Linear & Polynomial Regression

Introduction

$e_i = |f(x_i) - y_i|$ $\quad$ $e_i = (f(x_i) - y_i)^2$

exact하지만, 계산 힘듦.

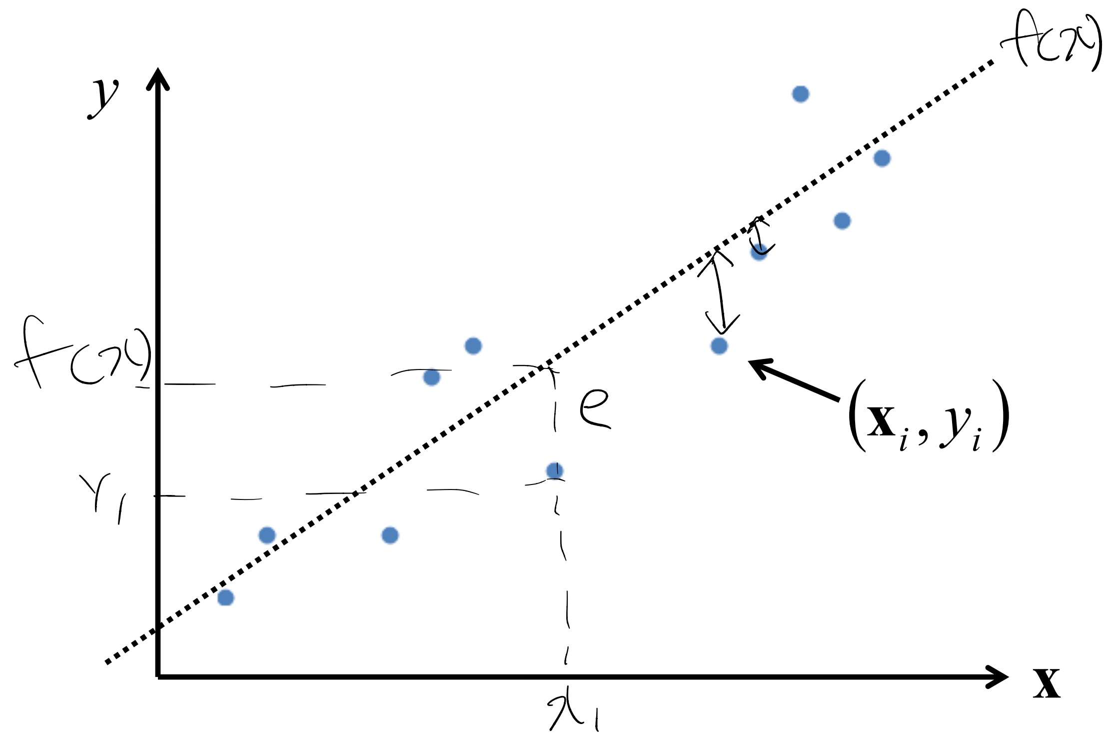
그래서 얘를 씀

$\boxed{\min \sum_{i=1}^{n} (f(x_i) - y_i)^2}$

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

vector

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

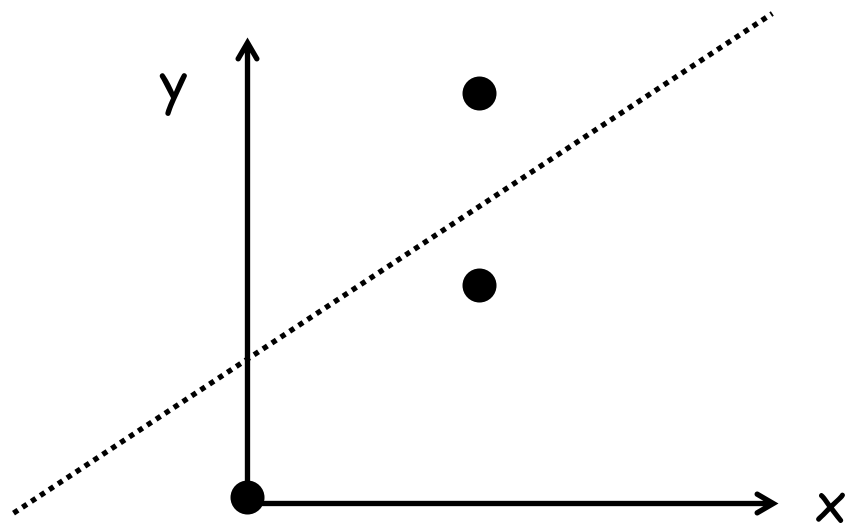


HOW?

Linear Regression

Simple Problem

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)\}$$



$$f(x) = w_1 x + w_0$$

Linear Regression

- **What is the "best-fit line" for the given samples?**

$$Data = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$$

$$f(x) = w_1 x + w_0$$

best-fit line의 조건

가능한 가장 많이

$f(x_1) = w_1 x_1 + w_0$ should be close to y_1 as much as possible

$f(x_2) = w_1 x_2 + w_0$ should be close to y_2 as much as possible

$f(x_3) = w_1 x_3 + w_0$ should be close to y_3 as much as possible

⇨ $|f(x_1) - y_1|$ is minimized ⇨ $(f(x_1) - y_1)^2$ is minimized

$|f(x_2) - y_2|$ is minimized $(f(x_2) - y_2)^2$ is minimized

$|f(x_3) - y_3|$ is minimized $(f(3) - y_3)^2$ is minimized

미니마이즈!

$$= \sum_{(x,y) \in Data} (f(x) - y)^2$$

Linear Regression

- **Find a line *f(x)* which minimizes *E*** = 선형회귀

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)\}$$

$$f(x) = w_1 x + w_0$$

$$E = \sum_{(\mathbf{x}, y) \in Data} \left(y - (w_1 x + w_0)\right)^2$$

- **But, how?**

Steps of Machine Learning

- **We have to find out *w_0* and *w_1* which can minimize *E***

$$E = \sum_{(\mathbf{x},y)\in Data} \left(y - (w_1 x + w_0)\right)^2$$

X,Y, W 에대한 함수

⇓

(x,y) 데이터를입력

⇓

W에대한함수

(W에대한 2차함수)

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)\}$$

x_1 y_1 x_2 y_2 x_3 y_3

$$E = \left(0.0 - f(0.0)\right)^2 + \left(1.0 - f(1.0)\right)^2 + \left(2.0 - f(1.0)\right)^2$$

$$E = (0.0 - w_0)^2 + \left(1.0 - (w_1 + w_0)\right)^2 + \left(2.0 - (w_1 + w_0)\right)^2$$

$$E = 2{w_1}^2 + 3{w_0}^2 - 6w_1 - 6w_0 + 4w_1 w_0 + 5$$

Steps of Machine Learning

- **We have to find out *w_0* and *w_1* which can minimize *E***

$$E = 2{w_1}^2 + 3{w_0}^2 - 6w_1 - 6w_0 + 4w_1 w_0 + 5$$

w1 에대한 함수최소
w0에 대한 함수최소) 둘다 최소일때

$$\frac{\partial E}{\partial w_1} = 4w_1 + 4w_0 - 6$$

$$\frac{\partial E}{\partial w_0} = 4w_1 + 6w_0 - 6$$

$$\boxed{\begin{aligned} 4w_1 + 4w_0 - 6 &= 0 \\ 4w_1 + 6w_0 - 6 &= 0 \end{aligned}}$$

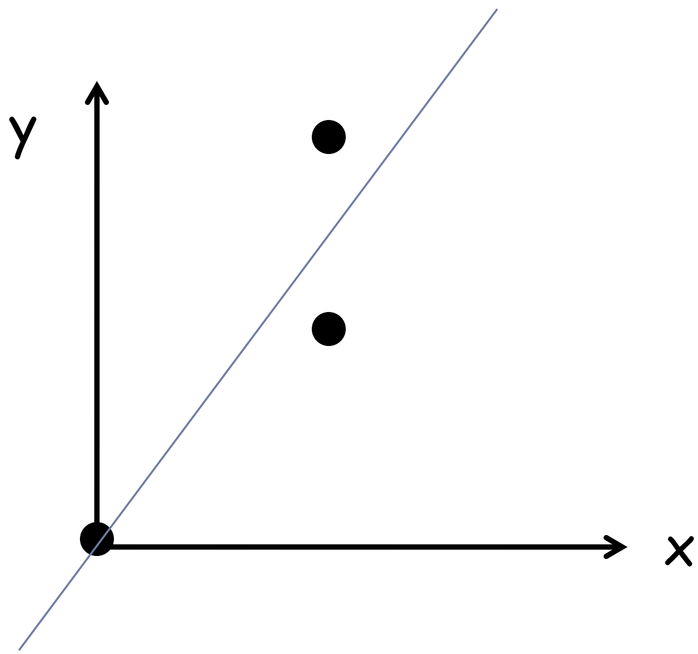
$$w_1 = 1.5$$

$$w_0 = 0.0$$

Steps of Machine Learning

- **The best-fit line is**

$$f(x)=1.5x + 0.0$$



Steps of Regression

- For given **Data** $= \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_\mathrm{n}, y_n)\}$

- Choose a model $f(\mathbf{x}; \mathbf{w})$

- Find w to minimize E

$$E(\mathbf{w}) = \sum_{(\mathbf{x},y) \in Data} \left(y - f(\mathbf{x}; \mathbf{w})\right)^2$$

Hmm, I have a question

- **Does f have to be a linear function of x?**

Find $w_1, w_2, \dots, w_m$ to minimize the followings:

$$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$

- For example, why not

$$f(x) = w_2 x^2 + w_1 x + w_0$$

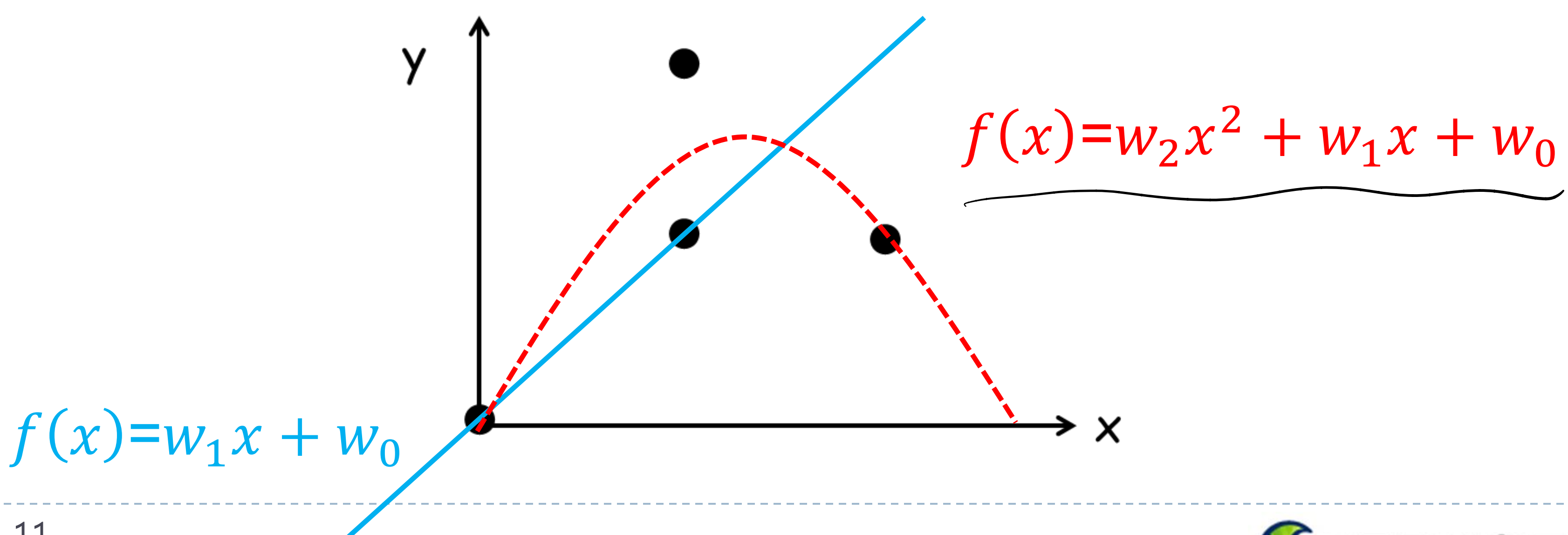
instead of

$$f(x) = w_1 x + w_0$$

Hmm, I have a question

- **Does f have to be a linear function of x?**

> Find $w_1, w_2, \dots, w_m$ to minimize the followings:
>
> $$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$



Hmm, I have a question

x, y, w에 대한 함수

- **Does f have to be a linear function of x?**

Find $w_1, w_2, \ldots, w_m$ to minimize the followings:

$$E(w_1, w_2, \ldots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \ldots, w_m)\right)^2$$

Quadratic Fun

x, y는 상수 => w만 변수 = (linear)^2

이차함수 of w

Quadratic function of x $f(x)=w_2x^2 + w_1x + w_0$

w에 대한 선형함수

Linear function of w's

Linear function of x $f(x)=w_1x + w_0$

E is not a function of x, but of w's -> E is a quadratic function of w's

= Best Fit 적용가능

Hmm, I have a question

- **Does f have to be a linear function of x?"**

Find $w_1, w_2, \dots, w_m$ to minimize the followings:

$$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$

- E is a quadratic function of w's. Let's apply the same method

$$\left.\begin{aligned} \frac{\partial E}{\partial w_1} &= 0 \\ \frac{\partial E}{\partial w_2} &= 0 \\ &\cdots \\ \frac{\partial E}{\partial w_m} &= 0 \end{aligned}\right\}$$

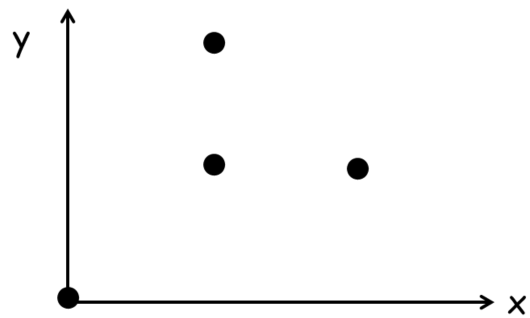
A system of linear equations

Yes!!
We can solve it.

Let's do it

- **Find the best-fit quadratic function of x**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0), (2.0, 1.0)\}$$



$$f(x; w_0, w_1, w_2) = w_2 x^2 + w_1 x + w_0$$

  - Determine w_0, w_1, w_2

Let's do it

- **For given data and function**

$$Data = \{(\overset{x_1}{0.0}, \overset{y_1}{0.0}), (\overset{x_2}{1.0}, \overset{y_2}{1.0}), (\overset{x_3}{1.0}, \overset{y_3}{2.0})(\overset{x_4}{2.0}, \overset{y_4}{1.0})\}$$

$$f(x; w_0, w_1, w_2) = w_2 x^2 + w_1 x + w_0$$

determine w_0, w_1, w_2 that minimize

$$E(w_0, w_1, w_2) = \sum_{(\mathbf{x},y)\in Data} \big(y - f(\mathbf{x}; w_0, w_1, w_2)\big)^2$$

Let's do it

- **For given data and function**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)(2.0, 1.0)\}$$

$$E(w_0, w_1, w_2) = \sum_{(\mathbf{x}, y) \in Data} \left(y - (w_2 x^2 + w_1 x + w_0)\right)^2$$

determine w_0, w_1, w_2 that minimize

$$\begin{aligned} E(w_0, w_1, w_2) = {} & (0 - w_0)^2 \\ & + \left(1 - (w_2 + w_1 + w_0)\right)^2 \\ & + \left(2 - (w_2 + w_1 + w_0)\right)^2 \\ & + \left(1 - (4\mathrm{w}_2 + 2\mathrm{w}_1 + \mathrm{w}_0)\right)^2 \end{aligned}$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**

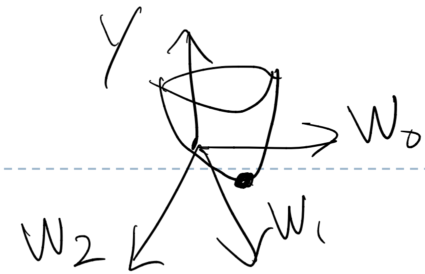
$$E(w_0, w_1, w_2) = (0 - w_0)^2 + \left(1 - (w_2 + w_1 + w_0)\right)^2$$
$$+\left(2 - (w_2 + w_1 + w_0)\right)^2 + \left(1 - (4\mathrm{w}_2 + 2\mathrm{w}_1 + \mathrm{w}_0)\right)^2$$

$$\begin{aligned} E(w_0, w_1, w_2) = {} & w_0^2 \\ & + w_0^2 + w_1^2 + w_2^2 + 2w_0w_1 + 2w_0w_2 + 2w_1w_2 - 2w_0 - 2w_1 - 2w_2 + 1 \\ & + w_0^2 + w_1^2 + w_2^2 + 2w_0w_1 + 2w_0w_2 + 2w_1w_2 - 4w_0 - 4w_1 - 4w_2 + 4 \\ & + w_0^2 + 4w_1^2 + 16w_2^2 + 4w_0w_1 + 8w_0w_2 + 16w_1w_2 - 2w_0 - 4w_1 - 8w_2 + 1 \end{aligned}$$

$$\begin{aligned} E(w_0, w_1, w_2) = {} & 4w_0^2 + 6w_1^2 + 18w_2^2 + 8w_0w_1 \\ & + 12w_0w_2 + 20w_1w_2 - 8w_0 - 10w_1 - 14w_2 + 6 \end{aligned}$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**



$$E(w_0, w_1, w_2) = 4w_0^2 + 6w_1^2 + 18w_2^2 + 8w_0w_1 + 12w_0w_2 + 20w_1w_2 - 8w_0 - 10w_1 - 14w_2 + 6$$

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 8w_0 + 8w_1 + 12w_2 - 8$$

$$\frac{\partial}{\partial w_1} E(w_0, w_1, w_2) = 8w_0 + 12w_1 + 20w_2 - 10$$

$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 12w_0 + 20w_1 + 36w_2 - 14$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 8w_0 + 8w_1 + 12w_2 - 8 = 0$$

$$\frac{\partial}{\partial w_1} E(w_0, w_1, w_2) = 8w_0 + 12w_1 + 20w_2 - 10 = 0$$

$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 12w_0 + 20w_1 + 36w_2 - 14 = 0$$

$$w_0 = 0, w_1 = \frac{5}{2}, w_2 = -1$$

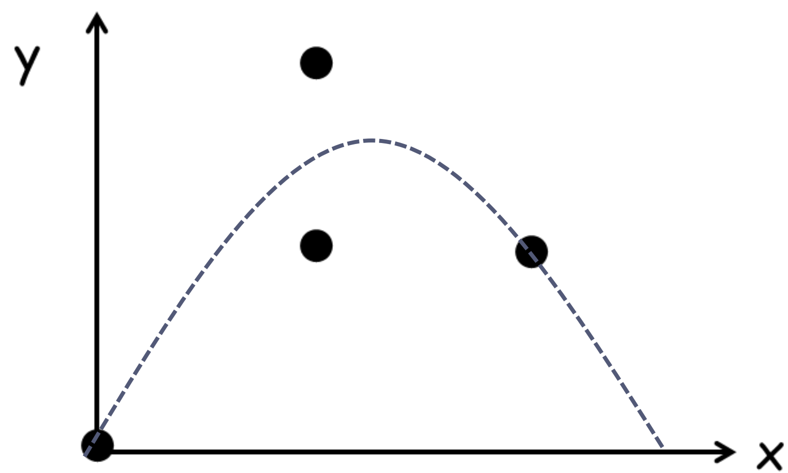
Let's do it

DNN

- **Find the best-fit quadratic function of x**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0), (2.0, 1.0)\}$$

Linear regression



$$f(x) = -x^2 + \frac{5}{2}x$$

I have another question

- Does f have to be a polynomial of x?

> Find $w_1, w_2, \dots, w_m$ to minimize the followings:
>
> $$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x}, y) \in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$

복잡한 함수여도 x,y가 상수기에 구할 수 있다!

- For example, why not

$$f(x) = w_2 2^x + w_1 \sin\frac{\pi}{2}x + w_0$$

instead of

$$f(x) = w_2 x^2 + w_1 x + w_0 \qquad \boxed{f(x) = w_1 x + w_0}$$

구할 수 있는데 (curve) 복잡한 함수는 사용하지 X

데이터에 가장 좋고 자주쓰는 curve는 찾아!

안에 들어갈 함수 = 내가 정하는 것

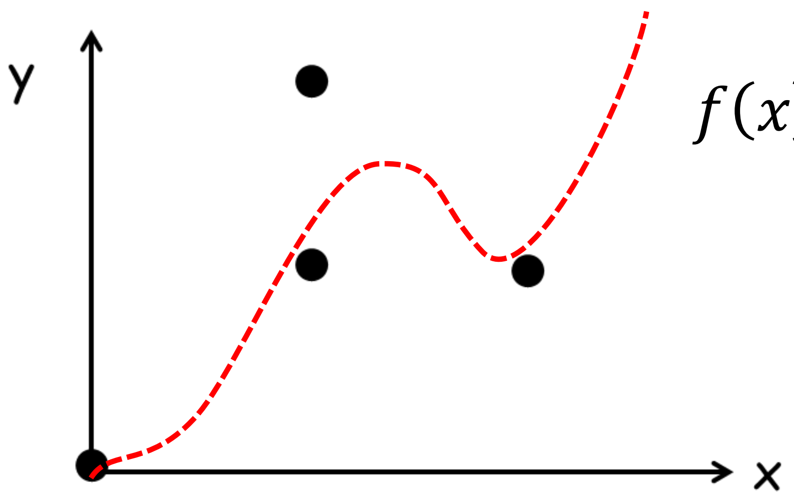
→ W(계수)만 best fit으로 구하면 된다

I have another question

- **Does f have to be a polynomial of x?**

Find $w_1, w_2, \dots, w_m$ to minimize the followings:

$$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$



$$f(x) = w_2 2^x + w_1 \sin\frac{\pi}{2}x + w_0$$

I have another question

- **Does f have to be a linear function of x?**

Find $w_1, w_2, \ldots, w_m$ to minimize the followings:

$$E(w_1, w_2, \ldots, w_m) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_2, \ldots, w_m)\right)^2$$

linear regression

Non-polynomial of x $\quad f(x) = w_2 2^x + w_1 \sin\frac{\pi}{2}x + w_0$

Quadratic function of x $\quad f(x) = w_2 x^2 + w_1 x + w_0$

Linear function of x $\quad f(x) = w_1 x + w_0$

Linear function of w's

선형회귀는 이것을 의미

E is not a function of x, but of w's -> E is a quadratic function of w's

I have another question

- **Does f have to be a linear function of x?"**

Find $w_1, w_2, \dots, w_m$ to minimize the followings:

$$E(w_1, w_2, \dots, w_m) = \sum_{(\mathbf{x},y) \in Data} \left(y - f(\mathbf{x}; w_1, w_2, \dots, w_m)\right)^2$$

- **E is a quadratic function of w's. Let's apply the same method**

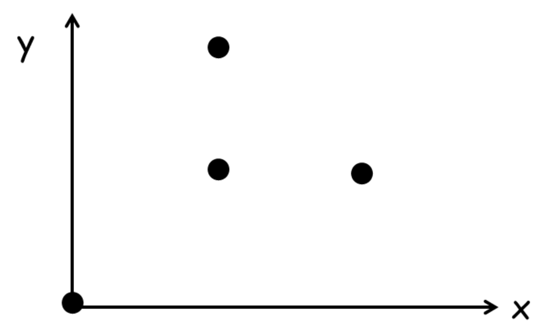
$$\left.\begin{array}{l} \dfrac{\partial E}{\partial w_1} = 0 \\ \dfrac{\partial E}{\partial w_2} = 0 \\ \cdots \\ \dfrac{\partial E}{\partial w_m} = 0 \end{array}\right\} \text{A system of linear equations}$$

Yes!!
We can solve it.

Let's do it

- **Find the best-fit quadratic function of x**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0), (2.0, 1.0)\}$$



$$f(x; w_0, w_1, w_2) = w_2 2^x + w_1 \sin\frac{\pi}{2} x + w_0$$

  - Determine w_0, w_1, w_2

Let's do it

- **For given data and function**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)(2.0, 1.0)\}$$

$$f(x; w_0, w_1, w_2) = w_2 2^x + w_1 \sin\frac{\pi}{2} x + w_0$$

determine w_0, w_1, w_2 that minimize

$$E(w_0, w_1, w_2) = \sum_{(\mathbf{x}, y) \in Data} \left(y - f(\mathbf{x}; w_0, w_1, w_2)\right)^2$$

Let's do it

- **For given data and function**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)(2.0, 1.0)\}$$

$$E(w_0, w_1, w_2) = \sum_{(\mathbf{x}, y) \in Data} \left(y - \left(w_2 2^x + w_1 \sin\frac{\pi}{2}x + w_0 \right) \right)^2$$

determine w_0, w_1, w_2 that minimize

$$\begin{aligned} E(w_0, w_1, w_2) = & \left(0 - (w_2 + w_0)\right)^2 \\ & + \left(1 - (2w_2 + w_1 + w_0)\right)^2 \\ & + \left(2 - (2w_2 + w_1 + w_0)\right)^2 \\ & + \left(1 - (4\mathrm{w}_2 + \mathrm{w}_0)\right)^2 \end{aligned}$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**

$$E(w_0, w_1, w_2) = \big(0 - (w_2 + w_0)\big)^2 + \big(1 - (2w_2 + w_1 + w_0)\big)^2$$
$$+\big(2 - (2w_2 + w_1 + w_0)\big)^2 + \big(1 - (4w_2 + w_0)\big)^2$$

$$\begin{aligned} E(w_0, w_1, w_2) = {} & w_0^2 + 2w_0w_2 + w_2^2 \\ & +w_0^2 + w_1^2 + 4w_2^2 + 2w_0w_1 + 4w_0w_2 + 4w_1w_2 - 2w_0 - 2w_1 - 4w_2 + 1 \\ & +w_0^2 + w_1^2 + 4w_2^2 + 2w_0w_1 + 4w_0w_2 + 4w_1w_2 - 4w_0 - 4w_1 - 8w_2 + 4 \\ & +w_0^2 + 16w_2^2 + 8w_0w_2 - 2w_0 - 8w_2 + 1 \end{aligned}$$

$$\begin{aligned} E(w_0, w_1, w_2) = {} & 4w_0^2 + 2w_1^2 + 25w_2^2 + 4w_0w_1 \\ & +18w_0w_2 + 8w_1w_2 - 8w_0 - 6w_1 - 20w_2 + 6 \end{aligned}$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**

$$E(w_0, w_1, w_2) = 4w_0^2 + 2w_1^2 + 25w_2^2 + 4w_0w_1 \\ +18w_0w_2 + 8w_1w_2 - 8w_0 - 6w_1 - 20w_2 + 6$$

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 8w_0 + 4w_1 + 18w_2 - 8$$

$$\frac{\partial}{\partial w_1} E(w_0, w_1, w_2) = 4w_0 + 4w_1 + 8w_2 - 6$$

$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 18w_0 + 8w_1 + 50w_2 - 20$$

Let's do it

- **Determine w_0, w_1, w_2 that minimize**

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 8w_0 + 4w_1 + 18w_2 - 8 = 0$$

$$\frac{\partial}{\partial w_1} E(w_0, w_1, w_2) = 4w_0 + 4w_1 + 8w_2 - 6 = 0$$

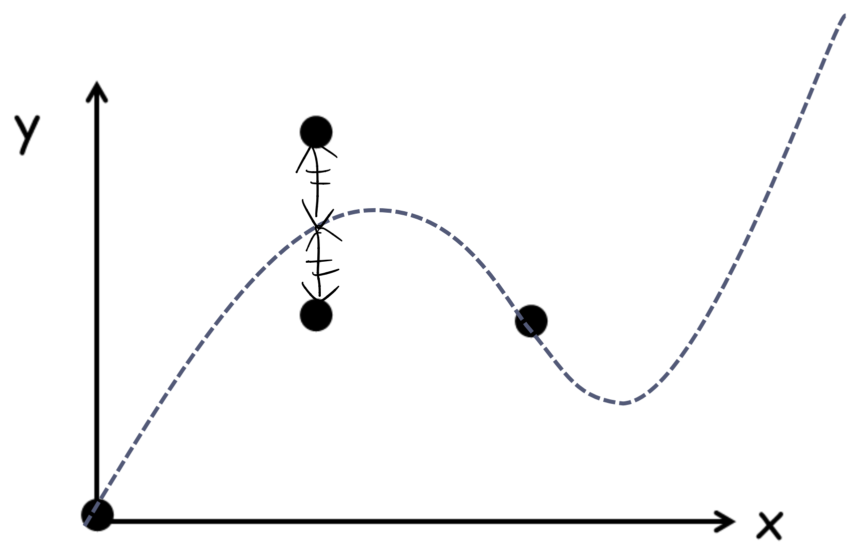
$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 18w_0 + 8w_1 + 50w_2 - 20 = 0$$

$$w_0 = -\frac{1}{3}, w_1 = \frac{7}{6}, w_2 = \frac{1}{3}$$

Let's do it

- **Find the best-fit quadratic function of x**

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0), (2.0, 1.0)\}$$



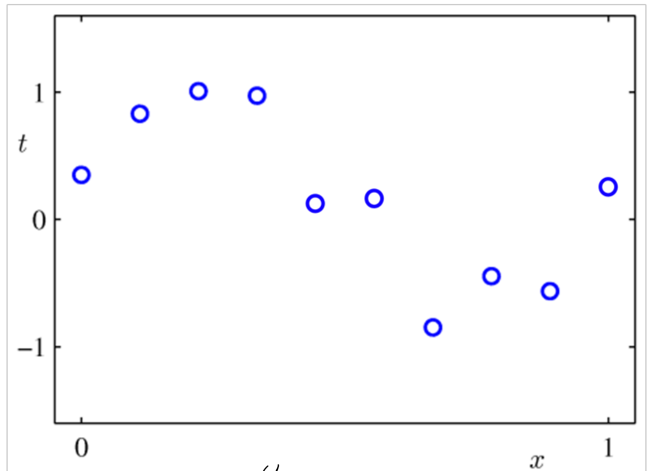
Best Fit

$$f(x) = \frac{1}{3}2^x + \frac{7}{6}\sin\frac{\pi}{2}x - \frac{1}{3}$$

Which model will be best?

- **What degree of polynomial?**

Sample



Error=0
→ Best ? (X)

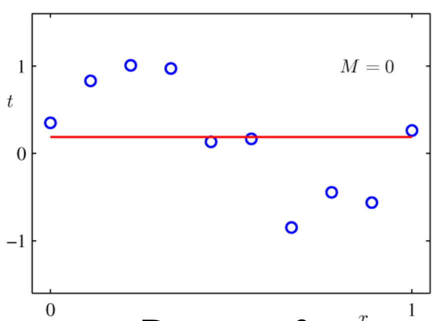
poly

0th

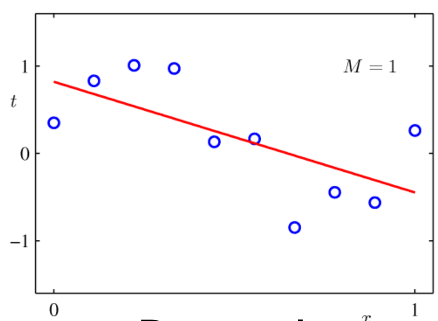
1st "w_1x+w_0"

"$w_3x^3+w_2x^2+w_1x+w_0$"

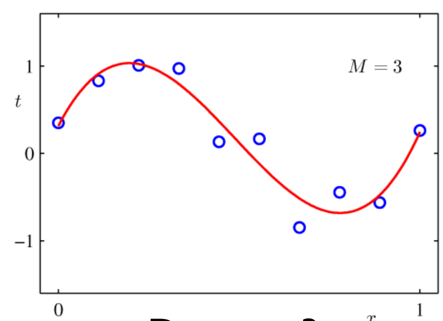
"$w_9x^9+\cdots+w_0$"



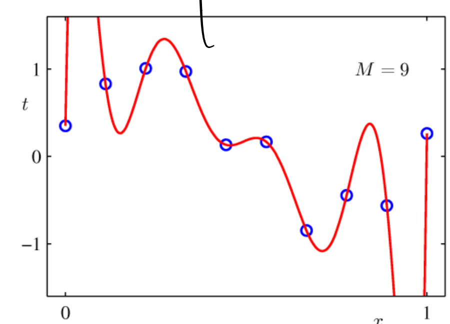
Degree=0



Degree=1



Degree=3



Degree=9

용어 잘기억하기 (overfitting)

Which model will be best?

- **Which degree of polynomial?**

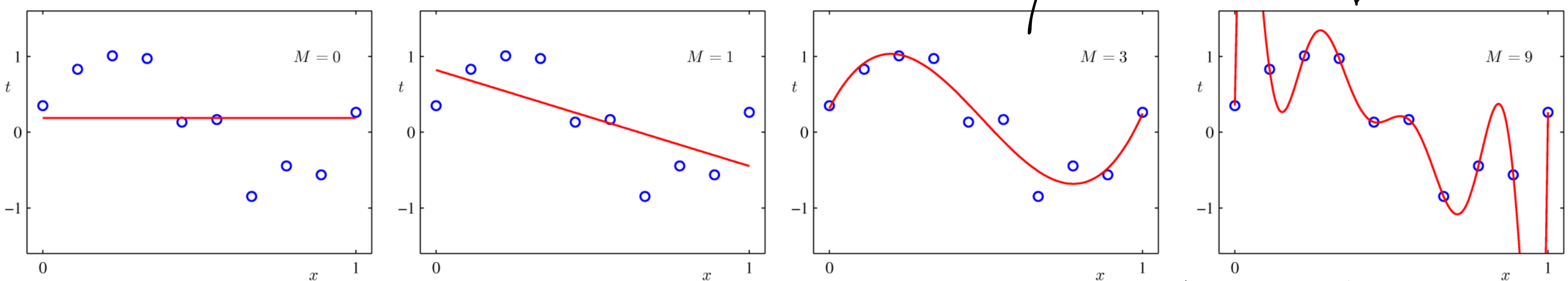
underfitted = stupid

이중에 Best Fit

Degree = hyperparameter

overfitted
과적합

Degree=0 Degree=1 Degree=3 Degree=9



" Complexity of Model " ← You have to choose

간단 복잡

Simple ↔ Complex

Fidelity to data

Error=0

Low ↔ High

같은말이다!

Sensitivity to noisy data

Low ↔ High

Question and Answer