# Curve - Fitting

- Given some data, for example experimental Conc. values at different time points, we may want to determine values at other time points.

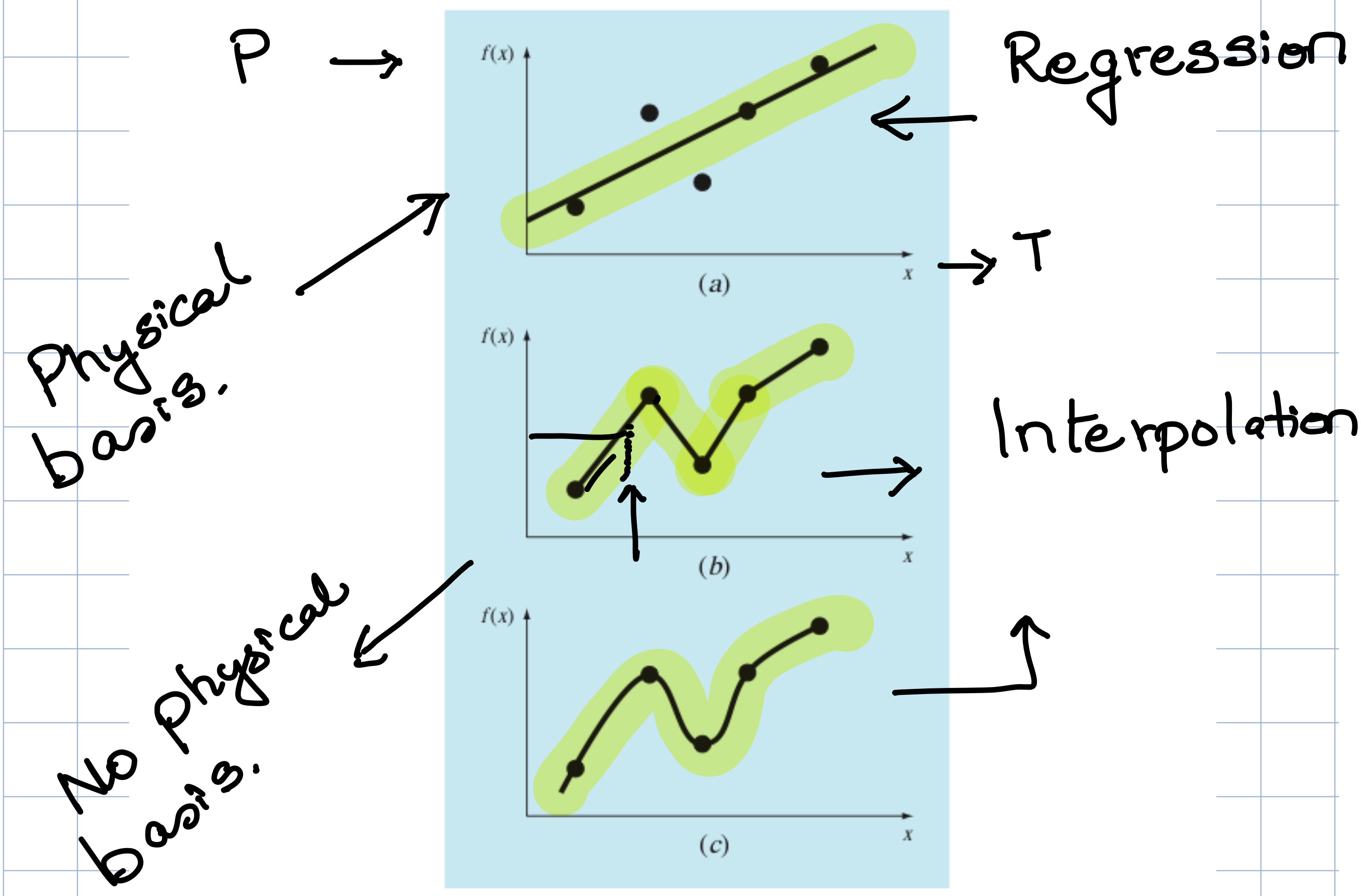



**FIGURE PT5.1**
Three attempts to fit a "best" curve through five data points. (*a*) Least-squares regression, (*b*) linear interpolation, and (*c*) curvilinear interpolation.

# Regression vs. Interpolation

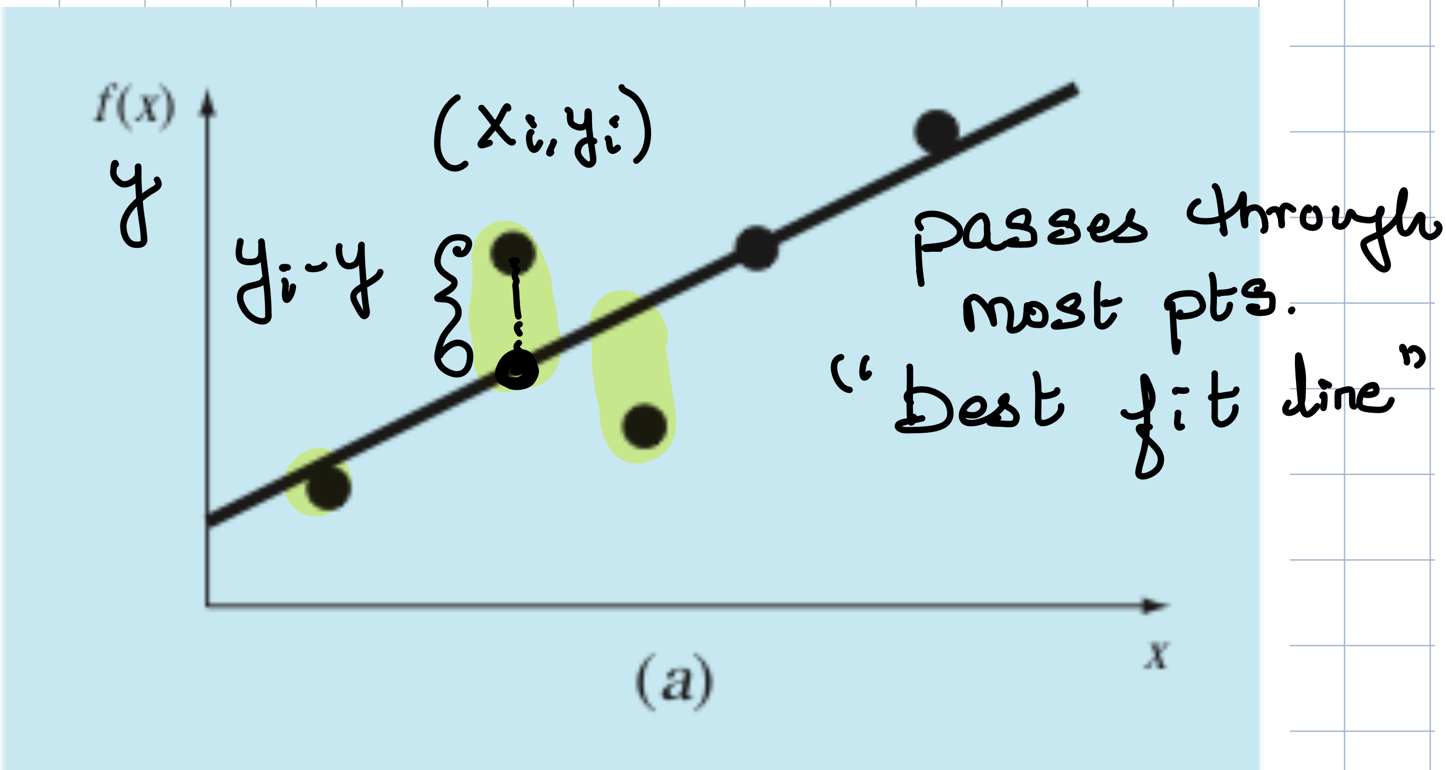
⇓

- Mathematical relationship between variables assumed.
  Linear regression → $y = mx + C$
- Deviation from this equation is due to experi- error.
- Best fit line that minimizes error.
- Data pts. more than # of unknown parameters.

## Interpolation

- Mathematical function/eq that passes through all data pts.

# Linear Regression




$(x_1, y_1)$ $(x_2, y_2)$ . . . . . $(x_n, y_n)$

Find $y = a_0 + a_1 x$

Residual : $y_i - y = y_i - (a_0 + a_1 x)$

Sum of squares of residuals

$$S_r = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i)\right)^2$$

Minimize $S_r$ to find $a_0, a_1$

$$\frac{\partial S_r}{\partial a_0} = 0 \ , \qquad \frac{\partial S_r}{\partial a_1} = 0$$

$$S_r = \sum \left(y_i - (a_0 + a_1 x_i)\right)^2$$

$$\frac{\partial S_r}{\partial a_0} = \sum 2\left(y_i - (a_0 + a_1 x_i)\right) - 1$$

$$= -2 \sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = \sum 2\left(y_i - (a_0 + a_1 x_i)\right)(-x_i)$$

$$= -2 \sum \left(x_i y_i - x_i (a_0 + a_1 x_i)\right)$$

$$-2 \sum y_i + 2 \sum a_0 + 2 \sum a_1 x_i = 0$$

$$-2 \sum x_i y_i + 2 \sum x_i a_0 + 2 \sum a_1 x_i^2 = 0$$

$$\sum a_0 + a_1 \sum x_i = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i$$

$$\sum a_0 + a_1 \sum x_i = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i$$

$$n a_0 + a_1 \sum x_i = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i$$

$$n a_0 \sum x_i + a_1 \left(\sum x_i\right)^2 = \sum x_i \sum y_i$$

$$n a_0 \sum x_i + n a_1 \sum x_i^2 = n \sum x_i y_i$$

$$\therefore \quad a_1 \left[\left(\sum x_i\right)^2 - n \sum x_i^2\right]$$

$$= \sum x_i \sum y_i - n \sum x_i y_i$$

$$a_1 = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - n \sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \sum_{i=1}^{n} x_i^2}$$

$$a_0 = \frac{\sum_{i=1}^{n} y_i - a_1 \sum_{i=1}^{n} x_i}{n}$$

$$a_0 + a_1 x_1 = y_1$$
$$a_0 + a_1 x_2 = y_2$$
$$\vdots$$
$$a_0 + a_1 x_n = y_n$$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$X A = Y$$
$$X^T X A = X^T Y$$
$$A = (X^T X)^{-1} X^T Y$$

Show, $A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ is

Same as least square estimate from before.

# Quantifying error of linear Regression

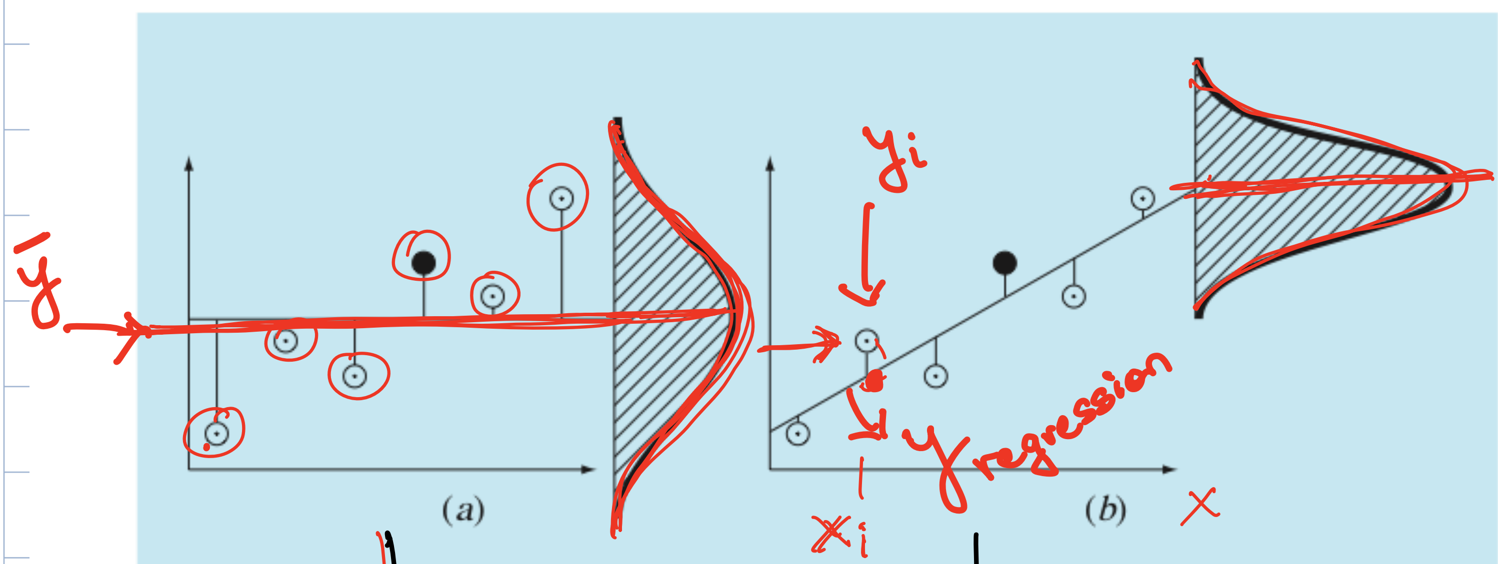
$$S_r = \sum_{i=1}^{n} e_i^2 = \sum (y_i - a_0 - a_1 x_i)^2$$

Coefficient of determination:

$$r^2 = \frac{S_t - S_r}{S_t}$$

$S_r = 0, \quad r^2 = 1$

$S_r = S_t, \quad r^2 = 0$

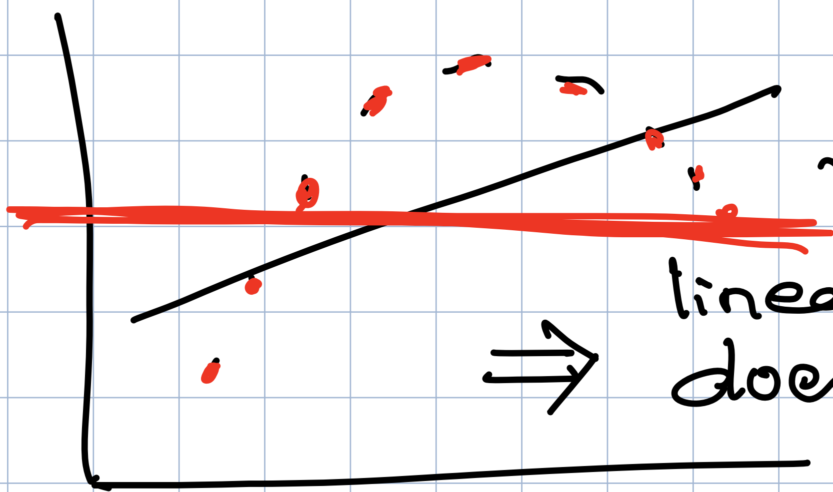



$\downarrow$ $S_t$ $\quad\quad$ $\downarrow$ $S_r$

$\hookrightarrow \sum (y_i - \bar{y})^2$

If $S_t = S_r \Rightarrow r^2 = 0$
$\Rightarrow$ No fit.

$S_r = 0 \Rightarrow r^2 = 1$
$\Rightarrow$ perfect fit



$r^2$ low

linear line $\Rightarrow$ does not describe this data.

# Linearization of nonlinear relationships

Linear regression can also be applied to nonlinear eqs.

* $y = \alpha_1 e^{\beta_1 x}$ exponential eq.

$$\ln y = \ln \alpha_1 + \beta_1 x$$

$$a_0 \equiv \ln \alpha_1 , \quad a_1 \equiv \beta_1$$

* $y = \alpha_2 x^{\beta_2}$ Power eqn.

$$\ln y = \ln \alpha_2 + \beta_2 \ln x$$

$\ln y$ vs. $\ln x$

$$a_0 = \ln \alpha_2 , \quad a_1 = \beta_2$$

# Polynomial Regression

If relationship is best described as polynomial,

$$y = a_0 + a_1 x + a_2 x^2$$

Sum of squares of residuals

$$S_r = \sum_{i=1}^{n} \left( y_i - (a_0 + a_1 x_i + a_2 x_i^2) \right)^2$$

To obtain best fit, compute

$$\frac{\partial S_r}{\partial a_0}, \frac{\partial S_r}{\partial a_1}, \frac{\partial S_r}{\partial a_2} = 0$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum \left( y_i - a_0 - a_1 x_i - a_2 x_i^2 \right)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum x_i \left( y_i - a_0 - a_1 x_i - a_2 x_i^2 \right)$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum x_i^2 \left( y_i - a_0 - a_1 x_i - a_2 x_i^2 \right)$$

Compute $a_0, a_1, a_2$ by setting above to 0.

Easier to frame this as a set of linear equations,

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & & \\ 1 & x_n & x_n^2 \end{bmatrix}}_{n \times 3} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}}_{3 \times 1} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{n \times 1}$$

$$X A = Y$$

$$X^T X A = X^T Y$$

$$A = (X^T X)^{-1} X^T Y$$

Same as $\frac{\partial S_r}{\partial a_i} = 0$

If $y$ is a fn of several variables,

$$y = f(x_1, x_2, x_3, \ldots)$$

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & & & \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or

$$S_r = \Sigma (y_i - a_0 - a_1 x_{1i} + a_2 x_{2i} - a_3 x_{3i})^2$$

Minimize $S_r$ wrt $a_0, a_1, a_2, a_3$.

$x_1$ – Variable

$x_{11}$, $x_{12}$ – readin for $x_1$

| $x_i$ | $y_i$ |
|---|---|
| 0 | 2.1 |
| 1 | 7.7 |
| 2 | 13.6 |
| 3 | 27.2 |
| 4 | 40.9 |
| 5 | 61.1 |

Find the best-fit linear or Polynomial eq for the given data.

Polynomial Regression: $y = a_0 + a_1 x + a_2 x^2$

Linear Regression : $y = a_0 + a_1 x$

| $x_i$ | $y_i$ |
|---|---|
| 0 | 2.1 |
| 1 | 7.7 |
| 2 | 13.6 |
| 3 | 27.2 |
| 4 | 40.9 |
| 5 | 61.1 |

Find the best-fit linear or Polynomial eq for the given data.

Polynomial Regression: $y = a_0 + a_1 x + a_2 x^2$

Linear Regression : $y = a_0 + a_1 x$

Linear Regression

$$\underset{X}{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \underset{Y}{\begin{bmatrix} 2.1 \\ 7.7 \\ \vdots \\ 61.1 \end{bmatrix}}$$

$$\begin{bmatrix} -3.7 \\ 11.6 \end{bmatrix}$$

$$X^T X A = X^T Y \Rightarrow A = \begin{bmatrix} -3.7 \\ 11.6 \end{bmatrix}$$

| $x_i$ | $y_i$ | $(y_i - \bar{y})^2$ | $(y_i - y)^2$ |
|---|---|---|---|
| 0 | 2.1 | 544.44 | 33.92 |
| 1 | 7.7 | 314.47 | 0.06 |
| 2 | 13.6 | 140.03 | 36.03 |
| 3 | 27.2 | 3.12 | 16.53 |
| 4 | 40.9 | 239.22 | 4.11 |
| 5 | 61.1 | 1272.11 | 42.36 |

$S_t \rightarrow \Sigma = 2513.39$ $\quad S_r = \Sigma\, 133$

$\bar{y} = 25.43$

$$r^2 = \frac{2513.39 - 133}{2513.39}$$

$$= 0.9470.$$

$$y = a_0 + a_1 x$$

$$y = -3.7 + 11.6x$$

$x_i = 0, \quad y_i = -3.7$

$(y_i - y)^2 = (2.1 + 3.7)^2 =$

# Polynomial Regression

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_6 & x_6^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.1 \\ \vdots \\ 61.1 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.4786 \\ 2.3593 \\ 1.8607 \end{bmatrix}$$

$$r^2 = 0.99$$

# Nonlinear Regression

Consider, $f(x) = a_0(1 - e^{-a_1 x})$

This equation cannot be linearized.

Determine $a_0, a_1$ given data $x_i, y_i$ where $y_i = f(x_i)$

Choose $a_0, a_1$ such that residual is minimized:

Residual: $\sum_{i=1}^{n}\left(y_i - a_0\left(1 - e^{-a_1 x_i}\right)\right)^2$

Gauss-Newton Method:

Approx $f(x)$ as taylor series around $a_0, a_1$

$$f(x_i, a_0^{(k+1)}, a_1^{(k+1)}) = f(x_i) + \frac{\partial f}{\partial a_0}\Delta a_0 + \frac{\partial f}{\partial a_1}\Delta a_1$$

If $a_0^{(k+1)}$ and $a_1^{(k+1)}$ are the true parameters $f(x_i, a_0^{(k+1)}, a_1^{(k+1)}) = y_i$

$$\therefore \quad y_i = f(x_i) + \left.\frac{\partial f}{\partial a_0}\right|_{x_i, a_0^{(k)}, a_1^{(k)}} \Delta a_0 + \left.\frac{\partial f}{\partial a_1}\right|_{x_i, a_0^{(k)}, a_1^{(k)}} \Delta a_1$$

for all $i = 1, \ldots n$

Above can be written as a

$$\underbrace{\begin{bmatrix} y_1 - f(x_1) \\ y_2 - f(x_2) \\ \vdots \\ y_n - f(x_n) \end{bmatrix}}_{D} = \underbrace{\begin{bmatrix} \left.\frac{\partial f}{\partial a_0}\right|_{x_1} & \left.\frac{\partial f}{\partial a_1}\right|_{x_1} \\ \left.\frac{\partial f}{\partial a_0}\right|_{x_2} & \left.\frac{\partial f}{\partial a_1}\right|_{x_2} \\ \vdots & \vdots \end{bmatrix}}_{Z} * \underbrace{\begin{bmatrix} \Delta a_0 \\ \Delta a_1 \end{bmatrix}}_{\Delta A}$$

$$\therefore \quad D = Z \Delta A$$

where $Z$ is not a square matrix

Multiply both sides by $Z^T$

or

$$Z^T D = Z^T Z \Delta A$$

$$Z^T D = Z^T Z \Delta A$$

or

$$\Delta A = (Z^T Z)^{-1} Z^T D$$

Use this to obtain improved estimate of $a_0$, $a_1$

$$a_0^{(k+1)} = a_0^{(k)} + \Delta a_0$$

$$a_1^{(k+1)} = a_1^{(k)} + \Delta a_1$$

where, $\Delta A = \begin{bmatrix} \Delta a_0 \\ \Delta a_1 \end{bmatrix}$

Continue iterations until solution converges, or

$$\varepsilon = \left| \frac{a_0^{(k+1)} - a_0^{(k)}}{a_0^{(k+1)}} \right| < tol$$

and

$$\left| \frac{a_1^{(k+1)} - a_1^{(k)}}{a_1^{(k+1)}} \right| < tol$$

and residual is less than tol.

Example, $f(x) = a_0(1 - e^{-a_1 x})$

Data

| | | | | | |
|---|---|---|---|---|---|
| x | 0.25 | 0.75 | 1.25 | 1.75 | 2.25 |
| y | 0.28 | 0.57 | 0.68 | 0.74 | 0.79 |

$a_0 = 1$, $a_1 = 1$ Initial Guess.

Solution
$$\frac{\partial f}{\partial a_0} = 1 - e^{-a_1 x}$$
$$\frac{\partial f}{\partial a_1} = + a_0 x e^{-a_1 x}$$

$$\text{Jacobian } (Z) = \begin{bmatrix} 1 - e^{-a_1 x_1} & a_0 x_1 e^{-a_1 x_1} \\ 1 - e^{-a_1 x_2} & a_0 x_2 e^{-a_1 x_2} \\ \vdots & \\ 1 - e^{-a_1 x_5} & a_0 x_2 e^{-a_1 x_5} \end{bmatrix}$$

$$D = \begin{bmatrix} y_1 - f(x_1) \\ \vdots \\ y_5 - f(x_5) \end{bmatrix} = \begin{bmatrix} 0.28 - a_0(1 - e^{-a_1 x_1}) \\ \vdots \\ 0.79 - a_0(1 - e^{-a_1 x_5}) \end{bmatrix}$$

$$D = J\,\Delta A \quad \text{where}$$

$$\Delta A = \begin{bmatrix} a_0^{k+1} - a_0^{k} \\ a_1^{k+1} - a_1^{k} \end{bmatrix}$$

$$a_0^1 = 1 \quad , \quad a_1^1 = 1$$

Solve, $D = J\,\Delta A$

Multiply $J^T$

$$J^T D = J^T J\,\Delta A$$

or

$$(J^T J)\,\Delta A = J^T D$$

$$\Delta A = (J^T J)^{-1} J^T D$$

Substitute $a_0^1$, $a_1^1$ above to get,

$$\Delta A = \begin{bmatrix} -0.2714 \\ 0.5019 \end{bmatrix}$$

or $$\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix} + \begin{pmatrix} -0.2714 \\ 0.5019 \end{pmatrix}$$

or $a_0^2 = 0.7286$, $a_1^2 = 1.5019$

For $a_0^1, a_1^1 \rightarrow S_r = 0.0248$

$a_0^1, a_1^1 \rightarrow S_r = 0.0242$

$\vdots$

$a_0 = 0.79186$, $a_1 = 1.6751$

$S_r = 0.000662$.