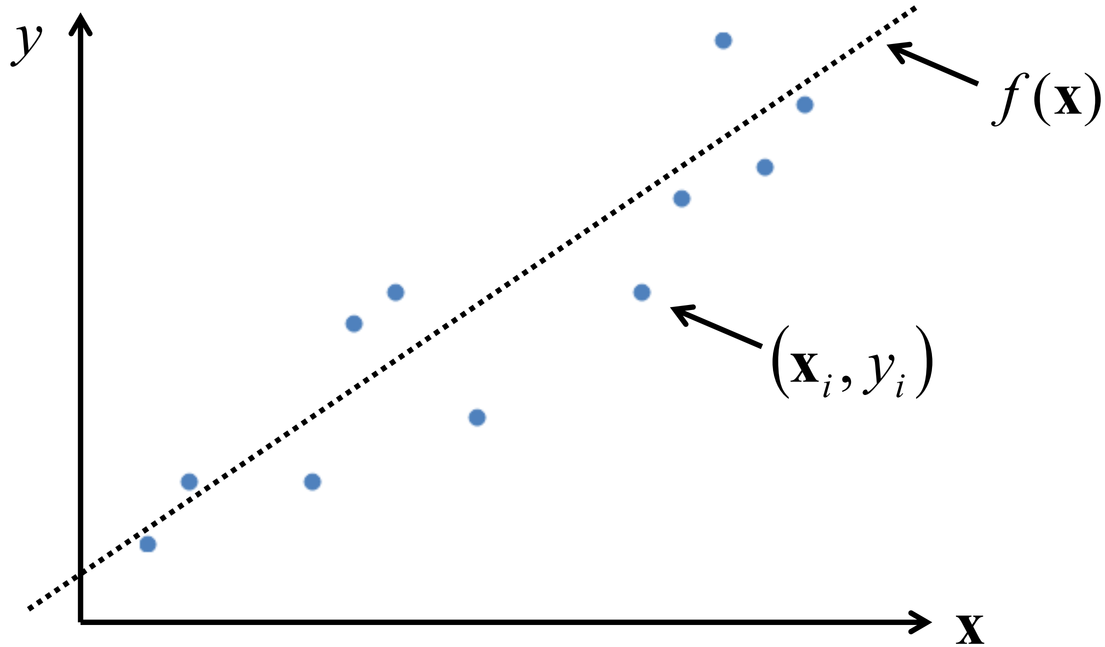
Linear Regression &
Additive Linear Model

Introduction

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

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



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

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

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

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

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

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

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

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

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

Introduction

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

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

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

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

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

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

Introduction

- **How to solve this?**

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

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

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

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

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

x and y are given values
E is a quadratic function of w's

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

Solution

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

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

  - That is

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

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

Solution

- **Quadratic Function Optimization**

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

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

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

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

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

Solution

- **Quadratic Function Optimization**

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

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

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

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

$$\Rightarrow$$

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

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

$$\cdots$$

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

Solution

- **Quadratic Function Optimization**

$$\mathbf{A}\mathbf{w} = \mathbf{b}$$

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

Solution

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

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

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

  - Then,

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

Solution

- **Quadratic Function Optimization**

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

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

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

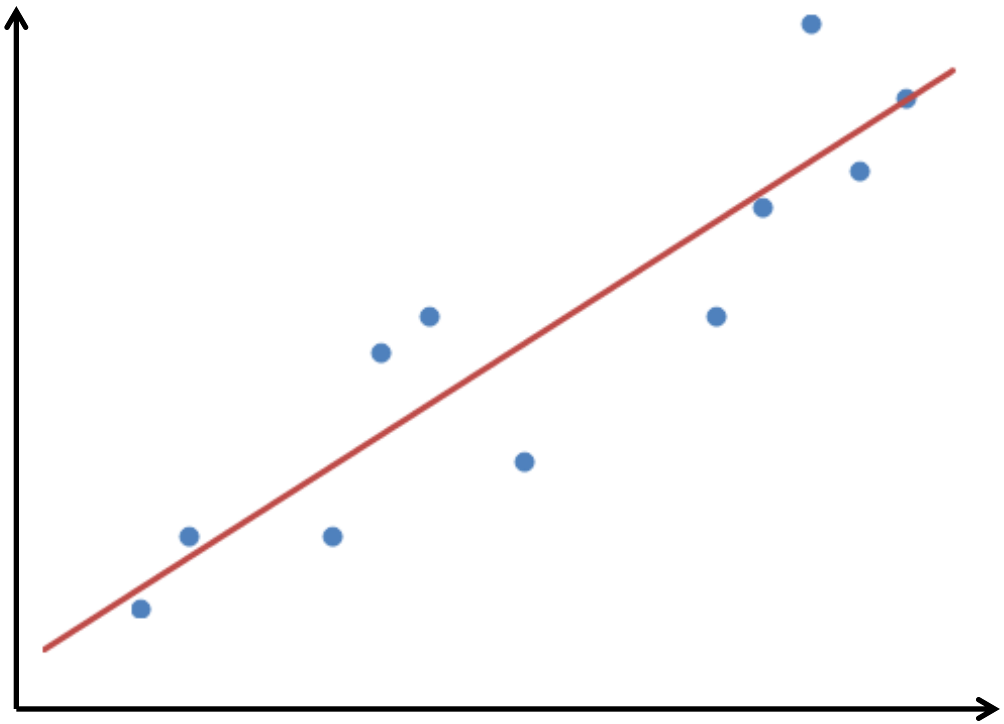
$$\mathbf{X} = \begin{pmatrix} 1, x_{11}, x_{12}, \cdots, x_{1d} \\ 1, x_{21}, x_{22}, \cdots, x_{2d} \\ \cdots \\ 1, x_{n1}, x_{n2}, \cdots, x_{nd} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

Solution

- **Example**

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

Y=0.840708+0.834071x



Solution

- **Example**

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

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

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

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

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

Additive Linear Model

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

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

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

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

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

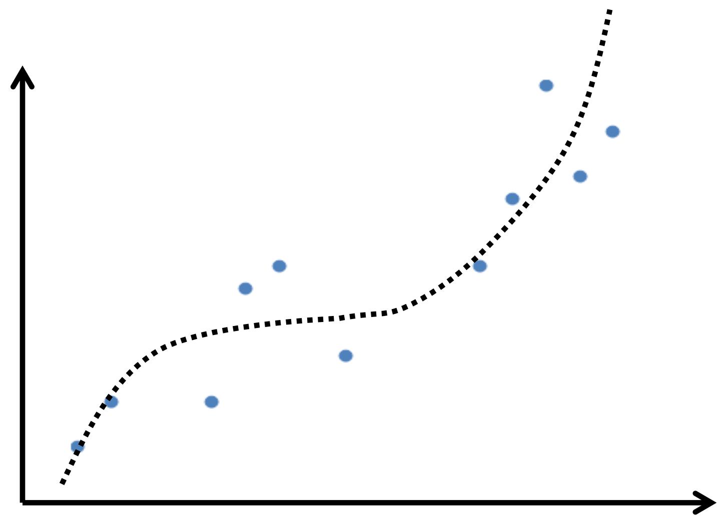
    - Find *w*'s so that *f(x)* best fits given data

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

Additive Linear Model

- **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)



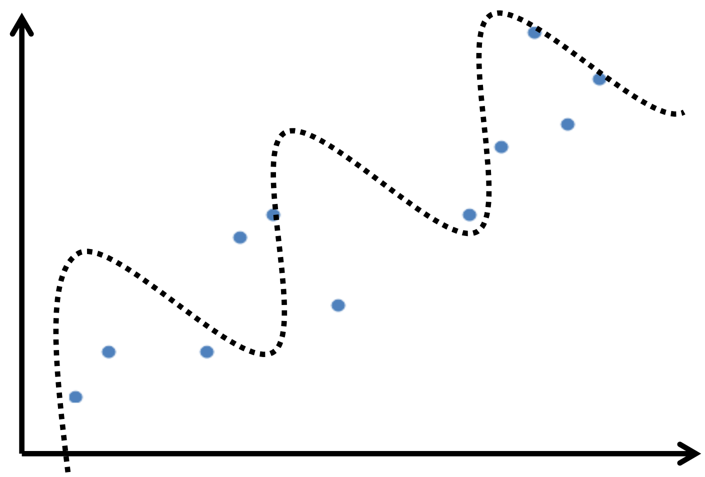
Use this model $f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)$

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 $\mathbf{w} = \left(w_0, w_1, \cdots, w_k\right)$ which minimizes $E\left(\mathbf{w}\right)$ where

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

$$f\left(\mathbf{x}_i\right) = w_0 h_0\left(\mathbf{x}_i\right) + w_1 h_1\left(\mathbf{x}_i\right) + \cdots + w_k h_k\left(\mathbf{x}_i\right) \quad \text{where } h_0\left(\mathbf{x}_i\right) = 1$$

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

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

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

Additive Linear Model

- **How to solve this**
  - Solution

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

  - Where

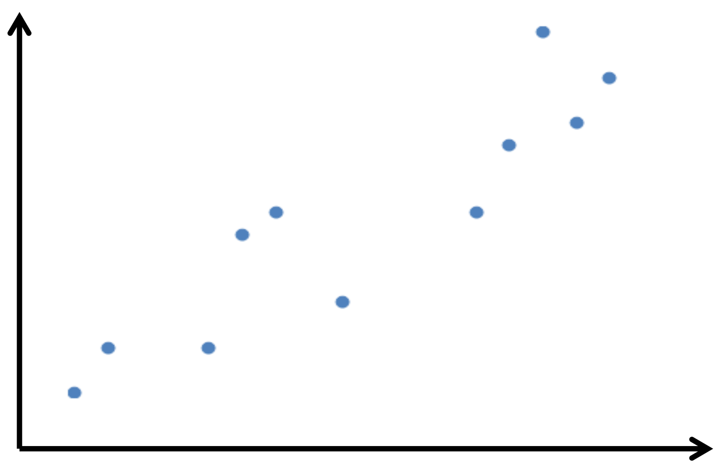
$$\mathbf{H} = \begin{pmatrix} h_0(\mathbf{x}_1), h_1(\mathbf{x}_1), \cdots, h_k(\mathbf{x}_1) \\ h_0(\mathbf{x}_2), h_1(\mathbf{x}_2), \cdots, h_k(\mathbf{x}_2) \\ \cdots \\ h_0(\mathbf{x}_n), h_1(\mathbf{x}_n), \cdots, h_k(\mathbf{x}_n) \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\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

Additive Linear Model

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

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

$$f(x) = w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)$$
$$\text{where, } h_1(x) = x,\ h_2(x) = x^2,\ \ h_3(x) = x^3$$

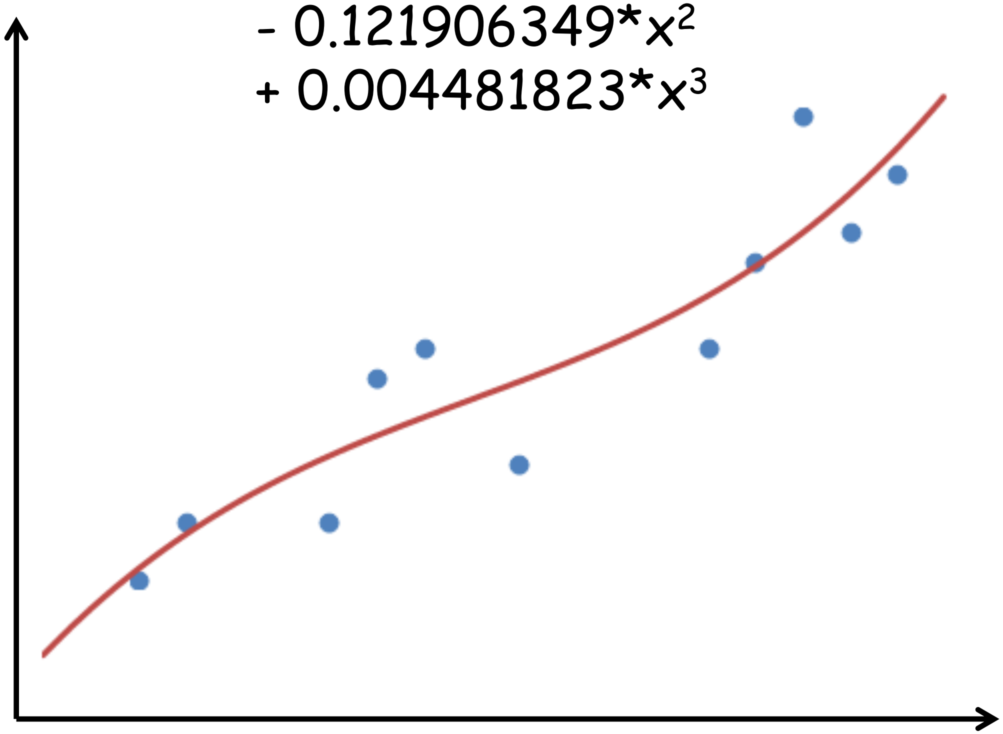


Additive Linear Model

- **Example**

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

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

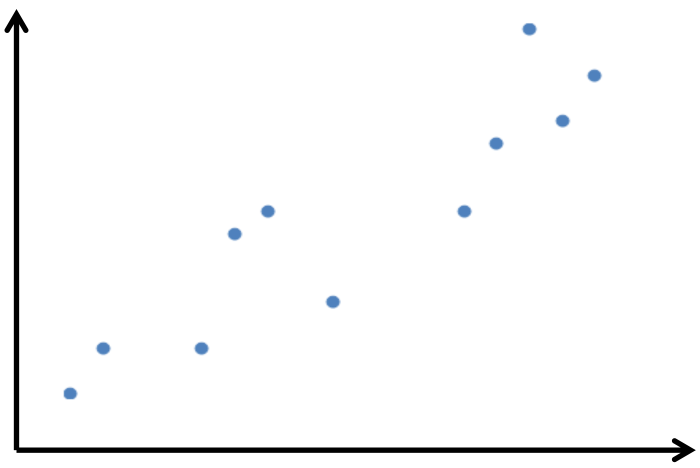


Additive Linear Model

- **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)

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

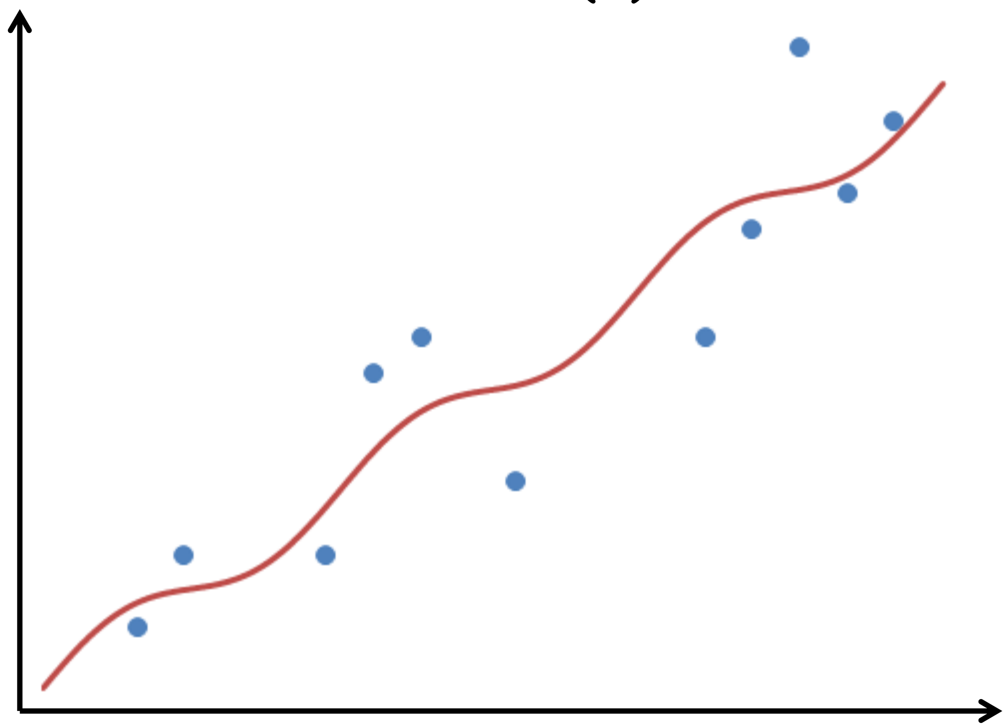


Additive Linear Model

- **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

- **Example: Kernel Regression**
  - 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)

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

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

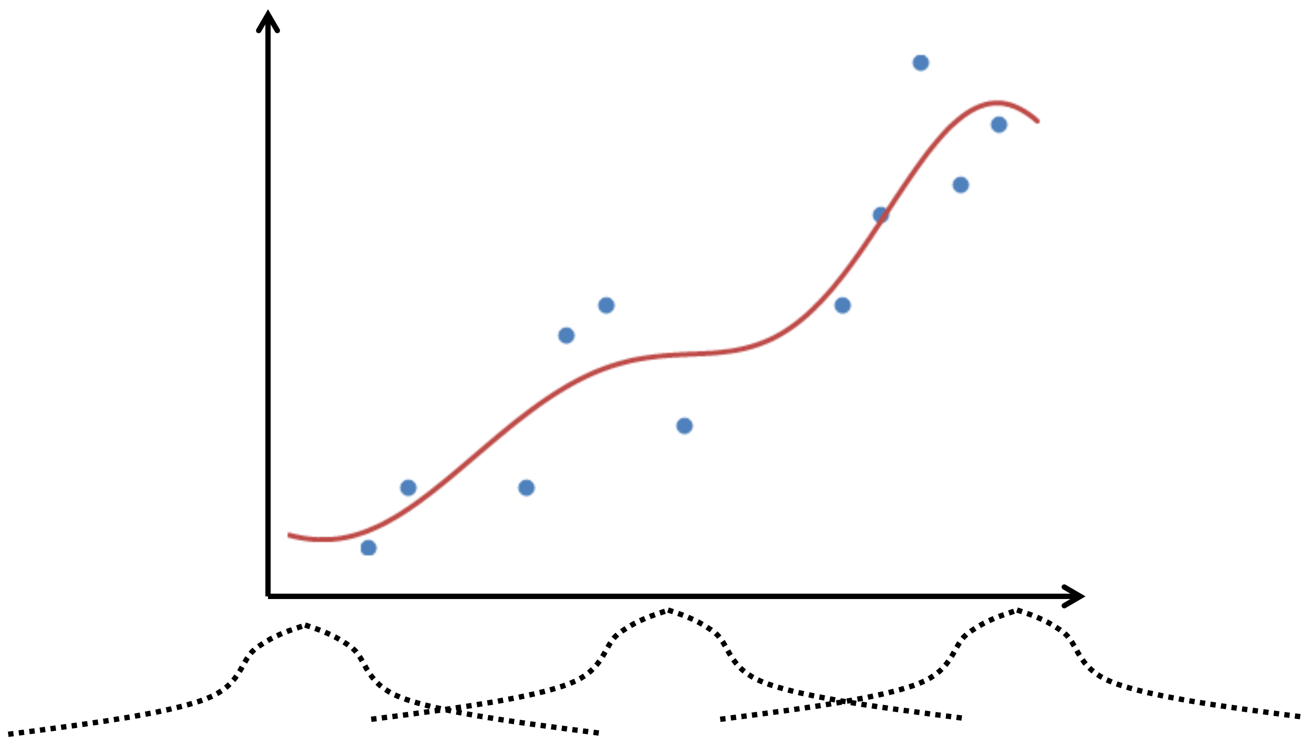
$$h_3(x) = \exp\left(\frac{-(x-18)^2}{18}\right)$$

Additive Linear Model

- **Example**

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

(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**
  - Simplest Approach for Linear Regression

- **Another Approaches to Solve Linear Regression**
  - 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