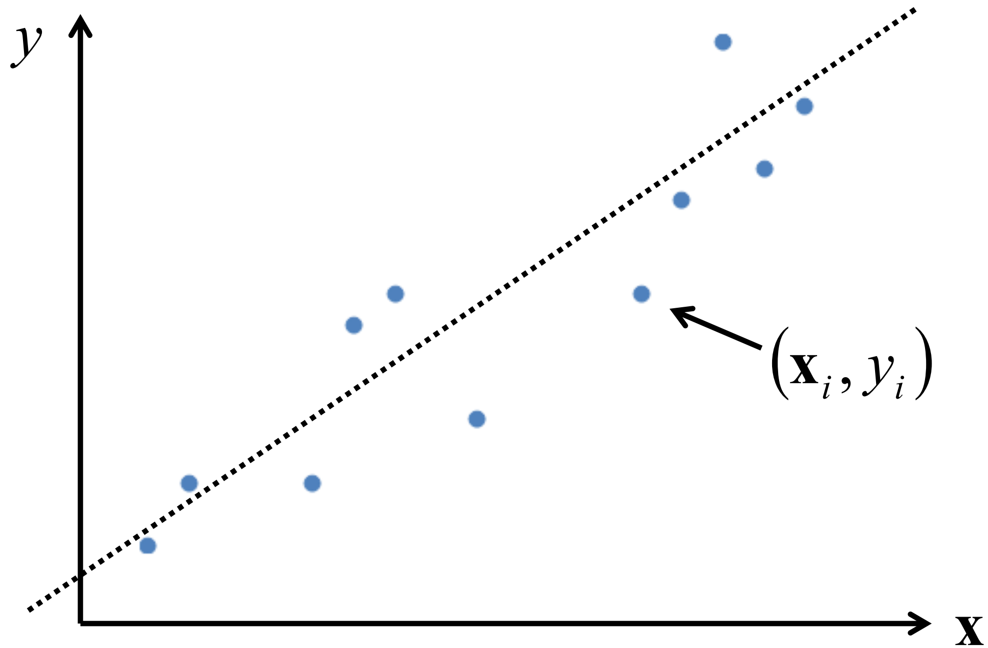
Linear & Polynomial Regression

Introduction

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

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

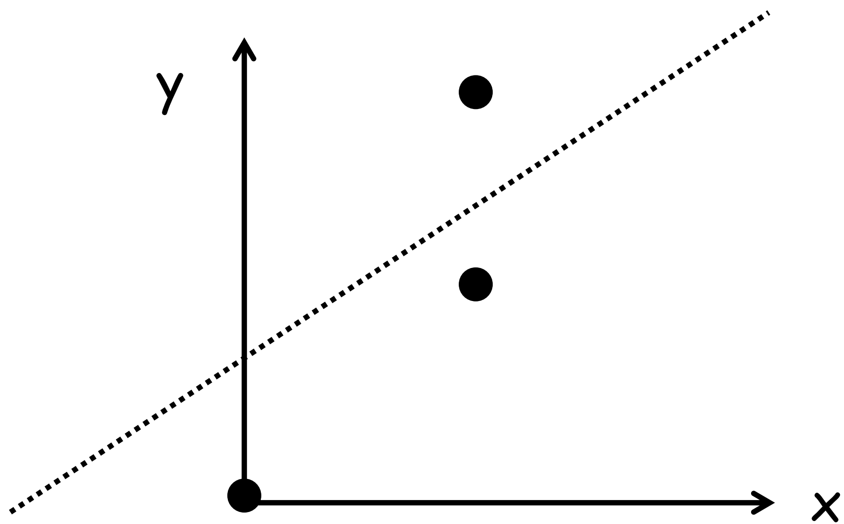


HOW?

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 line" for the given samples?**

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

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

$f(x_1)=w_1 x_1 + w_0$ is as close to y_1 as possible

$f(x_2)=w_1 x_2 + w_0$ is as close to y_2 as possible

$f(x_3)=w_1 x_3 + w_0$ is as close to y_3 as possible

⇨ $|f(x_1) - y_1|$ is minimized

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

$|f(x_3) - y_3|$ is minimized

Linear Regression

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

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

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

$\lvert f(x_1) - y_1 \rvert$	is minimized
$\lvert f(x_2) - y_2 \rvert$	is minimized
$\lvert f(x_3) - y_3 \rvert$	is minimized

$\Rightarrow$ $\displaystyle\sum_{(\mathbf{x},y)\in Data} \lvert y - f(\mathbf{x}) \rvert$ is minimized

$\Rightarrow$ $\displaystyle\sum_{(\mathbf{x},y)\in Data} \big(y - f(\mathbf{x})\big)^2$ is minimized

Linear Regression

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

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

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

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

- **But, how?**

Steps of Machine Learning

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

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

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

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

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

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

Steps of Machine Learning

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

$$E = 2{w_1}^2 + 3{w_0}^2 - 6w_1 - 6w_0 + 4w_1w_0 + 5$$

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

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

$$4w_1 + 4w_0 - 6 = 0$$

$$4w_1 + 6w_0 - 6 = 0$$

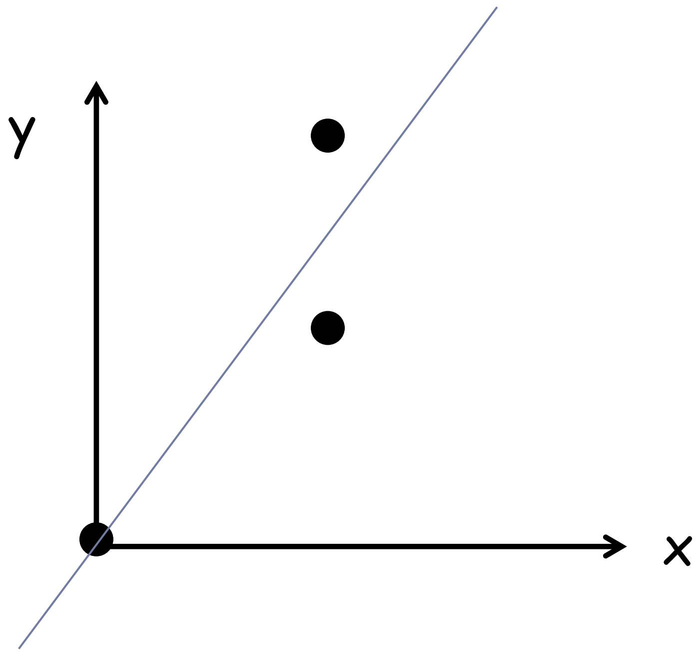
$$w_1 = 1.5$$

$$w_0 = 0.0$$

Steps of Machine Learning

- **The best-fit line is**

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



성균관대학교

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:

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

- **For example, why not**

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

instead of

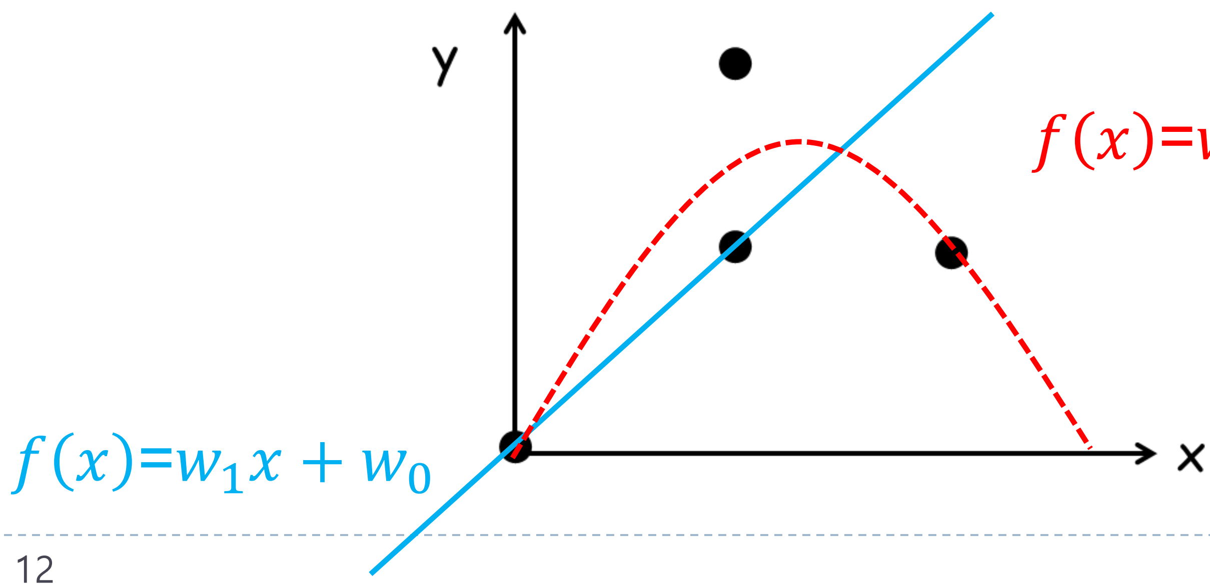
$$f(x)=w_1x + w_0$$

Hmm, I have a question

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

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

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



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:

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

$$\left.\begin{array}{ll} \text{Quadratic function of x} & f(x) = w_2 x^2 + w_1 x + w_0 \\ \\ \text{Linear function of x} & f(x) = w_1 x + w_0 \end{array}\right\} \text{Linear function of w's}$$

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

Hmm, I have a question

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

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

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

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

$$\left.\begin{array}{l} \dfrac{\partial E}{\partial w_1} = 0 \\ \dfrac{\partial E}{\partial w_2} = 0 \\ \cdots \\ \dfrac{\partial E}{\partial w_m} = 0 \end{array}\right\}$$

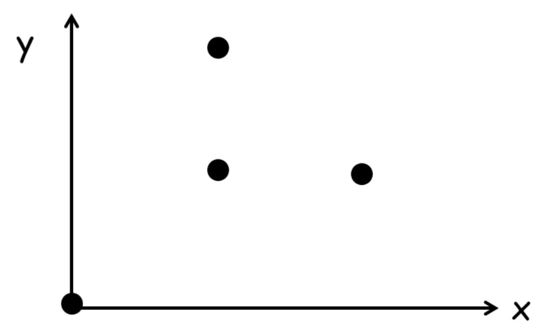
A system of linear equations

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**

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

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

determine w_0, w_1, w_2 to 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**

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

$$E(w_0, w_1, w_2) = \sum_{(\mathbf{x},y)\in Data} \left(y - (w_2x^2 + w_1x + w_0)\right)^2$$

determine w_0, w_1, w_2 to 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 to 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 to 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 to minimize**

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

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

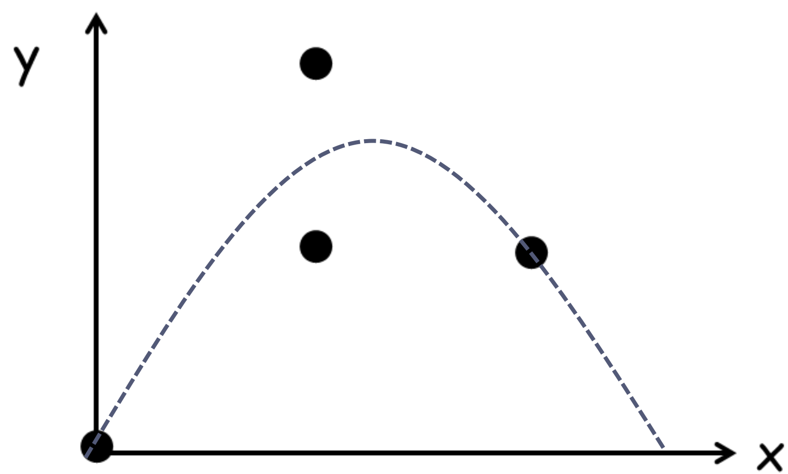
$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 12w_0 + 20w_1 + 36w_2 - 14 = 0$$

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

Let's do it

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

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



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

I have another question

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

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

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

- For example, why not

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

instead of

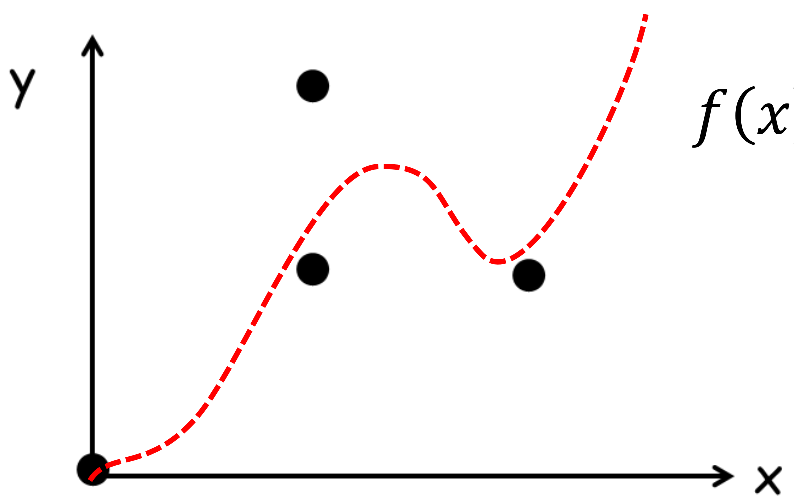
$$f(x) = w_2 x^2 + w_1 x + w_0 \qquad f(x) = w_1 x + w_0$$

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

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

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

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**

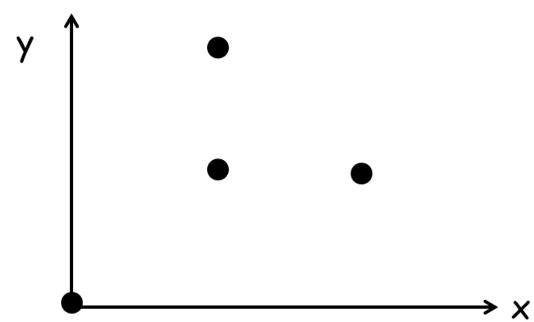
$$\left.\begin{aligned} \frac{\partial E}{\partial w_1} &= 0 \\ \frac{\partial E}{\partial w_2} &= 0 \\ &\cdots \\ \frac{\partial E}{\partial w_m} &= 0 \end{aligned}\right\} \text{A system of linear equations}$$

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 to 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 to 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 to minimize**

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

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

$$\begin{aligned} E(w_0, w_1, w_2) = {} & 2w_0^2 + 2w_1^2 + 23w_2^2 + 4w_0w_1 \\ & +14w_0w_2 + 8w_1w_2 - 8w_0 - 6w_1 - 20w_2 + 6 \end{aligned}$$

Let's do it

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

$$E(w_0, w_1, w_2) = 2w_0^2 + 2w_1^2 + 23w_2^2 + 4w_0w_1 + 14w_0w_2 + 8w_1w_2 - 8w_0 - 6w_1 - 20w_2 + 6$$

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 4w_0 + 4w_1 + 14w_2 - 8$$

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

$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 14w_0 + 8w_1 + 46w_2 - 12$$

Let's do it

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

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

$$\frac{\partial}{\partial w_1} E(w_0, w_1, w_2) = 4w_0 + 4w_1 + 8w_2 - 14 = 0$$

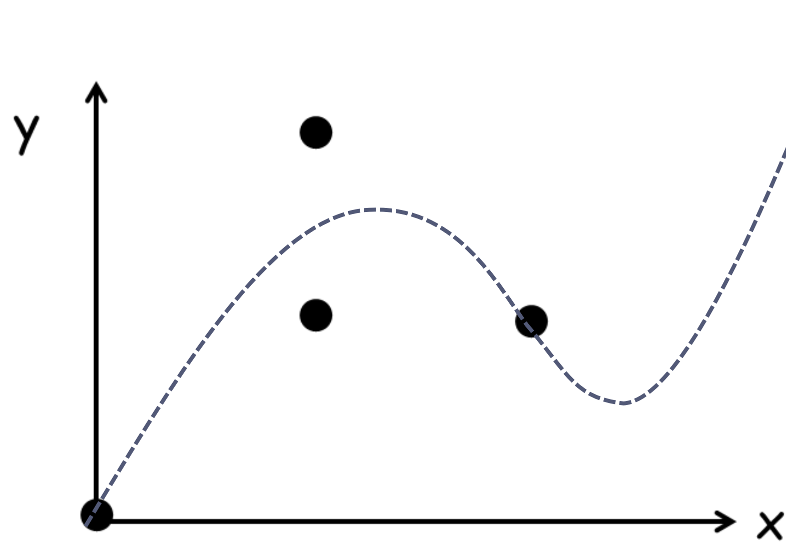
$$\frac{\partial}{\partial w_2} E(w_0, w_1, w_2) = 18w_0 + 8w_1 + 50w_2 - 12 = 0$$

$$w_0 = -\frac{1}{3}, w_1 = \frac{21}{18}, w_2 = \frac{1}{3}$$

Let's do it

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

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



$$f(x) = \frac{1}{3}2^x + \frac{21}{18}\sin\frac{\pi}{2}x - \frac{1}{3}$$

Question and Answer