1. Least Squares Line

Problem: Find the equation of a line \( y = f(x) = Ax + B \) that is the best fit to some data \( \{(x_i,y_i)\}_{i=1}^N \) in a least squares sense. Find \( A \) and \( B \) that minimizes

\[
R(A,B) = \frac{1}{2} \sum_{i=1}^{N} [(Ax_i + B) - y_i]^2.
\]

Technique: take the partial derivatives of \( R(A,B) \) with respect to \( A \) and \( B \), set them equal to zero, and solve for \( A \) and \( B \):

\[
\frac{\partial R}{\partial A} = \sum_{i=1}^{N} [(Ax_i + B) - y_i]x_i = A \sum_{i=1}^{N} (x_i)^2 + B \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i y_i = 0 \tag{1}
\]

\[
\frac{\partial R}{\partial B} = \sum_{i=1}^{N} [(Ax_i + B) - y_i] = A \sum_{i=1}^{N} x_i + BN - \sum_{i=1}^{N} y_i = 0 \tag{2}
\]

This leads to a 2 x 2 linear system of equations

\[
A \sum_{i=1}^{N} (x_i)^2 + B \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i \tag{3}
\]

\[
A \sum_{i=1}^{N} x_i + BN = \sum_{i=1}^{N} y_i
\]

which can be solved by any method you want.

2. Linearized Nonlinear Least Squares

Convert the potentially nonlinear relationship to a linear one.

Example Problem: Find the function \( f(x) = Ce^{Ax} \) that is the best fit to some data \( \{(x_i,y_i)\}_{i=1}^N \) in a least squares sense. Find \( C \) and \( A \) that minimizes

\[
R(C,A) = \frac{1}{2} \sum_{i=1}^{N} \left[ Ce^{Ax_i} - y_i \right]^2.
\]

Linearize the problem. If

\[ y_i = Ce^{Ax_i} \]

then

\[ Y_i = \ln(y_i) = Ax_i + \ln(C) = Ax_i + B \]

is a least squares line problem for the data \( \{(x_i,Y_i)\}_{i=1}^N \). Do this as above, find \( A \) and \( B \), and then

\[ C = e^B \]

This same technique can be applied to a surprising number relationships such

\[
f(x) = Ce^{Ax} \quad \text{let} \quad Y_i = \ln(y_i) \quad \text{and} \quad X_i = x_i \tag{4}
\]

\[
f(x) = Cx^A \quad \text{let} \quad Y_i = \ln(y_i) \quad \text{and} \quad X_i = \ln(x_i) \tag{5}
\]

\[
f(x) = \frac{L}{1 + Ce^{Ax}} \quad \text{let} \quad Y_i = \ln \left( \frac{L}{y_i} - 1 \right) \quad \text{and} \quad X_i = x_i \quad \text{where} \quad L \text{ is given} \tag{6}
\]
3. Non-Linear Least Squares (Gauss-Newton Method)

Let $\gamma$ be a column vector of function parameters. In the previous examples $\gamma = [\gamma_1, \gamma_2]^T = [A, B]^T$. We want to minimize $R(\gamma)$ renamed $F(\gamma)$.

$$R(\gamma) = \frac{1}{2} \sum_{i=1}^{N} [f(\gamma, x_i) - y_i]^2$$  \hspace{1cm} (7)

$$F(\gamma) = \frac{1}{2} \sum_{i=1}^{N} [f_i(\gamma)]^2 = \frac{1}{2} f^T(\gamma) f(\gamma)$$  \hspace{1cm} (8)

where $f(\gamma) = [f_1(\gamma), f_2(\gamma), \ldots, f_N(\gamma)]^T$. Notice $F: \mathbb{R}^m \rightarrow \mathbb{R}$, where $m$ is the number of parameters.

Let $g$ denote the gradient of $F$ and $G$ denote the Hessian of $F$. Furthermore, let $J$ denote the Jacobian of $f$ and $G_i$ denote the Hessian of $f_i$. It can then be shown that

$$g = J^T f$$

$$G = J^T J + Q$$

$$Q = \sum_{i=1}^{N} f_i G_i$$

Now we want to take $g$, the gradient of $F$, and set it equal to zero.

$$g(\gamma) = [g_1, \ldots, g_m]^T = \left[ \frac{\partial F}{\partial \gamma_1}, \ldots, \frac{\partial F}{\partial \gamma_m} \right]^T \quad g: \mathbb{R}^m \rightarrow \mathbb{R}^m \quad \text{solve } g(\gamma) = 0 \text{ for } \gamma$$

It can be further shown that the Jacobian of $g$ is equal to the Hessian of $F$. Linearizing the gradient about some initial guess at the parameters $= \gamma_0$, and setting equal to zero yields:

$$g(\gamma_0 + h) \approx g(\gamma_0) + G(\gamma_0) \cdot h = 0$$  \hspace{1cm} (9)

Plugging in our special form of $G$ yields

$$(J^T J + Q) \cdot h_0 = -J^T(\gamma_0) f(\gamma_0)$$  \hspace{1cm} (10)

$$\gamma_1 = h + \gamma_0$$

The small residual assumption implies $Q \rightarrow 0$ as $\gamma$ approaches the solution so we continue the following procedure until a convergence condition is satisfied.

$$J^T(\gamma_k) J(\gamma_k) \cdot h_k = -J^T(\gamma_k) f(\gamma_k)$$  \hspace{1cm} (11)

$$\gamma_{k+1} = h_k + \gamma_k$$

where $J_{ij} = \frac{\partial f_i}{\partial \gamma_j}$ is an $N$ by $m$ matrix.