total least squares (TLS) estimators. Clearly, they cannot be obtained in explicit forms. They have to be obtained by using some optimization routine. But in this case the objective function is differentiable. Hence, some of the standard method like Newton-Raphson method may be used to obtain the estimators.

Problem 5: Prove that if $\beta_1 = 2$, the TLS estimator of β_0 can be obtained in explicit form.

Problem 6: Prove that the TLS estimators of β_0 and β_1 can be obtained by solving one dimensional optimization problem.

Note that a robust version of the TLS estimators also can be obtained by minimizing $Q_{RTLS}(\beta_0, \beta_1)$, where

$$Q_{RTLS}(\beta_0, \beta_1) = \sum_{i=1}^{n} \left| \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{1 + \beta_1^2} \right| = \frac{1}{|1 + \beta_1^2|} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|. \quad (3)$$

Problem 7: Prove that if $\beta_1 = 1$, the RTLS estimator of β_0 can be obtained in explicit form.

Problem 8: Prove that the RTLS estimators of β_0 and β_1 can be obtained by solving one dimensional optimization problem.

Now I would discuss some numerical methods by which TLS estimators can be obtained. Since (2) is a differentiable function, and the function has a minimum, hence the minimum can be obtained by solving two partial derivatives to be equal to zero. I will be discussing first a very general numerical method which can be used to solve a non-linear problem of the form:

$$\frac{d}{dx} f(x) = 0, \quad (4)$$

where $f : \mathbb{R} \to \mathbb{R}$, and it is a sufficiently differentiable function, at least twice. Also, it is assumed that there exists a $\widehat{x}$, such that $f'(\widehat{x}) = 0$. Let us consider the following function $f(x) = 2e^x + \frac{3}{2}e^{-2x}$ and it has been plotted in Figure 1.

The above function $f(x)$ is a sufficiently smooth function and it has a unique minimum in the range $[-1, 2]$. We want to find the point $\widehat{x}$, such that $f(\widehat{x}) \leq f(x)$, for all $x \in [-1, 2]$, equivalently we want to find the point $\widehat{x}$,

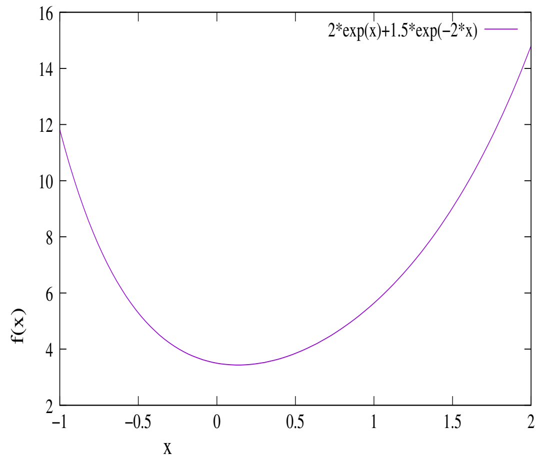

Figure 1: The function $f(x) = 2e^x + \frac{3}{2}e^{-2x}$

such that $f'(\widehat{x}) = 0$. In this case

$$\frac{d}{dx}f(x) = f'(x) = 2e^x - 3e^{-2x} = 0, \tag{5}$$

is a non-linear equation and the function $f'(x) = g(x)$ has been plotted in Figure 2. Clearly, the solution cannot be obtained explicitly.

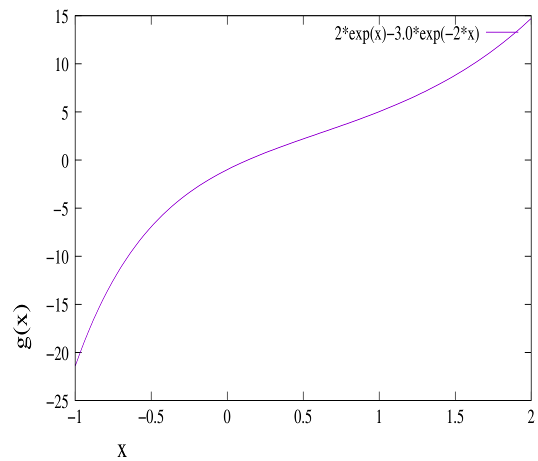

Figure 2: The function $f'(x) = g(x) = 2e^x - 3e^{-2x}$

We propose a numerical method to find an approximate solution of the non-linear equation (5). Suppose $x^{(0)}$ is an approximate value of $\widehat{x}$, a solution of (5), then we make Taylor series approximation of $f'(\widehat{x})$ around the point $x^{(0)}$ as follows:

$$f'(\widehat{x}) - f'(x^{(0)}) \approx (\widehat{x} - x^{(0)}) f''(x^{(0)}). \tag{6}$$

Since $f'(\widehat{x}) = 0$, therefore, (6) can be written as

$$\widehat{x} \approx x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})}, \tag{7}$$

provided $f''(x^{(0)}) \neq 0$. Hence, we can take

$$x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})}.$$

Hence, from the i-th step to the $(i+1)$-th step the operation becomes

$$x^{(i+1)} = x^{(i)} - \frac{f'(x^{(i)})}{f''(x^{(i)})}. \tag{8}$$

We stop the iteration when ever $|x^{(i+1)} - x^{(i)}| \leq \epsilon$, where ϵ is some pre-assigned chosen precision depending on the user.

Example: In the above example we have $f''(x) = 2e^{x} + 6e^{-2x}$ and $x^{(0)} = 0$. We can start with $x^{(0)} = 0$ and we can take $\epsilon = 0.001$, then

$$x^{(1)} = 0.125, \quad x^{(2)} = 0.13510, \quad x^{(3)} = 0.13516$$

Since $|x^{(3)} - x^{(2)}| < \epsilon$, we stop the iteration. Hence, the approximate value of $\widehat{x}$ is 0.13516. It may be noted that $f'(x^{(3)}) = 0.00003$.