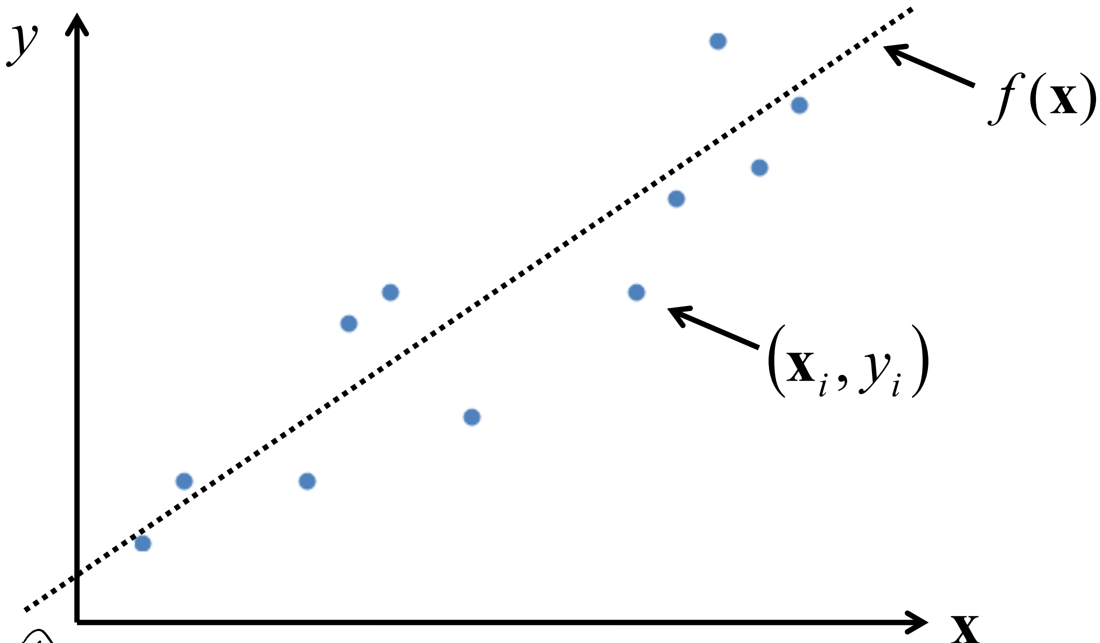
General solution using matrix

Linear Regression & Additive Linear Model

Introduction

- **Find the line which best fits the data**
  - We want to find a line which generalizes the given data

$$(x_{11}, x_{12}, \dots, x_{1d}, y_1)$$
$$(x_{21}, x_{22}, \dots, x_{2d}, y_2)$$
$$\dots$$
$$(x_{i1}, x_{i2}, \dots, x_{id}, y_i)$$
$$\dots$$
$$(x_{n1}, x_{n2}, \dots, x_{nd}, y_n)$$



$$f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d = w_0 + \sum_{j=1}^{d} w_j x_j$$

Introduction

- **We have *n* sample data**

$$D = \{D_1, D_2, \cdots, D_n\} \text{ where } D_i = (\mathbf{x}_i, y_i)$$

$\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})$ is an input in d dimensional space

y_i is the output for $\mathbf{x}_i$

Find $\mathbf{w} = (w_0, w_1, \cdots, w_d)$ which minimizes $E(\mathbf{w})$

$$E(\mathbf{w}) = \sum_{i=1}^{n} (f(\mathbf{x_i}) - y_i)^2$$

$$f(\mathbf{x}_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = w_0 + \sum_{j=1}^{d} w_j x_{ij}$$

Introduction

X_i 1-D sample
d dimensional
d 차원

- **In Other Words**
  - So, our problem can be stated as follows:

<Formulation>

co-efficient model

Find $\mathbf{w} = (w_0, w_1, \cdots, w_d)$ which minimizes $E(\mathbf{w})$ where

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i \right)^2$$

you

$$f(\mathbf{x}_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = w_0 + \sum_{j=1}^{d} w_j x_{ij}$$

linear

$D = \{D_1, D_2, \cdots, D_n\}$ where $D_i = (\mathbf{x}_i, y_i)$

training sample

$\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})$ is an input in d dimensional space

y_i is the output for $\mathbf{x}_i$

Introduction

- **In Other Words**
  - So, our problem can be stated as follows:

Error

Find $\mathbf{w} = (w_0, w_1, \cdots, w_d)$ which minimizes $E(\mathbf{w})$ where

target Value

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i \right)^2$$

$w_j x_{ij}$ 의 형태로 전부 만들려고

$$f(\mathbf{x}_i) = w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = \sum_{j=0}^{d} w_j x_{ij}$$

$$D = \{D_1, D_2, \cdots, D_n\} \text{ where } D_i = (\mathbf{x}_i, y_i)$$

$$\mathbf{x}_i = (x_{i0}, x_{i1}, x_{i2}, \cdots, x_{id}) \text{ where } \boxed{x_{i0} = 1} \text{ for } i = 1, \cdots, \cancel{d}\, n$$

fixed constant

1대입

Introduction

- **How to solve this?**

$$\text{Find } \mathbf{w} = (w_0, w_1, \cdots, w_d) \text{ which minimizes } E(\mathbf{w}) \text{ where}$$

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i\right)^2$$

$w_j x_{ij}$

$$f(\mathbf{x}_i) = w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = \sum_{j=0}^{d} w_j x_{ij}$$

$$D = \{D_1, D_2, \cdots, D_n\} \text{ where } D_i = (\mathbf{x}_i, y_i)$$

$$\mathbf{x}_i = (x_{i0}, x_{i1}, x_{i2}, \cdots, x_{id}) \text{ where } x_{i0} = 1 \text{ for } i = 1, \cdots, n$$

constant

x and y are given values

E is a quadratic function of w's w에대한 함수

- There are various ways to solve
- Here, the SIMPLEST one will be presented

Solution

- **Quadratic Function Optimization**
  - At the point which minimizes $E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i\right)^2$

x_{ij}

$\rightarrow 2\left(f(x_i) - y_i\right) \times \frac{\partial\left(f(x_i) - y_i\right)}{\partial w_j}$

$= 2x_{ij}\left(f(x_i) - y_i\right)$

$$\boxed{\frac{\partial}{\partial w_j} E(w_0, w_1, \cdots, w_d) = 0 \ \text{ for } j = 0, \cdots, d}$$

  - That is

W 각각에 대해 미분한 값은 0

$$\frac{\partial}{\partial w_j} E(\mathbf{w}) = \sum_{i=1}^{n} \boxed{\left(\frac{\partial}{\partial w_j}\left(f(\mathbf{x}_i) - y_i\right)^2\right)} = \sum_{i=1}^{n} 2x_{ij}\left(f(\mathbf{x}_i) - y_i\right) = 0 \quad j \geq 0$$

$w_j x_{ij}$ 항에 대해서만 미분되니까

$$f(\mathbf{x}_i) = w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = \sum_{j=0}^{d} w_i x_{ij}$$

$\rightarrow x_{ij}$ 만 남음

Solution

- **Quadratic Function Optimization**

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i \right)^2$$

$$\left.\begin{aligned}
\frac{\partial}{\partial w_0} E(\mathbf{w}) &= \sum_{i=1}^{n} x_{i0} \left(f(\mathbf{x_i}) - y_i \right) = 0 \\
\frac{\partial}{\partial w_1} E(\mathbf{w}) &= \sum_{i=1}^{n} x_{i1} \left(f(\mathbf{x_i}) - y_i \right) = 0 \\
\frac{\partial}{\partial w_d} E(\mathbf{w}) &= \sum_{i=1}^{n} x_{id} \left(f(\mathbf{x_i}) - y_i \right) = 0
\end{aligned}\right\}$$

There are *d* variables *d* equations.
If we solve the equation system,
we can obtain

$$\mathbf{w} = (w_0, w_1, \cdots, w_d)$$

$$f(\mathbf{x}_i) = w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = \sum_{j=0}^{d} w_i x_{ij}$$

Solution

▪ Quadratic Function Optimization

이항

$$\sum_{i=1}^{n} x_{i0}(f(\mathbf{x_i}) - y_i) = 0 \longrightarrow \sum_{i=1}^{n} x_{i0}(w_0 x_{i0} + w_1 x_{i1} + \cdots + w_d x_{id} - y_i) = 0$$

$$\sum_{i=1}^{n} x_{i1}(f(\mathbf{x_i}) - y_i) = 0 \longrightarrow \sum_{i=1}^{n} x_{i1}(w_0 x_{i0} + w_1 x_{i1} + \cdots + w_d x_{id} - y_i) = 0$$

$$\cdots \qquad\qquad \cdots$$

$$\sum_{i=1}^{n} x_{id}(f(\mathbf{x_i}) - y_i) = 0 \longrightarrow \sum_{i=1}^{n} x_{id}(w_0 x_{i0} + w_1 x_{i1} + \cdots + w_d x_{id} - y_i) = 0$$

$$\Rightarrow$$

$$w_0 \sum_{i=1}^{n} x_{i0}x_{i0} + w_1 \sum_{i=1}^{n} x_{i0}x_{i1} + w_2 \sum_{i=1}^{n} x_{i0}x_{i2} + \cdots + w_d \sum_{i=1}^{n} x_{i0}x_{id} = \sum_{i=1}^{n} x_{i0}y_i$$

$$w_0 \sum_{i=1}^{n} x_{i1}x_{i0} + w_1 \sum_{i=1}^{n} x_{i1}x_{i1} + w_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + w_d \sum_{i=1}^{n} x_{i1}x_{id} = \sum_{i=1}^{n} x_{i1}y_i$$

$$\cdots$$

$$w_0 \sum_{i=1}^{n} x_{id}x_{i0} + w_1 \sum_{i=1}^{n} x_{id}x_{i1} + w_d \sum_{i=1}^{n} x_{id}x_{i2} + \cdots + w_d \sum_{i=1}^{n} x_{id}x_{id} = \sum_{i=1}^{n} x_{id}y_i$$

Solution

- **Quadratic Function Optimization**

$$\mathbf{Aw} = \mathbf{b}$$

$$\mathbf{A} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0} x_{i0}, \sum_{i=1}^{n} x_{i0} x_{i1}, \cdots, \sum_{i=1}^{n} x_{i0} x_{id} \\ \sum_{i=1}^{n} x_{i1} x_{i0}, \sum_{i=1}^{n} x_{i1} x_{i1}, \cdots, \sum_{i=1}^{n} x_{i1} x_{id} \\ \cdots \\ \sum_{i=1}^{n} x_{id} x_{i0}, \sum_{i=1}^{n} x_{id} x_{i1}, \cdots, \sum_{i=1}^{n} x_{id} x_{id} \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \cdots \\ w_d \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \cdots \\ \sum_{i=1}^{n} x_{id} y_i \end{pmatrix}$$

Solution

- **Quadratic Function Optimization**
  - Using given data, let's define **X** and **Y**

$D_1 = (1, x_{11}, x_{12}, \dots, x_{1d}, y_1)$

$D_2 = (1, x_{21}, x_{22}, \dots, x_{2d}, y_2)$

$\dots$

$D_3 = (1, x_{n1}, x_{n2}, \dots, x_{nd}, y_n)$

$$\mathbf{X} = \begin{pmatrix} x_{10}, x_{11}, x_{12}, \cdots, x_{1d} \\ x_{20}, x_{21}, x_{22}, \cdots, x_{2d} \\ \cdots \\ x_{n0}, x_{n1}, x_{n2}, \cdots, x_{nd} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

  - Then,

$$\mathbf{A} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0}x_{i0}, \sum_{i=1}^{n} x_{i0}x_{i1}, \cdots, \sum_{i=1}^{n} x_{i0}x_{id} \\ \sum_{i=1}^{n} x_{i1}x_{i0}, \sum_{i=1}^{n} x_{i1}x_{i1}, \cdots, \sum_{i=1}^{n} x_{i1}x_{id} \\ \cdots \\ \sum_{i=1}^{n} x_{id}x_{i0}, \sum_{i=1}^{n} x_{id}x_{i1}, \cdots, \sum_{i=1}^{n} x_{id}x_{id} \end{pmatrix} = \mathbf{X}^T\mathbf{X} \qquad \mathbf{b} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0}y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \cdots \\ \sum_{i=1}^{n} x_{id}y_i \end{pmatrix} = \mathbf{X}^T\mathbf{Y}$$

$y = w_0 + w_1 x$

Solution

$Aw = b$

- **Quadratic Function Optimization**

$X^T X w = X^T Y$

$\boxed{w = (X^T X)^{-1} X^T Y}$

가장 최적, 정확한 솔루션

$$\mathbf{w} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}\right)$$

$$f(\mathbf{x}_i) = w_0 x_{i0} + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} = \sum_{j=0}^{d} w_j x_{ij}$$

Training samples

$(1, x_{11}, x_{12}, \ldots, x_{1d}, y_1)$
$(1, x_{21}, x_{22}, \ldots, x_{2d}, y_2)$
$\ldots$
$(1, x_{n1}, x_{n2}, \ldots, x_{nd}, y_n)$

Cut

$$\mathbf{X} = \begin{pmatrix} 1, x_{11}, x_{12}, \cdots, x_{1d} \\ 1, x_{21}, x_{22}, \cdots, x_{2d} \\ \cdots \\ 1, x_{n1}, x_{n2}, \cdots, x_{nd} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

Solution

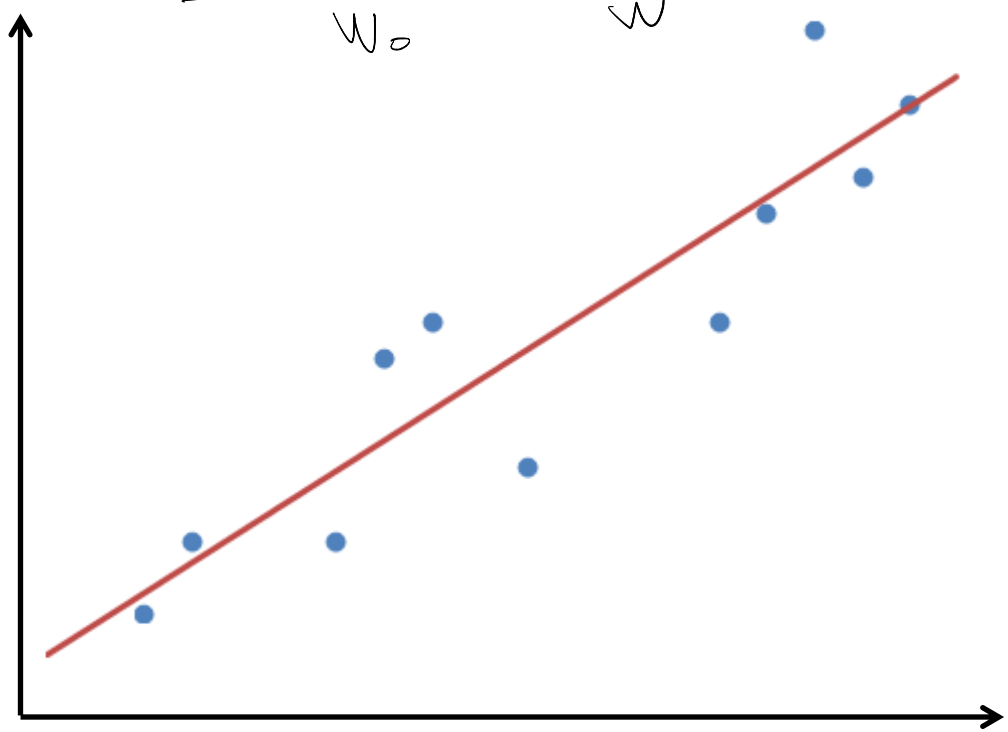
$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = (X^TX)^{-1}X^TY$$

- **Example**

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

matrix x y

$Y = \underbrace{0.840708}_{w_0} + \underbrace{0.834071}_{w}x$



Solution

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 & \cdots \\ 2 & 3 & 6 & 7 & 8 & \end{matrix}$$

- **Example**

x_0

rewrite

$2 \times 10 \times 10 \times 2$

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

$$\mathbf{X} = \begin{pmatrix} 1,2 \\ 1,3 \\ 1,6 \\ 1,7 \\ 1,8 \\ 1,10 \\ 1,14 \\ 1,15 \\ 1,16 \\ 1,17 \\ 1,18 \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} 2 \\ 4 \\ 4 \\ 9 \\ 10 \\ 6 \\ 10 \\ 13 \\ 18 \\ 14 \\ 16 \end{pmatrix}$$

$$\mathbf{A} = \mathbf{X}^T\mathbf{X} = \begin{pmatrix} 11, 116 \\ 116, 1552 \end{pmatrix}$$

$$\mathbf{b} = \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} 106 \\ 1392 \end{pmatrix}$$

$$\mathbf{w} = (A)^{-1}(b) = \begin{pmatrix} 0.840708 \\ 0.834071 \end{pmatrix}$$

Additive Linear Model

- **Generalized Version of Linear Regression**
  - Linear Regression

$y = w_0 + w_1 x$

$$f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d = w_0 + \sum_{j=1}^{d} w_j x_j$$

둘다 w에 대해 linear함

    - Find *w*s so that *f(x)* best fits given data

instead

$y = w_0 + w_1 \sin \pi x + w_2 e^x$

  - Generalized Linear Regression
    - Instead of single variables, let's use pre-determined functions of **x**, *h(x)*s

$$f(\mathbf{x}) = w_0 + w_1 h_1(\mathbf{x}) + w_2 h_2(\mathbf{x}) + \cdots + w_d h_d(\mathbf{x}) = w_0 + \sum_{j=1}^{d} w_j h_j(\mathbf{x})$$

    - Find *w*s so that *f(x)* best fits given data

For example, $f(x) = w_0 + w_1 \sin(\pi x) + w_2 e^x$

Additive Linear Model

ex

x가 성분이 2개이상!

x_1	x_2	y
1	1	1
1	2	2
2	3	3
2	4	4
3	2	5

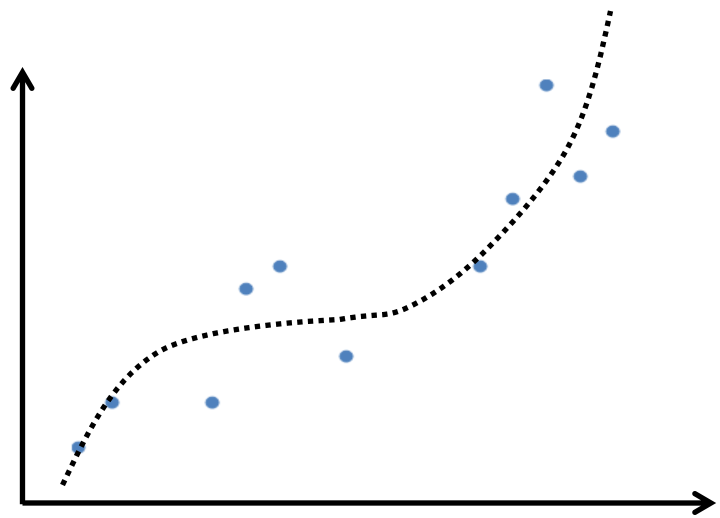
2nd order

$$f(x_1, x_2) = w_0 + w_1 \underbrace{x_1}_{h_1(x_1,x_2)} + w_2 \underbrace{x_2}_{h_2(x_1,x_2)} + w_3 \underbrace{x_1^2}_{h_3(x_1,x_2)} + w_4 \underbrace{x_2^2}_{h_4(x_1,x_2)} + w_5 \underbrace{x_1 x_2}_{h_5(x_1,x_2)}$$

이걸 계산해주면됨

- **Example**
  - Find the 3rd order Polynomial which best fits the data

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)



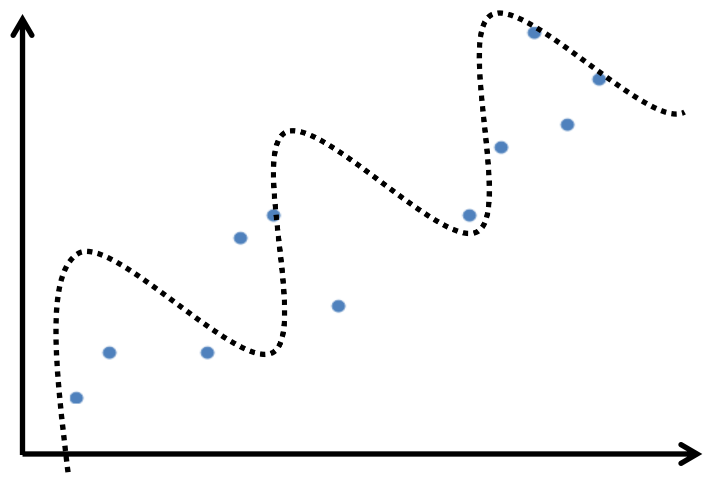
Use this model $f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)$

where, $h_1(x) = x$, $h_2(x) = x^2$, $h_3(x) = x^3$

Additive Linear Model

- **Example**
  - You think that the given data are periodic

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)



You may use $f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x)$

where, $h_1(x) = x,\ h_2(x) = \sin(ax)$

Additive Linear Model

- **How to solve this???**

Find $\mathbf{w} = (w_0, w_1, \cdots, w_d)$ which minimizes $E(\mathbf{w})$ where

$$E(\mathbf{w}) = \sum_{i=1}^{n} \left(f(\mathbf{x}_i) - y_i \right)^2$$

$$f(\mathbf{x}_i) = w_0 h_0(\mathbf{x}_i) + w_1 h_1(\mathbf{x}_i) + \cdots + w_d h_d(\mathbf{x}_i) \quad \text{where } \underbrace{h_0(\mathbf{x}_i) = 1}_{\text{가중치를 설정해줄 필요X}}$$

$$D = \{D_1, D_2, \cdots, D_n\}$$

$$D_i = (\mathbf{x}_i, y_i)$$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})$$

Additive Linear Model

- **How to solve this**
  - Solution

$$\mathbf{w} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\left(\mathbf{H}^T\mathbf{Y}\right)$$

$$\approx \left(X^T X\right)^{-1}\left(X^T Y\right)$$

  - Where

$$\mathbf{H} = \begin{pmatrix} h_0(\mathbf{x}_1), h_1(\mathbf{x}_1), \cdots, h_k(\mathbf{x}_1) \\ h_0(\mathbf{x}_2), h_1(\mathbf{x}_2), \cdots, h_k(\mathbf{x}_2) \\ \cdots \\ h_0(\mathbf{x}_n), h_1(\mathbf{x}_n), \cdots, h_k(\mathbf{x}_n) \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

Additive Linear Model

- **Example**
  - Find the 3rd order Polynomial which best fits the data

x_i , y

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

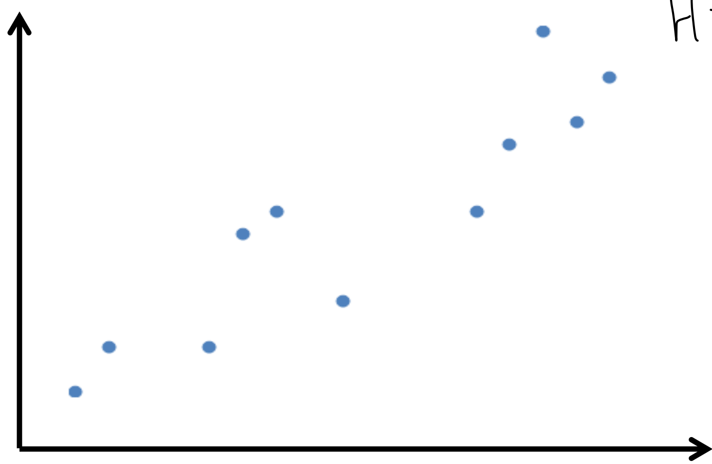
$$f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)$$

$$\text{where, } h_1(x) = x,\ h_2(x) = x^2,\ h_3(x) = x^3$$

$$y = \begin{pmatrix} 2 \\ 4 \\ \vdots \\ 14 \\ 16 \end{pmatrix}$$

$$\begin{array}{cccccc} & h_0(x) & h_1(x) & h_2(x) & h_3(x) \\ H = & \left(\begin{array}{cccc} h_0(2) & h_1(2) & h_2(2) & h_3(2) \\ h_0(3) & h_1(3) & h_2(3) & h_3(3) \\ & \vdots & & \\ h_0(18) & h_1(18) & \cdots & \end{array}\right) \end{array}$$

$$= \begin{pmatrix} 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 6 & 36 & 216 \\ & \vdots & & \end{pmatrix}$$

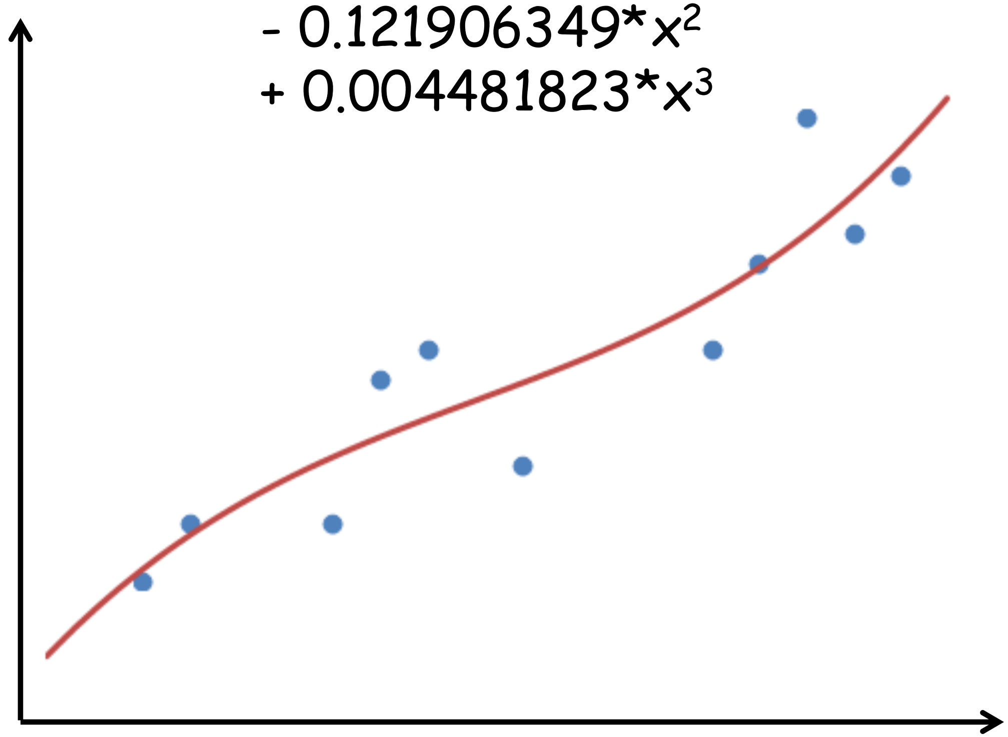


Additive Linear Model

- **Example**

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

$$f(x) = -0.63143704 + 1.713506327*x - 0.121906349*x^2 + 0.004481823*x^3$$

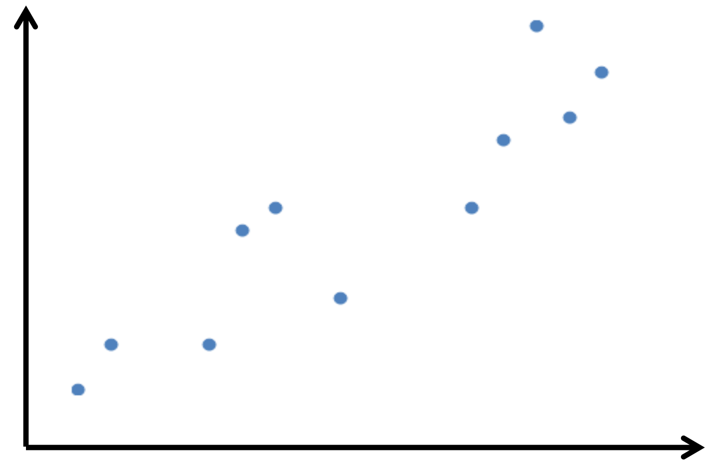


Additive Linear Model

- **Example**
  - You think that the given data are periodic

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

$$f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x)$$
$$\text{where, } h_1(x) = x,\ h_2(x) = \sin(ax)$$

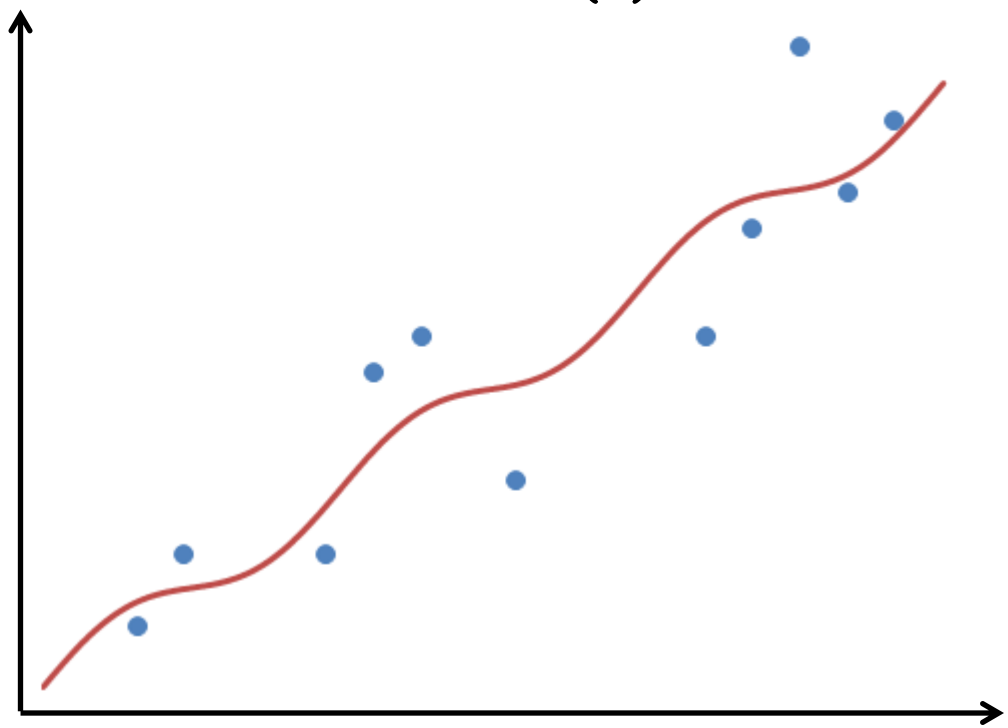


Additive Linear Model

- **Example**

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

f(x) =0.328770262
+0.873739195*x
+0.680191174*sin(x)



Additive Linear Model

$\rightarrow \exp\left(\frac{(x-a)^2}{b}\right)$ 이런 형태를 kernel 이라 함

- **Example: Kernel Regression** Good choice
  - You can find a linear combination of given kernel functions

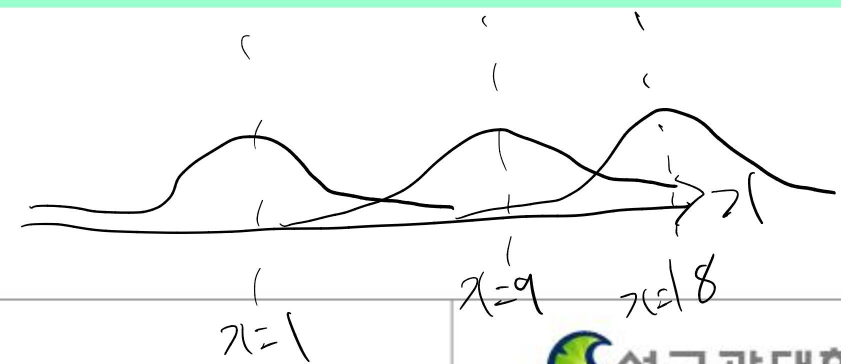
(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)

$$f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)$$

$$\text{where, } h_1(x) = \exp\left(\frac{-(x-1)^2}{18}\right), h_2(x) = \exp\left(\frac{-(x-9)^2}{18}\right),$$

$= e^x$

$$h_3(x) = \exp\left(\frac{-(x-18)^2}{18}\right)$$



Additive Linear Model

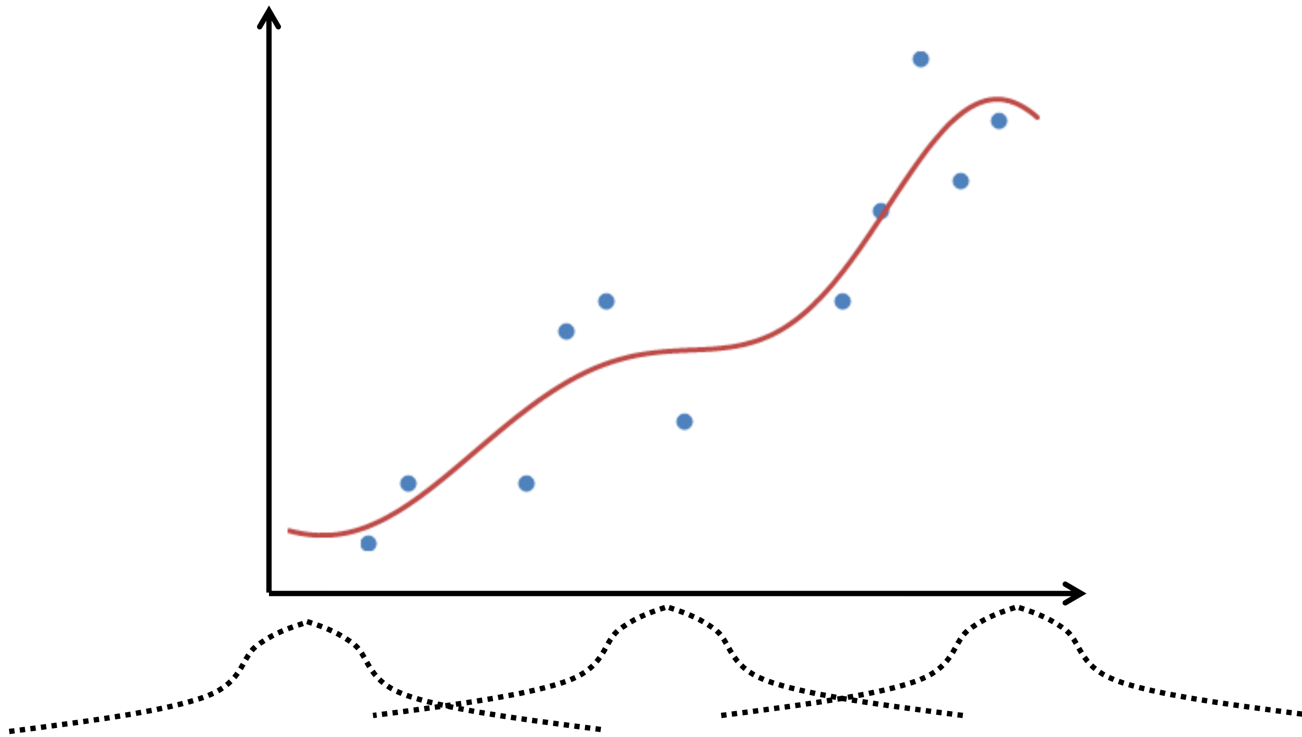
hyper parameter ① # of kernel ② center point

- **Example**

→ kernel을 사용하면 2개의 파라미터로 모든 모양의 커브 표현 가능

$$f(x) = 5.76 - 3.64\exp\left(\frac{-(x-1)^2}{18}\right) + 2.40\exp\left(\frac{-(x-9)^2}{18}\right) + 10.82\exp\left(\frac{-(x-18)^2}{18}\right)$$

(2, 2)
(3, 4)
(6, 4)
(7, 9)
(8, 10)
(10, 6)
(14, 10)
(15, 13)
(16, 18)
(17, 14)
(18, 16)



Another Approaches

- **Solving Linear Equations** ← Exact solution
  - Simplest Approach for Linear Regression

- **Another Approaches to Solve Linear Regression** 중요하다 (DNN과 관계있다)
  - Gradient Descent Approach
  - Maximum Likelihood Estimation Approach

  => Both give the same solution as SLE gives

- **Then, Why?**
  - Both are well-known, useful and important methods to solve problems in ML domain