Non Linear Regression

Non linear regression can be formulated as :

$$\min_X E(X) : X \in \mathcal{R}^M$$

where $E(X) \;= \frac{1}{2}\sum_{n=1}^{N}(f(X,A_n)-b_n)^2 \;= \frac{1}{2}R^T R$

$$R = \begin{bmatrix} f(X,A_1)-b_1 \\ f(X,A_2)-b_2 \\ \dots \\ f(X,A_N)-b_N \end{bmatrix}$$

f is the non linear regression model $\mathfrak{R}^K \rightarrow \mathfrak{R}^1$, which maps row vector A_n (with size K) into scalar b_n, for a total of N observations. Non linear regression is usually an optimization that involves Jacobian of E and Hession of E *w.r.t.* parameter X. We will find out how the Jacobian and Hessian of E related to the Jacobian and Hessian of f. Jacobian and Hessian of E are defined as :

$$\nabla_X E = [\partial_{x_1}E \quad \partial_{x_2}E \quad \dots \quad \partial_{x_M}E] \qquad \textit{Jacobian}$$

$$\nabla_X^2 E = \begin{bmatrix} \partial^2_{x_1x_1}E & \partial^2_{x_1x_2}E & \dots & \partial^2_{x_1x_M}E \\ \partial^2_{x_2x_1}E & \partial^2_{x_2x_2}E & \dots & \partial^2_{x_2x_M}E \\ \dots & \dots & \dots & \dots \\ \partial^2_{x_Mx_1}E & \partial^2_{x_Mx_2}E & \dots & \partial^2_{x_Mx_M}E \end{bmatrix} \qquad \textit{Hessian}$$

Lets expand Jacobian :

$$\nabla_X^T E = \begin{bmatrix} \partial_{x_1}E \\ \partial_{x_2}E \\ \dots \\ \partial_{x_M}E \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{n=1}^{N}(f(X,A_n)-b_n)\partial_{x_1}f(X,A_n) \\ \sum_{n=1}^{N}(f(X,A_n)-b_n)\partial_{x_2}f(X,A_n) \\ \dots \\ \sum_{n=1}^{N}(f(X,A_n)-b_n)\partial_{x_M}f(X,A_n) \end{bmatrix}$$

$$= \begin{bmatrix} \partial_{x_1}f(X,A_1) & \partial_{x_1}f(X,A_2) & \dots & \partial_{x_1}f(X,A_N) \\ \partial_{x_2}f(X,A_1) & \partial_{x_2}f(X,A_2) & \dots & \partial_{x_2}f(X,A_N) \\ \dots & \dots & \dots & \dots \\ \partial_{x_M}f(X,A_1) & \partial_{x_M}f(X,A_2) & \dots & \partial_{x_M}f(X,A_N) \end{bmatrix} \begin{bmatrix} f(X,A_1)-b_1 \\ f(X,A_2)-b_2 \\ \dots \\ f(X,A_N)-b_N \end{bmatrix}$$

$$= (\nabla_X^T f)R \qquad \textit{relation between Jacobian of E and f}$$

where
$$\nabla_X f = \begin{bmatrix} \partial_{x_1}f(X,A_1) & \partial_{x_2}f(X,A_1) & \dots & \partial_{x_M}f(X,A_1) \\ \partial_{x_1}f(X,A_2) & \partial_{x_2}f(X,A_2) & \dots & \partial_{x_M}f(X,A_2) \\ \dots & \dots & \dots & \dots \\ \partial_{x_1}f(X,A_N) & \partial_{x_2}f(X,A_N) & \dots & \partial_{x_M}f(X,A_N) \end{bmatrix}$$

Matrix convention in this document :

- *different data in different rows*
- *different dimension in different columns*
- *number of data N*
- *number of dimension M*

Lets expand Hessian :

$$\nabla_{XX}^2 E \;= \begin{bmatrix} \partial_{x_1x_1}^2 E & \partial_{x_1x_2}^2 E & \dots & \partial_{x_1x_M}^2 E \\ \partial_{x_2x_1}^2 E & \partial_{x_2x_2}^2 E & \dots & \partial_{x_2x_M}^2 E \\ \dots & \dots & \dots & \dots \\ \partial_{x_Mx_1}^2 E & \partial_{x_Mx_2}^2 E & \dots & \partial_{x_Mx_M}^2 E \end{bmatrix}$$

Consider the i^{th} row and j^{th} column :

$$\begin{aligned} [\nabla_{XX}^2 E]_{i,j} &= \partial_{x_j} \textstyle\sum_{n=1}^{N} (f(X, A_n) - b_n) \partial_{x_i} f(X, A_n) \\ &= \textstyle\sum_{n=1}^{N} (f(X, A_n) - b_n) \partial_{x_i x_j}^2 f(X, A_n) + \sum_{n=1}^{N} \partial_{x_j} f(X, A_n) \partial_{x_i} f(X, A_n) \end{aligned}$$

Hence we have :

$$\nabla_{XX}^2 E \;= \textstyle\sum_{n=1}^{N} (f(X, A_n) - b_n) \nabla_{XX}^2 f(X, A_n) + (\nabla_X^T f) \nabla_X f$$

$$where \quad \nabla_{XX}^2 f(X, A_n) \;= \begin{bmatrix} \partial_{x_1x_1}^2 f(X, A_n) & \partial_{x_1x_2}^2 f(X, A_n) & \dots & \partial_{x_1x_M}^2 f(X, A_n) \\ \partial_{x_2x_1}^2 f(X, A_n) & \partial_{x_2x_1}^2 f(X, A_n) & \dots & \partial_{x_2x_M}^2 f(X, A_n) \\ \dots & \dots & \dots & \dots \\ \partial_{x_Mx_1}^2 f(X, A_n) & \partial_{x_Mx_2}^2 f(X, A_n) & \dots & \partial_{x_Mx_M}^2 f(X, A_n) \end{bmatrix}$$

Please note that Jacobian and Hessian matrices are of different sizes :

$$\begin{aligned} &sizeof(\nabla_X E) = 1 \times M \\ &sizeof(\nabla_{XX}^2 E) = M \times M \\ &sizeof(\nabla_X f) = N \times M \\ &sizeof(\nabla_{XX}^2 f) \quad = M \times M \end{aligned}$$

Reference

https://neos-guide.org/guide/types/least-squares/