lecture 26 — linear curve

$(x_1, y_1), \ldots, (x_n, y_n)$, n data set

$\hat{y} = \hat{y}(x) = a_1 + a_2 x$, $\hat{y}_i = \hat{y}(x_i) = a_1 + a_2 x_i$

$(x_1, \hat{y}_1), \ldots, (x_n, \hat{y}_n)$ output from model

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = E(a_1, a_2)$$

find a_1, a_2 s.t. E is minimum

$$\frac{\partial E}{\partial a_1} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - \hat{y}_i)\,1 = 0 \Rightarrow \boxed{a_1 n + a_2\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} y_i}$$

$$\frac{\partial E}{\partial a_2} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - \hat{y}_i)\,x_i = 0 \Rightarrow \boxed{a_1\left(\sum_{i=1}^{n} x_i\right) + a_2\left(\sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n} x_i y_i}$$

$$\Rightarrow J a = b, \quad a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad b = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$J = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$a = J \backslash b$$

$$\hat{y}(x) = \underbrace{[1, \; x]}_{1\times 2} \underbrace{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{2\times 1}, \quad \underbrace{\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}}_{n\times 2} \underbrace{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}}_{2\times 1}$$

B matrix

$$\bullet\; B^T B = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = J$$

$$\bullet\; B^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = b$$

Quadratic curve

$$\hat{y} = \hat{y}(x) = a_1 + a_2 x + a_3 x^2 \quad , \hat{y}_i = a_1 + a_2 x_i + a_3 x_i^2$$

$$\Downarrow$$

$$\frac{\partial \hat{y}_i}{\partial a_1} = 1$$

$$\frac{\partial \hat{y}_i}{\partial a_2} = x_i \, , \frac{\partial \hat{y}_i}{\partial a_3} = x_i^2$$

$$E = \sum (y_i - \hat{y}_i)^2 = E(a_1, a_2, a_3)$$

$$\frac{\partial E}{\partial a_1} = 0 \Rightarrow \sum 2(y_i - \hat{y}_i)\Big(\cancelto{0}{\frac{\partial y_i}{\partial a_1}} - \cancelto{1}{\frac{\partial \hat{y}_i}{\partial a_1}}\Big) = 0$$

$$\Rightarrow (-2) \sum (y_i - \hat{y}_i)\, 1 = 0$$

$$\Rightarrow \sum (y_i - a_1 - a_2 x_i - a_3 x_i^2) = 0$$

$$\Rightarrow \boxed{a_1 n + a_2 (\sum x_i) + a_3 (\sum x_i^2) = \sum y_i}$$

$$\frac{\partial E}{\partial a_2} = 0 \Rightarrow \boxed{a_1 (\sum x_i) + a_2 (\sum x_i^2) + a_3 (\sum x_i^3) = \sum x_i y_i}$$

$$\frac{\partial E}{\partial a_3} = 0 \Rightarrow \boxed{a_1 (\sum x_i^2) + a_2 (\sum x_i^3) + a_3 (\sum x_i^4) = \sum x_i^2 y_i}$$

$J a = b$ where $a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$, $b = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$

$$J = \begin{bmatrix} n & \sum x_i , & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix}$$

$$\hat{y}(x) = \underbrace{[1, x, x^2]_{1\times 3}}_{z(x)} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = z(x)\, a \quad , a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} — & Z(x_1) & — \\ — & Z(x_2) & — \\ & \vdots & \\ — & Z(x_n) & — \end{bmatrix}_{n\times 3} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}_{3\times 1}$$

$$= \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & & \\ 1 & x_n & x_n^2 \end{bmatrix}}_{B} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

$$J = B^T B, \qquad b = B^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\Rightarrow \boxed{Ja = b}$$

General Linear regression Least square method

$$\hat{y} = \hat{y}(x) = Z(x)\, a = [z_1(x), z_2(x), \ldots, z_m(x)]_{1\times m} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}_{m\times 1}$$

<u>n data</u> $(x_1, y_1), \ldots, (x_n, y_n)$

Solve the following problem to get a

$$Ja = b,$$

$$J = B^T B, \qquad b = B^T y = B^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$B = \begin{bmatrix} — Z(x_1) — \\ — Z(x_2) — \\ \vdots \\ — Z(x_n) — \end{bmatrix}_{n \times m} = \begin{bmatrix} z_1(x_1) & z_2(x_1) & \cdots & z_m(x_1) \\ z_1(x_2) & z_2(x_2) & \cdots & z_m(x_2) \\ \vdots & & & \\ z_1(x_n) & z_2(x_n) & \cdots & z_m(x_n) \end{bmatrix}$$

Examples of basis vectors

(i) Linear curve $z = [1, x]$, $z_1(x) = 1$, $z_2(x) = x$

(ii) quadratic curve $z = [1, x, x^2]$, $z_1 = 1$, $z_2 = x$, $z_3 = x^2$

(iii) $(m-1)^{th}$ polynomial curve

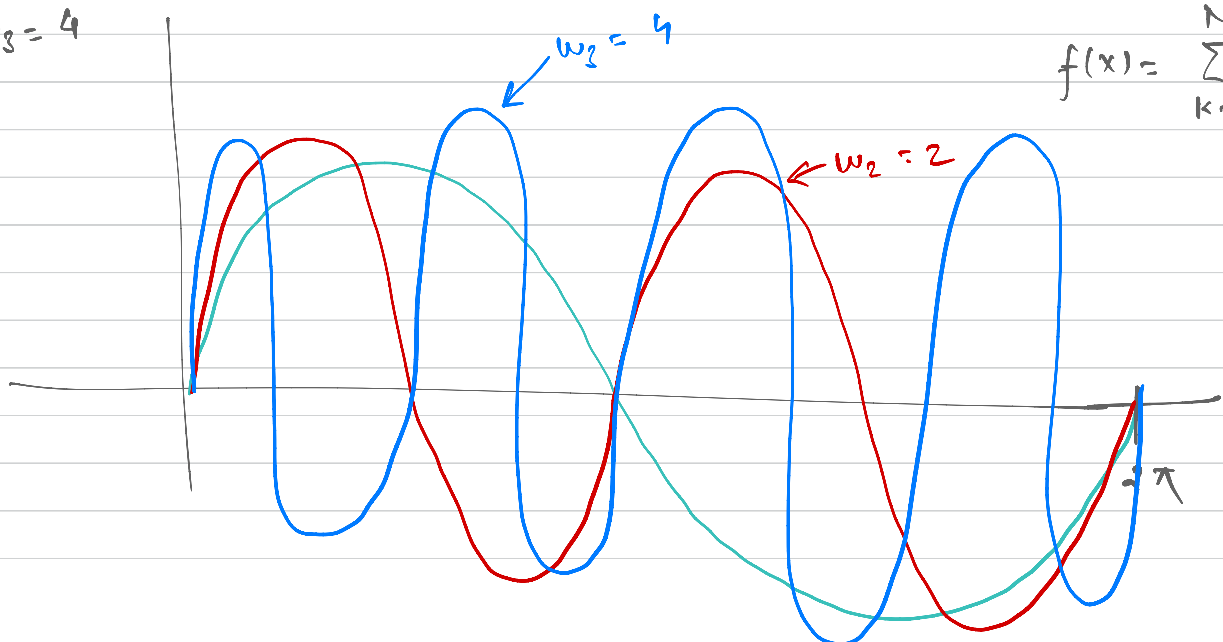
$$z = [1, x, x^2, \ldots, x^{m-1}]_{1 \times m}$$

(iv) Fourier basis (using m bases)

$$z(x) = [1, \sin(w_1 x), \sin(w_2 x), \ldots, \sin(w_{m-1} x)]_{1 \times m}$$

& here $w_1, w_2, \ldots, w_{m-1}$ are fixed frequencies of basis

$w_1 = 1$
$w_2 = 2$
$w_3 = 4$



$$f(x) = \sum_{k=1}^{N} \alpha_k \sin(w_k x)$$

$w_k = k - 1$

(iv) Exponential basis

$$z(x) = [1, e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots,]$$

$i < j$

$$z_i = \sin(\omega_i x) = z_j = \sin(\omega_j x) \rightarrow \omega_i = \omega_j.$$

$$\hat{y}(x) = a_1 + a_2 \sin(\omega_2 x) + \ldots + a_{i-1} \sin(\omega_{i-1} x)$$

$$+ a_i \sin(\omega_i x)$$

$$+ \ldots + a_{j-1} \sin(\omega_{j-1} x)$$

$$+ a_j \sin(\omega_j x)$$

$$+ \ldots + a_m \sin(\omega_m x)$$

$$a_i \sin(\omega_i x)$$

$$+ a_j \sin(\omega_j x)$$

$$= \underbrace{(a_i + a_j)}_{10} \sin(\omega_j x)$$

$a_i = 10, \quad a_j = 0 \quad , \quad a_i = 5, \quad a_j = 5$

Post lecture: Deriving matrix-vector $Ja = b$ for general linear regression

$\hat{y}(x) = z(x)\, a$ $\qquad z(x) = [z_1(x), z_2(x), \ldots, z_m(x)]_{1\times m}$

basis $\qquad$ vector of unknown $\qquad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}_{m\times 1}$

model for data $(x_1, y_1), \ldots, (x_n, y_n)$

let $\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$, $\quad \underline{\hat{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n\times 1} = \underbrace{\begin{bmatrix} z_1(x_1) & z_2(x_1) & \cdots & z_m(x_1) \\ z_1(x_2) & z_2(x_2) & \cdots & z_m(x_2) \\ \vdots & & & \\ z_1(x_n) & z_2(x_n) & \cdots & z_m(x_n) \end{bmatrix}_{n\times m}}_{B} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}_{m\times 1}$

$\Rightarrow \underline{\hat{y}} = B\, a \qquad B = \begin{bmatrix} - & z(x_1) & - \\ - & z(x_2) & - \\ & \vdots & \\ - & z(x_n) & - \end{bmatrix}_{n\times m}$

Squared errors $\qquad E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$$= (\underline{y} - \underline{\hat{y}})^T (\underline{y} - \underline{\hat{y}})$$

for $j = 1, 2, \ldots, m$

$$\frac{\partial E}{\partial a_j} = \sum_{i=1}^{n} \frac{\partial}{\partial a_j} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} 2(y_i - \hat{y}_i)\left(-\frac{\partial \hat{y}_i}{\partial a_j}\right)$$

where $\frac{\partial y_i}{\partial a_j} = 0$, & $\frac{\partial \hat{y}_i}{\partial a_j} = \frac{\partial}{\partial a_j}\left[z(x_i)\, a\right]$

$$= \frac{\partial}{\partial a_j}\left\{ [z_1(x_i), z_2(x_i), \ldots, z_m(x_i)] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \right\}$$

$$= \frac{\partial}{\partial a_j}\left\{ z_1(x_i)\, a_1 + z_2(x_i)\, a_2 + z_3(x_i)\, a_3 + \dots + z_m(x_i)\, a_m \right\}$$

$$= z_j(x_i)$$

for $j = 1, 2, \dots, m$

$$\frac{\partial E}{\partial a_j} = \sum_{i=1}^{n} 2\,(y_i - \hat{y}_i)\; z_j(x_i)$$

$$= 2\left[\sum_{i=1}^{n} y_i\, z_j(x_i) - \sum_{i=1}^{n} \hat{y}_i\, z_j(x_i)\right]$$

$$\Rightarrow \frac{\partial E}{\partial a_j} = 0 \Rightarrow \sum_{i=1}^{n} \hat{y}_i\, z_j(x_i) = \sum_{i=1}^{n} y_i\, z_j(x_i)$$

$$\because \hat{y}_i = z(x_i)\, a$$

$$\therefore \hat{y}_i\, z_j(x_i) = z_j(x_i)\; z(x_i)\; a$$

$$\Rightarrow \sum_{i=1}^{n} \hat{y}_i\, z_j(x_i) = \sum_{i=1}^{n} z_j(x_i)\; z(x_i)\, a$$

$$= z_j(x_1)\, z(x_1)\, a + z_j(x_2)\, z(x_2)\, a + \dots + z_j(x_n)\, z(x_n)\, a$$

$$\because z(x_i) = [z_1(x_i), z_2(x_i), \dots, z_m(x_i)]$$

$$= \left(z_j(x_1)\, z(x_1)\right) a + \left(z_j(x_2)\, z(x_2)\right) a + \dots + \left(z_j(x_n)\, z(x_n)\right) a$$

$$= \left([z_j(x_1), z_j(x_2), \dots, z_j(x_n)]_{1\times n} \underbrace{\begin{bmatrix} z(x_1) \\ z(x_2) \\ \vdots \\ z(x_n) \end{bmatrix}_{n\times m}}_{B_{n\times m}} \right) a_{m\times 1}$$

$$= [z_j(x_1), \dots, z_j(x_n)]\; B\; a$$

$\Rightarrow$ $\quad [z_j(x_1), \ldots, z_j(x_n)]\, B\, a = \underbrace{\sum_{i=1}^{n} z_j(x_i)\, y_i}_{b_j}$ $\Rightarrow$ b_j, $j = 1, 2, \ldots, m$

Stacking above gives, $j = 1, 2, \ldots, m$,

$$\underbrace{\begin{bmatrix} z_1(x_1) & z_1(x_2) & \ldots & z_1(x_n) \\ z_2(x_1) & z_2(x_2) & \ldots & z_2(x_n) \\ \vdots & & & \\ z_m(x_1) & z_m(x_2) & \ldots & z_m(x_n) \end{bmatrix}}_{\text{it is } B^T \text{ (check!)}} B\, a = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$\Rightarrow$ $\quad B^T B\, a = b$

Now $\quad b_j = \sum_{i=1}^{n} z_j(x_i)\, y_i$

$$= [z_j(x_1), z_j(x_2), \ldots, z_j(x_n)] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \rightarrow \underline{y}$$

$\therefore$

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \underbrace{\begin{bmatrix} z_1(x_1) & z_1(x_2) & \ldots & z_1(x_n) \\ \vdots & & & \\ z_m(x_1) & z_m(x_2) & \ldots & z_m(x_n) \end{bmatrix}}_{\text{it is } B^T} \underline{y}$$

$\Rightarrow$ $\quad b = B^T \underline{y}$

Then solve for $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ using $\quad \underbrace{B^T B}_{J \text{ matrix}}\, a = \underbrace{B^T \underline{y}}_{b \text{ vector}}$

- **Linear regression with multiple variables**

So far we considered data of type $(x_1, y_1), \ldots, (x_n, y_n)$ where x_i, y_i are scalars.

We could have data of the form

$(x_1, y_1), \quad (x_2, y_2), \quad \ldots, \quad (x_n, y_n)$

where y_i scalar

$$x_i = \begin{bmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^m \end{bmatrix} \quad \text{vector of } m \text{ components.}$$

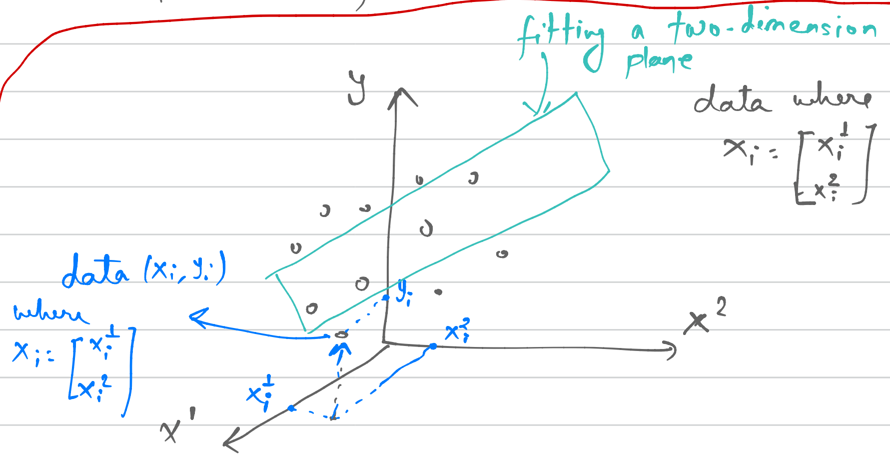
To fit this "m-dimensional" data using an m-dimensional plane in $m+1$ - dimensional space (m from x_i and 1 from y_i)

Take the model (m-dimensional plane, i.e., linear function of m arguments)

$$\hat{y} = \hat{y}(x) = a_1 + a_2 x^1 + a_3 x^2 + \ldots + a_{m+1} x^m$$

where

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{bmatrix}$$



in this case, $m = 2$ and plot of data is in $m+1 = 3$ dimension.

$\downarrow$

fit 2-dimension plane (linear in x^1 and x^2)

Defining basis vector $z = z(x)$ as

$$z(x) = [1, x^1, x^2, \dots, x^m]_{1 \times (m+1)}$$ → linear in m arguments ↓ extension of linear curve

where

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{bmatrix}$$

Our general linear regression formula still holds

find $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{m+1} \end{bmatrix}_{(m+1) \times 1}$ s.t.

$$J a = b$$

where

$$J_{(m+1)\times(m+1)} = B^T B, \quad b_{(m+1)\times 1} = B^T y, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

&

$$B_{n \times (m+1)} = \begin{bmatrix} \text{———} & z(x_1) & \text{———} \\ \text{———} & z(x_2) & \text{———} \\ & \vdots & \\ \text{———} & z(x_n) & \text{———} \end{bmatrix}$$

where $x_i = \begin{bmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^m \end{bmatrix}_{m \times 1}$

$$= \begin{bmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^m \\ 1 & x_2^1 & x_2^2 & \dots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n^1 & x_n^2 & \dots & x_n^m \end{bmatrix}$$

We could consider also polynomial curve in m- arguments

E.g. take $m = 2$ and consider 2^{nd} order polynomial of two variables

$$\hat{y}(x) = \underbrace{\begin{bmatrix} 1 & x^1 & x^2 & (x^1)^2 & (x^2)^2 & x^1 x^2 \end{bmatrix}}_{Z(x)} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}$$

Then $Ja = b$, where $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_6 \end{bmatrix}$, $b = B^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $J = B^T B$

and

$$B = \begin{bmatrix} \text{———} & Z(x_1) & \text{———} \\ \text{———} & Z(x_2) & \text{———} \\ & \vdots & \\ \text{———} & Z(x_n) & \text{———} \end{bmatrix}$$

- ## Nonlinear regression

Consider following model for n data $(x_1, y_1), \dots, (x_n, y_n)$

(assume x_i, y_i are scalars)

$$\hat{y}(x) = a_1 + a_2 \sin(a_3 x)$$

where a_1, a_2, a_3 need to be computed.

This is an example of nonlinear regression

This leads to minimization of nonlinear function

Total error $E = E(a_1, a_2, a_3)$

$$= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad , \quad \hat{y}_i = \hat{y}(x_i)$$

find a_1, a_2, a_3 s.t. $E(a_1, a_2, a_3)$ is minimum

$\Downarrow$

Solve nonlinear equations

$$\frac{\partial E}{\partial a_1} = 0 \Rightarrow \sum 2(y_i - \hat{y}_i)(-1) = 0 \Rightarrow \sum \left[y_i - (a_1 + a_2 \sin(a_3 x_i)) \right] = 0$$

$$\frac{\partial E}{\partial a_2} = 0 \Rightarrow \sum 2(y_i - \hat{y}_i)(-\sin(a_3 x_i)) = 0 \Rightarrow \sum \left[y_i - (a_1 + a_2 \sin(a_3 x_i)) \right] \sin(a_3 x_i) = 0$$

$$\frac{\partial E}{\partial a_3} = 0 \Rightarrow \sum 2(y_i - \hat{y}_i) \underbrace{(-a_1 \cos(a_3 x_i)\, x_i)}_{\text{check!!}} = 0$$

$$\Rightarrow \sum \left[y_i - (a_1 + a_2 \sin(a_3 x_i)) \right] a_1 \cos(a_3 x_i)\, x_i = 0$$

Read section 15.5 and example 15.5

$\Downarrow$

"fminsearch" MATLAB method to find minima of nonlinear function.