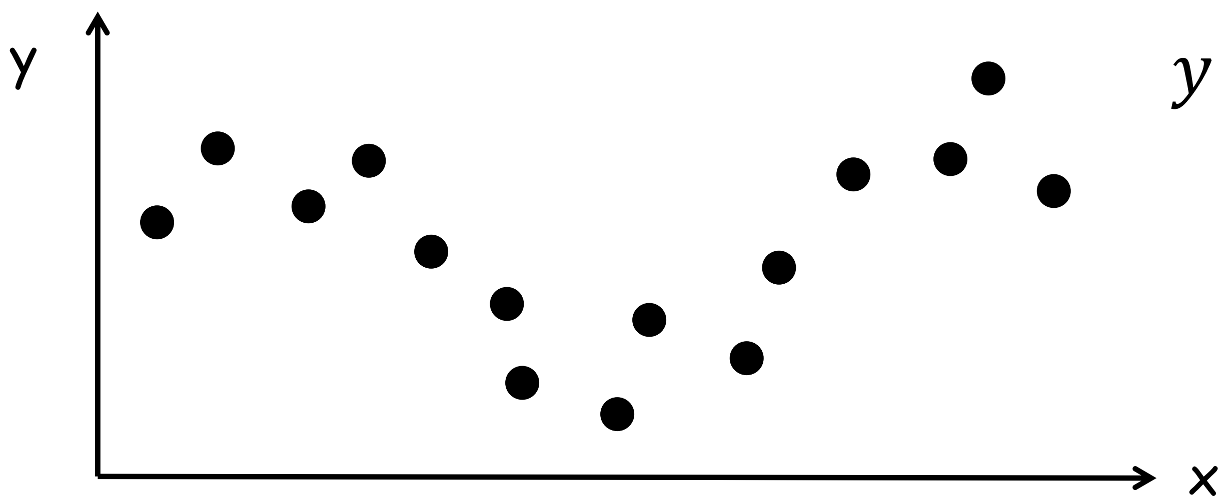
Basic Idea of Learning

Goal of Machine Learning

주어진 데이터를 가장 잘 설명하는 함수를 찾아라



$$y = f(x)$$

Steps of Machine Learning

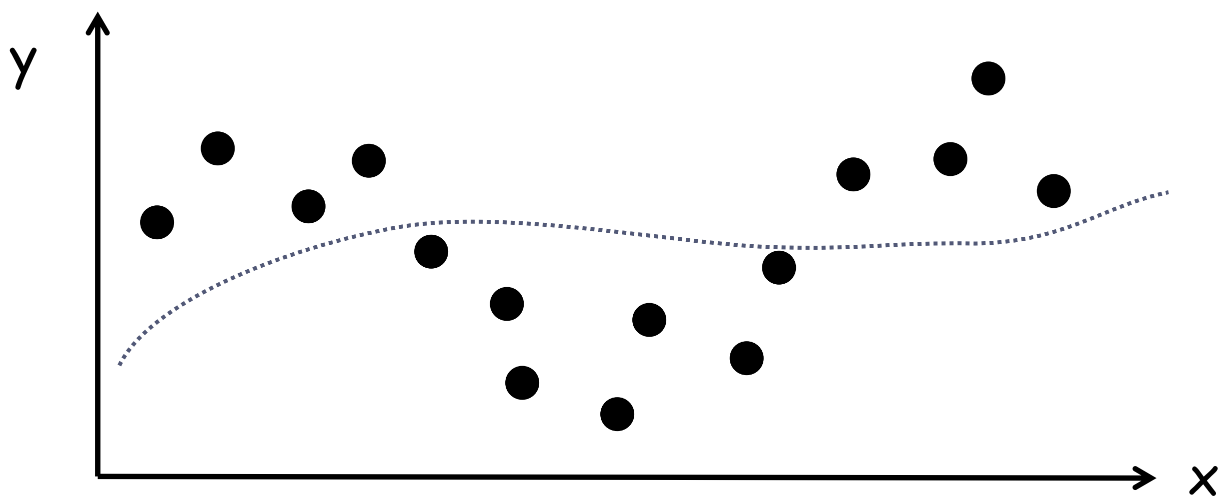
- **Step 1: 여러 종류의 모델 중에서 하나를 선택**
  - Neural Network, Deep Learning, Decision Tree, Random Forest 등

$$f(x_1, x_2, \dots, x_n; w_1, w_2, \dots, w_m)$$

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

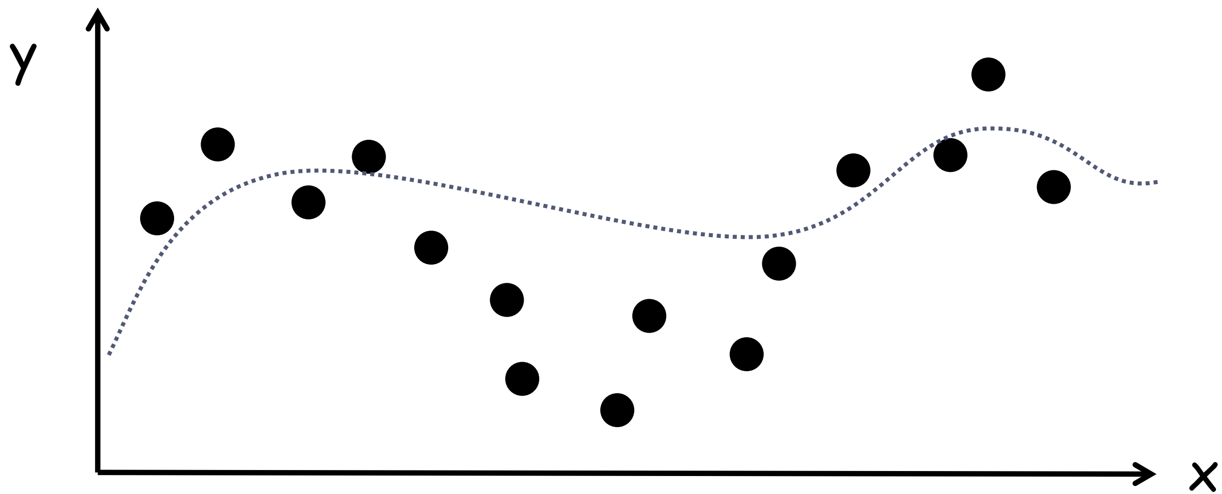
$$f(x; 1.0, 1.0, 1.0)$$



Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

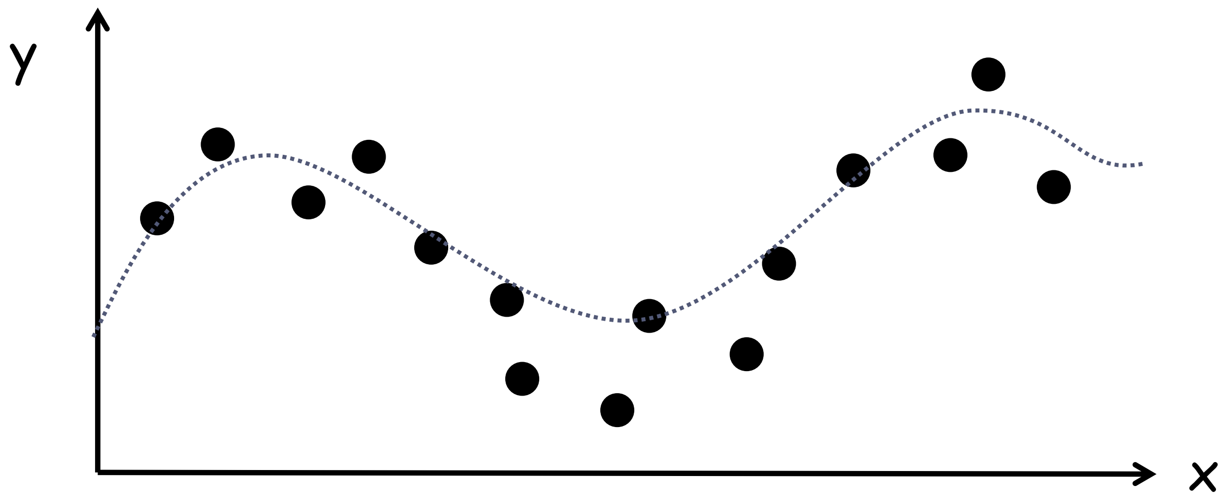
$$f(x; 1.5, 1.0, 1.5)$$



Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

$$f(x; 1.5, -1.0, 1.5)$$

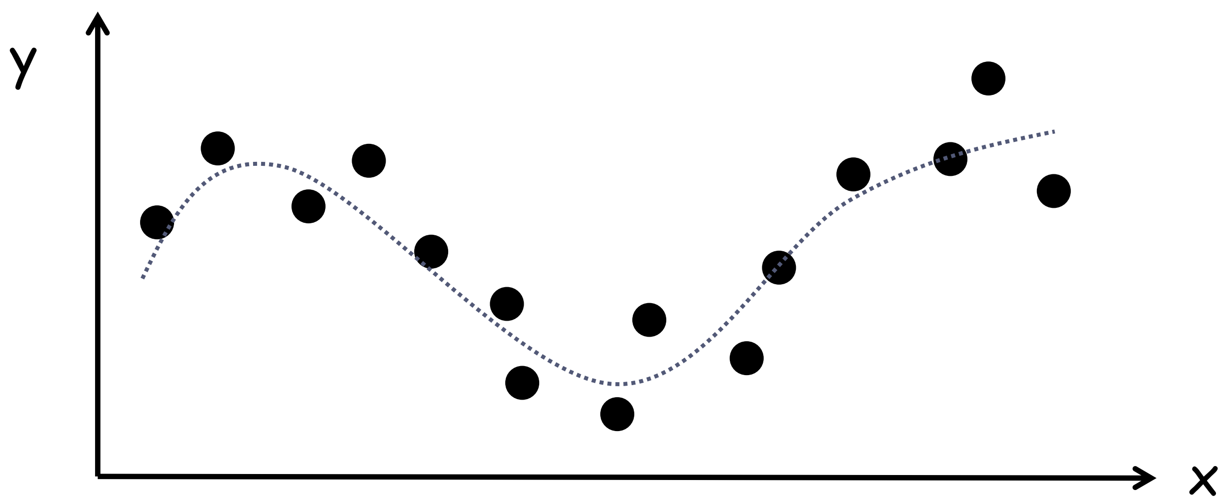


성균관대학교

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

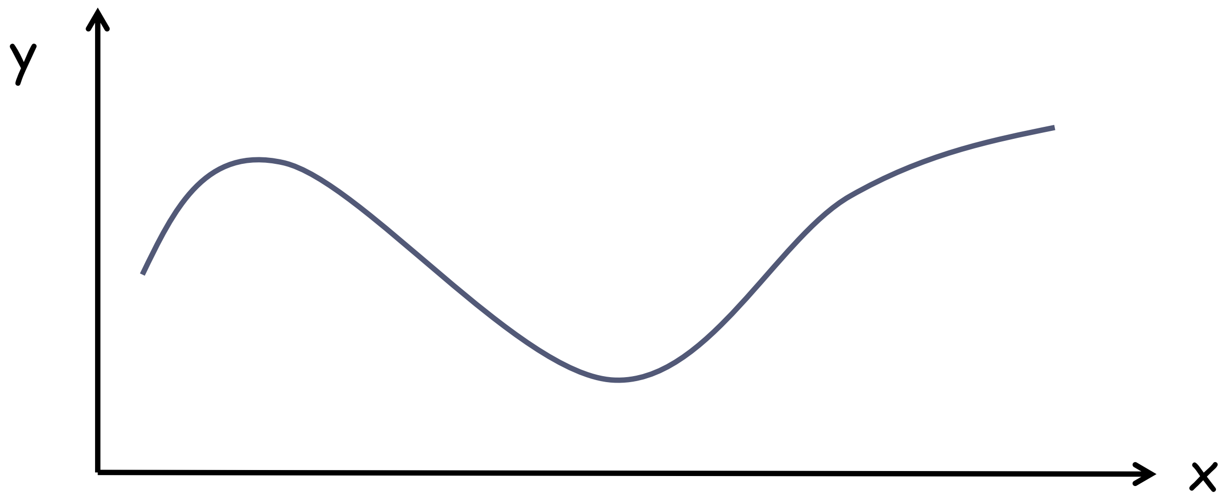
$$f(x; 1.7, -1.2, 1.6)$$



Steps of Machine Learning

- **Step 3: 결정된 f를 이용하여 값을 예측한다**

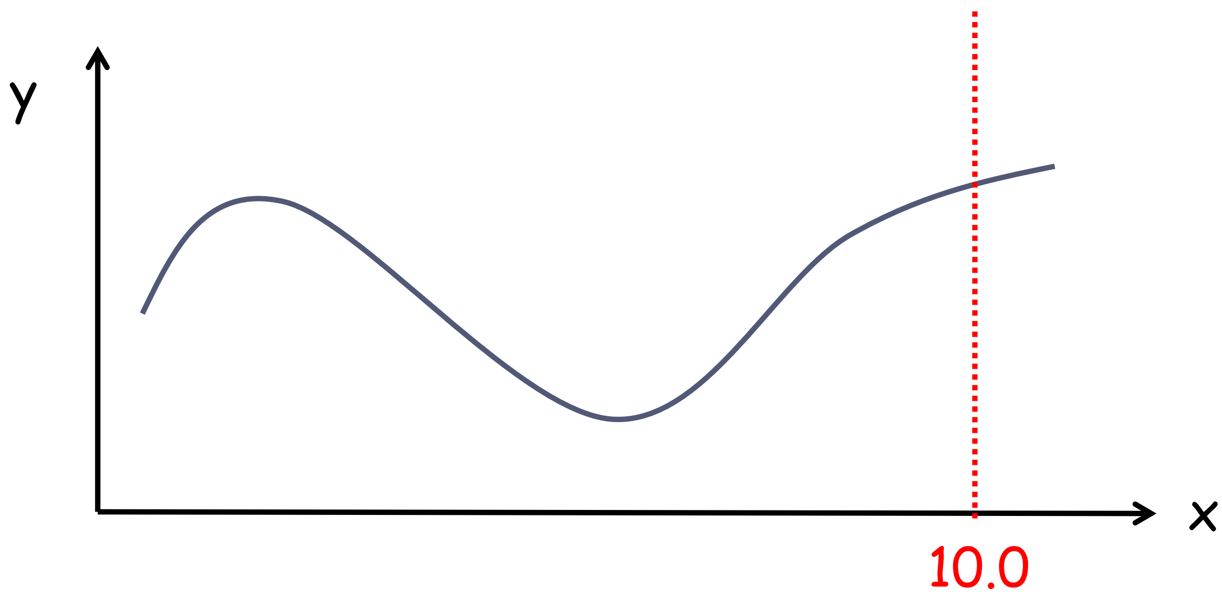
$$f(x; 1.7, -1.2, 1.6)$$



Steps of Machine Learning

- **Step 3: 결정된 f를 이용하여 값을 예측한다**

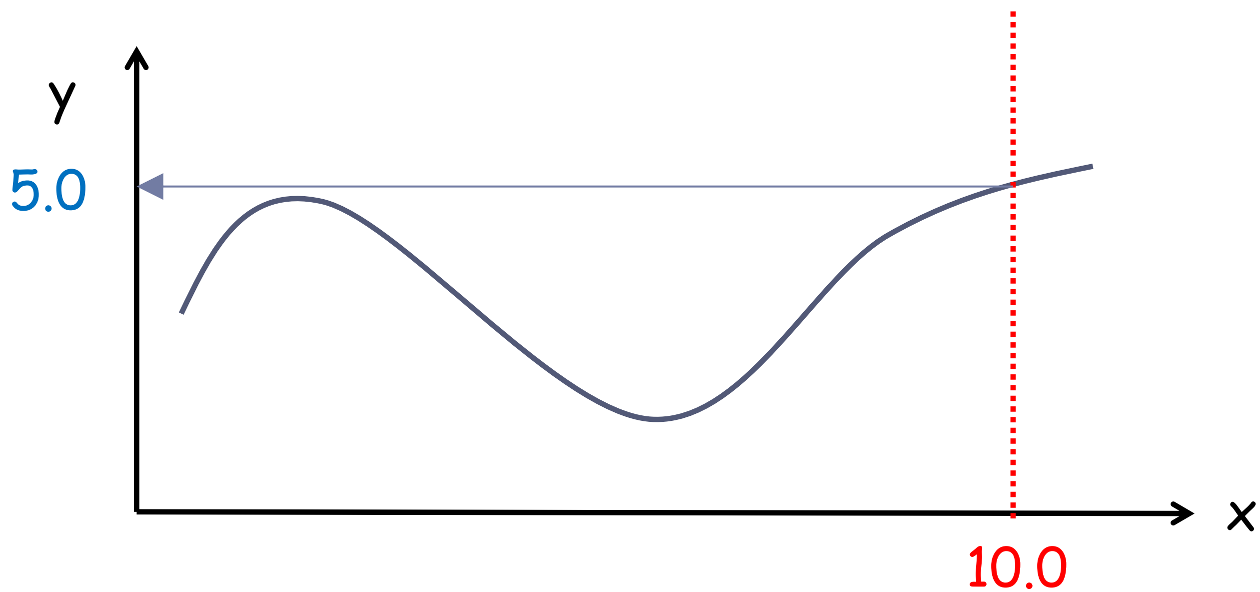
$$f(10.0\,;1.7,-1.2,1.6)$$



Steps of Machine Learning

- **Step 3: 결정된 f를 이용하여 값을 예측한다**

$$5.0 = f(10.0\,; 1.7, -1.2, 1.6)$$



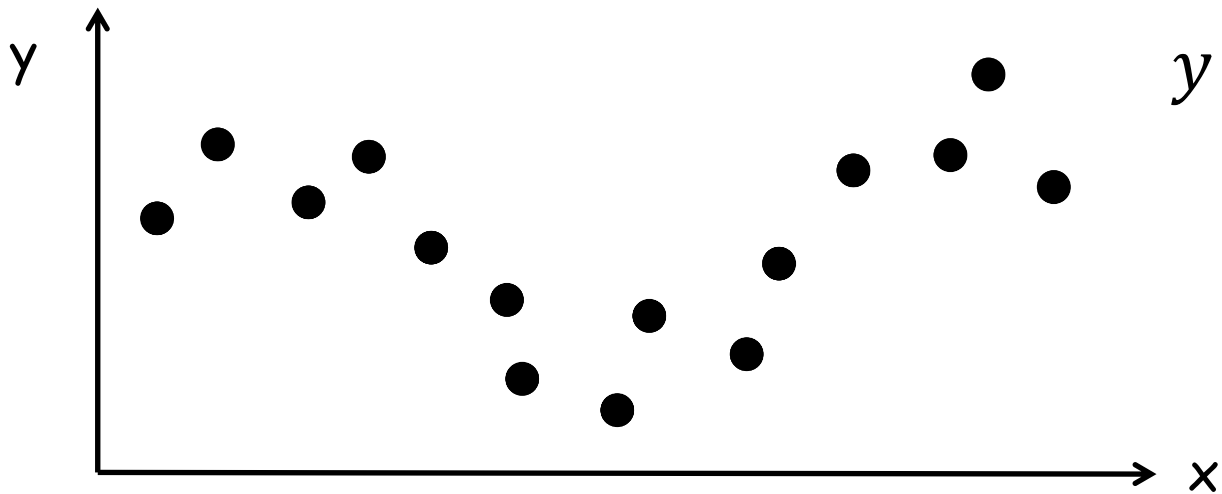
Some Issues

몇가지 Issue들

주어진 데이터를 가장 잘 설명하는 함수모양을 찾아라

이게 무슨 소리?

어떻게?

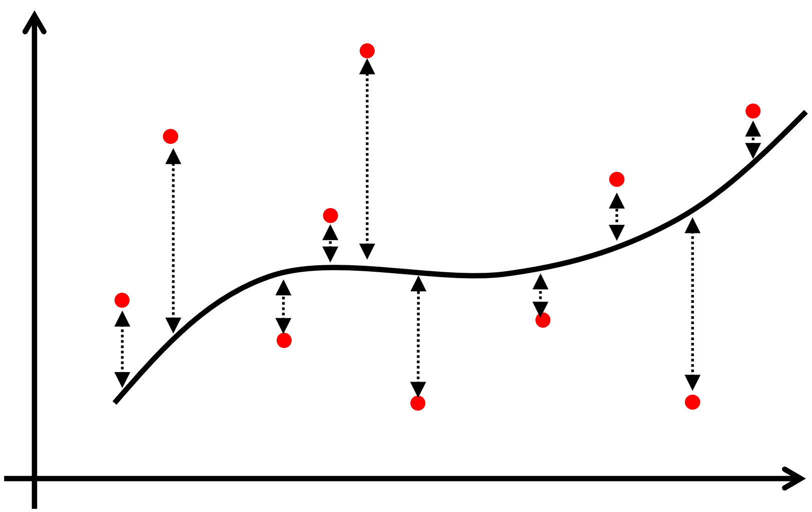


$$y = f(x)$$

성균관대학교

가장 잘 설명하는

- 데이터를 가장 잘 설명하는?
  - 주어진 데이터와 오류를 최소하는 함수
  - 오류는 무엇인가

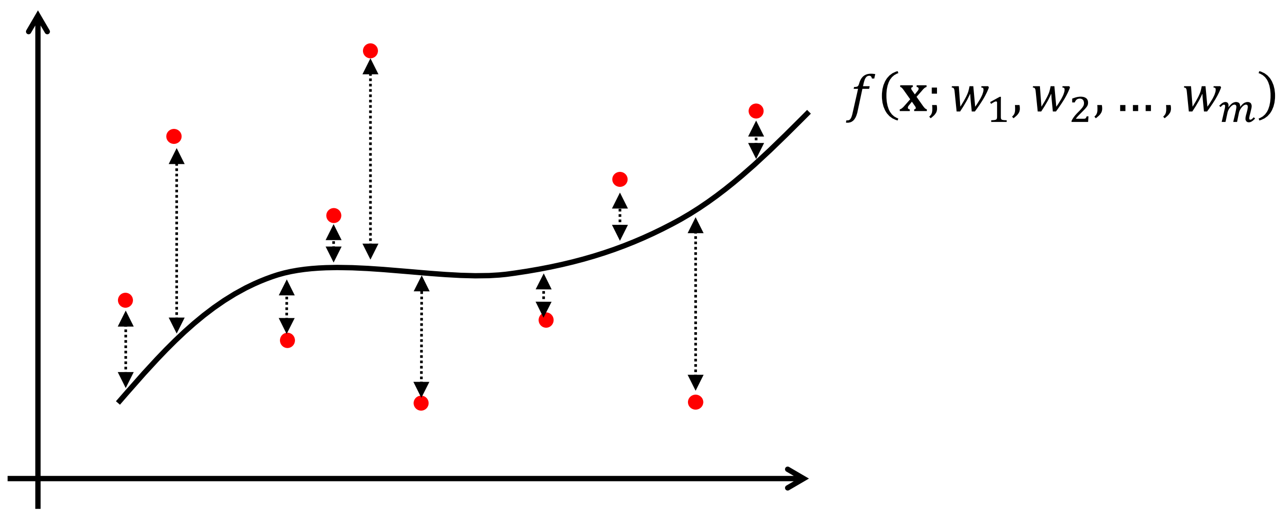


$f(\mathbf{x}; w_1, w_2, \ldots, w_m)$

가장 잘 설명하는

- **주어진 데이터와 오류를 최소화하는 함수**

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



가장 잘 설명하는 함수를 찾아라

- **주어진 데이터와 오류를 최소화하는 함수는 어떻게 찾지?**

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

- **함수모양은 $w_1, w_2, \ldots, w_m$가 결정하므로, 다시 말하면**

Error를 최소화하는 $w_1, w_2, \ldots, w_m$를 찾으면,
주어진 데이터와 오류를 최소화 하는 함수를 찾을 수 있다.

가장 잘 설명하는 함수를 찾아라

- 즉, "주어진 데이터를 가장 잘 설명하는 함수를 찾아라" 라는 문제는

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

문제로 변환될 수 있다

가장 잘 설명하는 함수를 찾아라

- 즉, 우리는 아래 문제를 해결하면 기계학습 문제를 해결할 수 있다

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

- 그런데, 어떻게??
  - 수단과 방법을 가리지 말고, 아무튼 찾아라.

Minimizing Error Functions

고등학교 때의 기억을

- **이런 류는 너무 어려울 수 있으니까..**

아래 함수를 최소화하는 $w_1, w_2, \ldots, w_m$를 찾아라

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

- **대신, 아래 문제를 해결하시오**

아래 함수를 최소화하는 w_1 을 찾아라

$$E(w_1) = {w_1}^2 + 2w_1 + 3$$

고등학교 때의 기억을

아래 문제를 해결하시오

아래 함수를 최소화하는 w_1 을 찾아라

$$E(w_1) = {w_1}^2 + 2w_1 + 3$$

흠.. 너무 쉽다

1. 우선 E를 w_1에 대해서 미분하고

$$\frac{dE(w_1)}{dw_1} = 2w_1 + 2$$

2. 이것을 0으로 만드는 w_1을 찾으면

$$2w_1 + 2 = 0$$

3. 그게 답이지

$$w_1 = -1$$

고등학교 때의 기억을

- **아.. 그러면 아래 문제라고 별 거 있겠어??**

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

- **우선 E를 w_i들에 대해서 편미분을 하고, 이것을 모두 0으로 만드는 w_i를 찾으면 되겠네..**

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

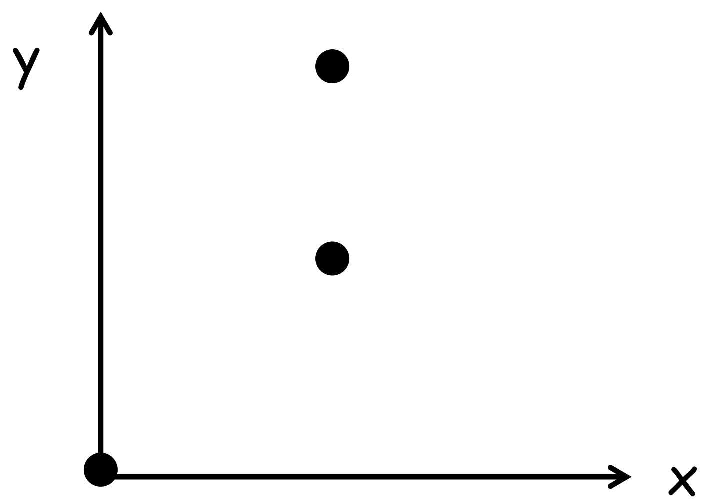
를 동시에 만족하는 $w_1, w_2, \ldots, w_m$를 찾으면 된다..

기계학습 끝!!

그럼, 한번 해 봅시다

다음 데이터에 가장 부합하는 모델은?

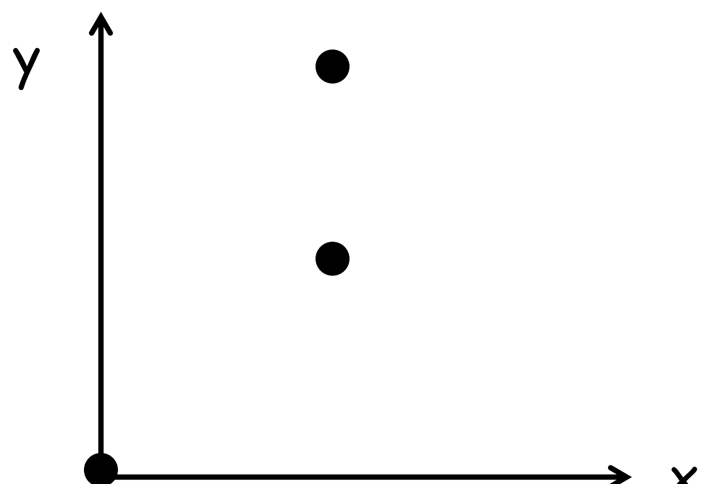
$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)\}$$



Steps of Machine Learning

- **Step 1: 여러 종류의 Approximator 중에서 하나를 선택**
  - Approximator로 직선을 선택

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

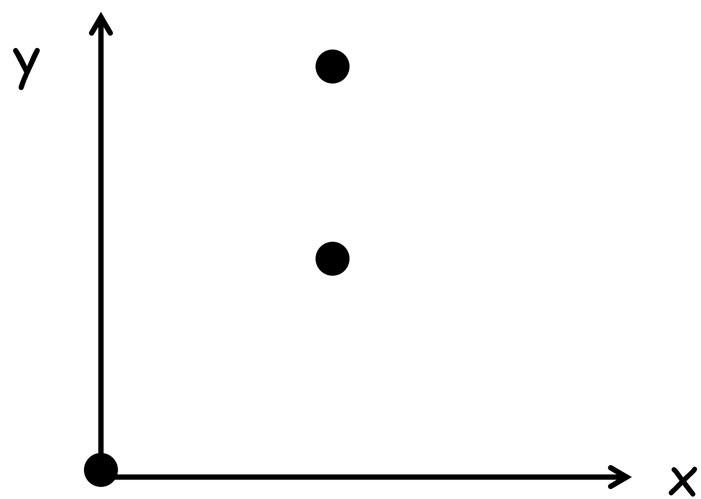


성균관대학교

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

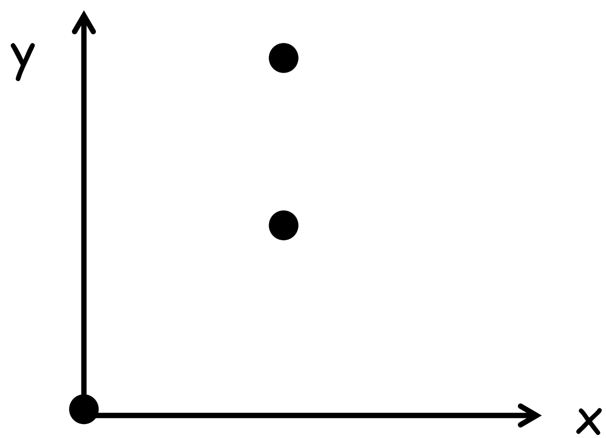


Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

$(0.0, 0.0)$ $(1.0, 1.0)$ $(1.0, 2.0)$



$$f(0.0; w_0, w_1) \approx 0.0$$

$$f(1.0; w_0, w_1) \approx 1.0$$

$$f(1.0; w_0, w_1) \approx 2.0$$

이렇게 되도록 w_0 와 w_1을 결정하라

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

$$[f(0.0; w_0, w_1) - 0.0]^2$$

$$[f(1.0; w_0, w_1) - 1.0]^2$$

$$[f(1.0; w_0, w_1) - 2.0]^2$$

의 합이 최소가 되도록
w_0 와 w_1을 결정하라

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

를 최소가 되도록 w_0 와 w_1을 결정하라

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

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

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

를 최소가 되도록 w_0 와 w_1을 결정하라

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

를 최소가 되도록 w_0 와 w_1을 결정하라

$$Data = \{(0.0, 0.0), (1.0, 1.0), (1.0, 2.0)\} \qquad f(x; w_0, w_1) = w_1 x + w_0$$

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

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

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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

를 최소가 되도록 w_0 와 w_1을 결정하라

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

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

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

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

의 해가 바로 우리가 찾는 것이다.

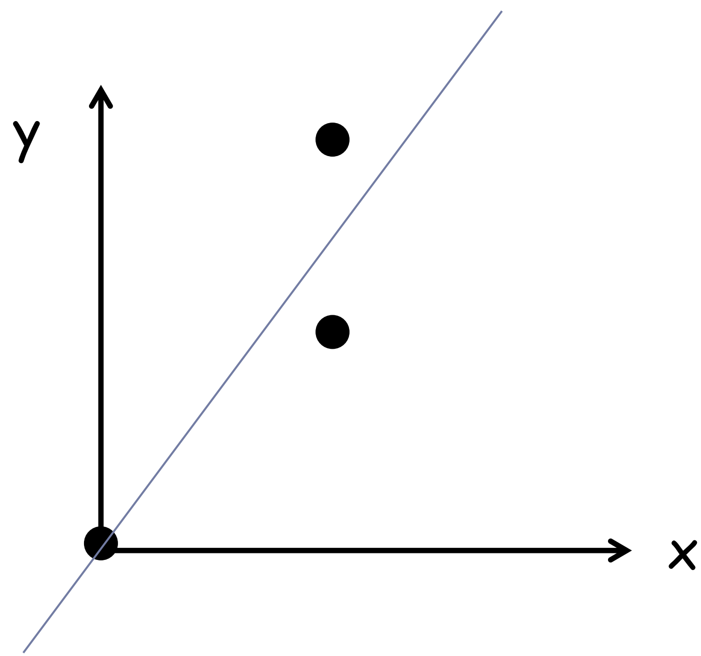
$$w_1 = 1.5$$

$$w_0 = 0.0$$

Steps of Machine Learning

- 주어진 데이터에 가장 부합하는 직선은 바로

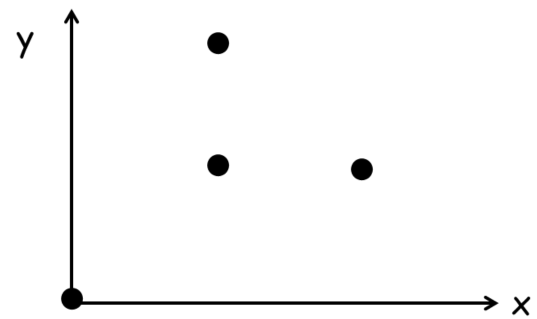
$$f(x) = 1.5x + 0.0$$



숙제 1

- 주어진 데이터에 가장 부합하는 모델을 찾아라

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



- 기본 모델로 2차 함수를 사용하자

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

  - w_0, w_1, w_2를 결정하라

성균관대학교

숙제 1

- **우리가 풀어야할 문제는**

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

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

일 때 아래 함수가 최소가 되도록 하는 w_0, w_1, w_2를 결정하라

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

Machine Learning 이란

- 아래의 문제를 해결하는 것이다.

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

여러분이 선택한 모델 $f(\mathbf{x}; \mathbf{w})$

에 대하여 아래 함수가 최소가 되도록 하는 **w**를 결정하라

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

Learning with Gradient Descent Method

고등학교 때의 기억을

- **아.. 그러면 아래 문제라고 별 거 있겠어??**

아래 함수를 최소화하는 $w_1, w_2, \ldots, w_m$를 찾아라

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

- **우선 E를 w_i들에 대해서 편미분을 하고, 이것을 모두 0으로 만드는 w_i를 찾으면 되겠네..**

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

를 동시에 만족하는 $w_1, w_2, \ldots, w_m$를 찾으면 된다..

아.. 쉽다??

고등학교 때의 기억을

- **그런데, 좀 심란하다..**

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

를 동시에 만족하는 $w_1, w_2, \ldots, w_m$를 찾아라.

- **답은 항상 존재하나?**
  - 존재하지 않으면 어떻게 하지?
- **내가 어떤 모델을 선택해도 항상 풀 수 있나?**
  - 풀 수 없으면 어떻게 하지?

고등학교 때의 기억을

- 그런데, 좀 심란하다..
  - 이런 문제에 답은 항상 존재하나?
    - 네.. 문제가 아래처럼 제곱의 형태이기 때문에..

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

  - 내가 어떤 모델을 선택해도 항상 풀 수 있나?
    - 해결하지 못 할 수도 있습니다
  - 풀 수 없으면 어떻게 하지?
    - 답을 구하는 다른 방법을 인간은 알고 있지 않습니다 ㅠㅠ

그래서 우리는 바꿉니다

- **우리의 원래 문제 대신**

아래 함수를 최소화하는 **w**를 찾아라

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

- **아래 문제를 해결합니다**

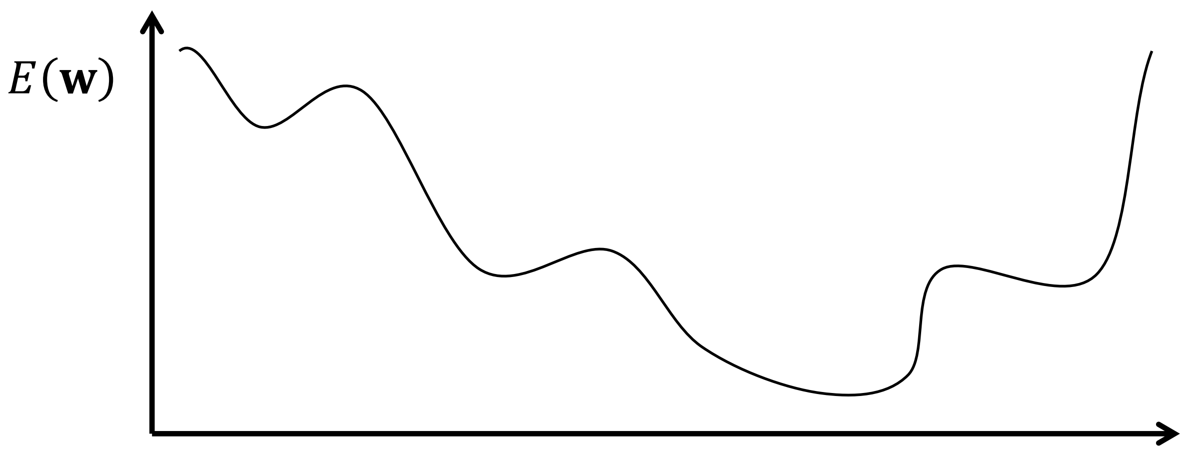
아래 함수를 minimum스럽게 만드는 **w**를 찾아라

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

그래서 우리는 바꿉니다

- **그런데, 무슨 뜻인가요?**

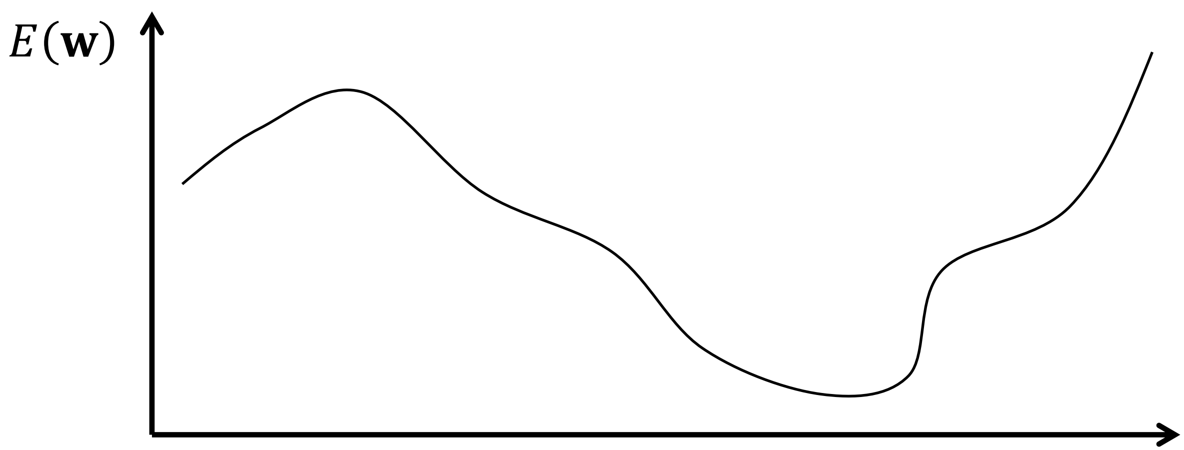
> 아래 함수를 minimum스럽게 만드는 **w**를 찾아라
>
> $$E(\mathbf{w}) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x};\mathbf{w})\right)^2$$



그래서 우리는 바꿉니다

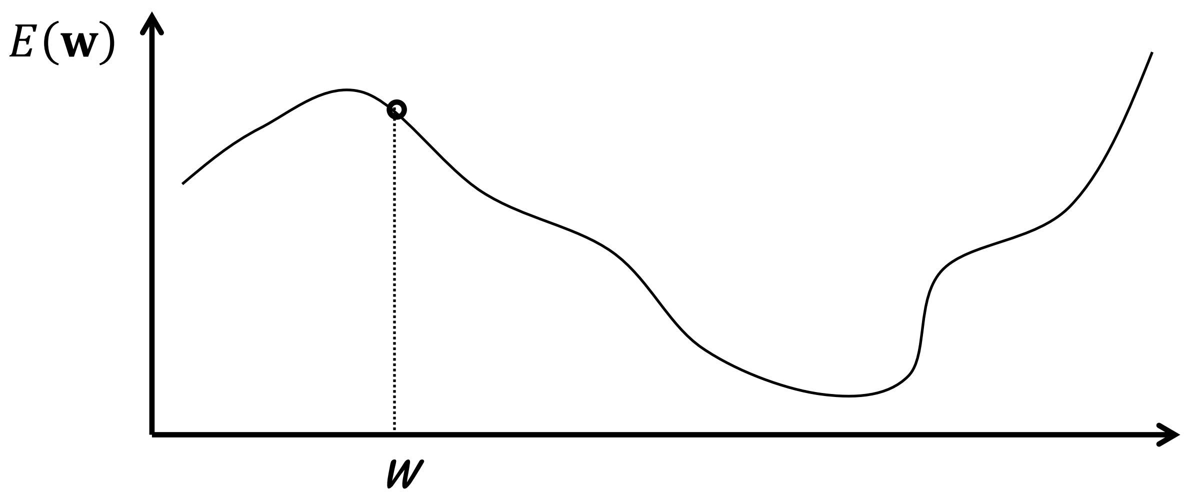
- **이 문제는 생각보다 풀기 쉽습니다**

경사면 따라 내려가기



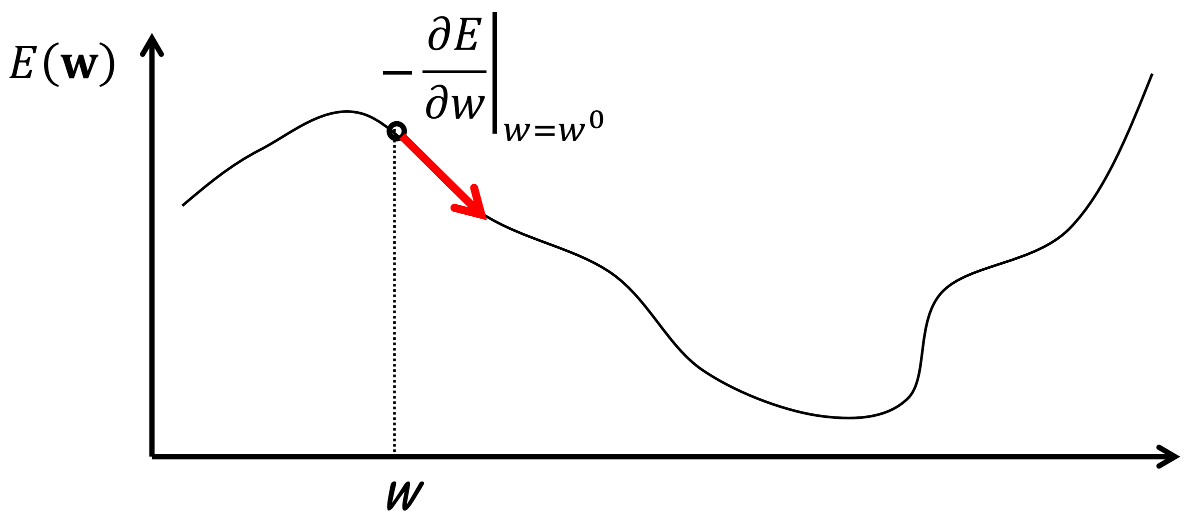
경사면 따라 내려가기

- **1. 임의의 시작점을 Random하게 잡기**



경사면 따라 내려가기

- **2. 현재의 위치에서 경사를 구하기**
  - 미분은 가능하기 때문에 현재 위치에서의 경사는 구할 수 있음
  - 우리는 내려가기 때문에 기울기의 반대 방향이 필요함

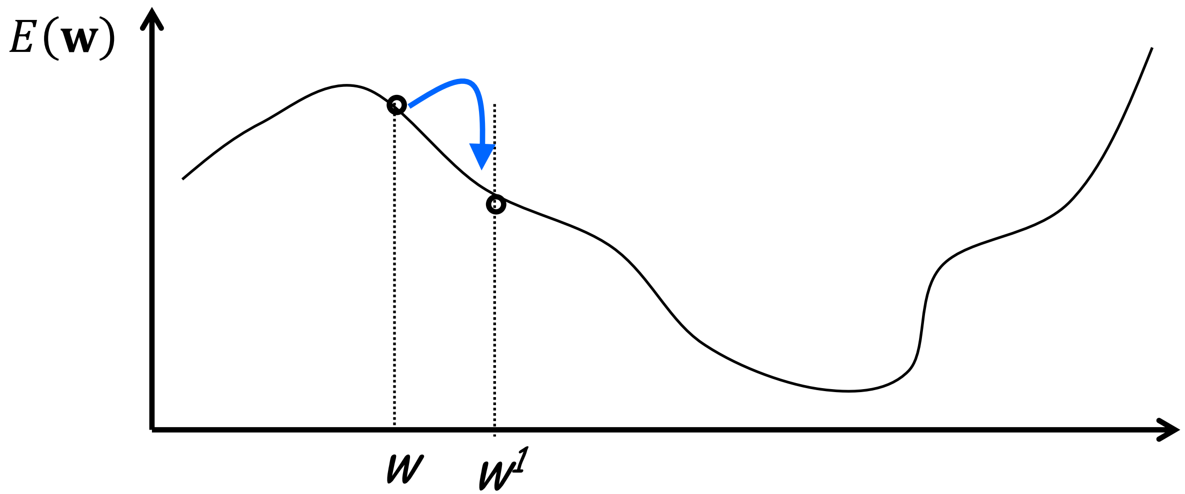


성균관대학교

경사면 따라 내려가기

- 3. 그 방향으로 매우 조금 이동하기
  - 크게 점프하면 건너 편으로 갈 수도 있음

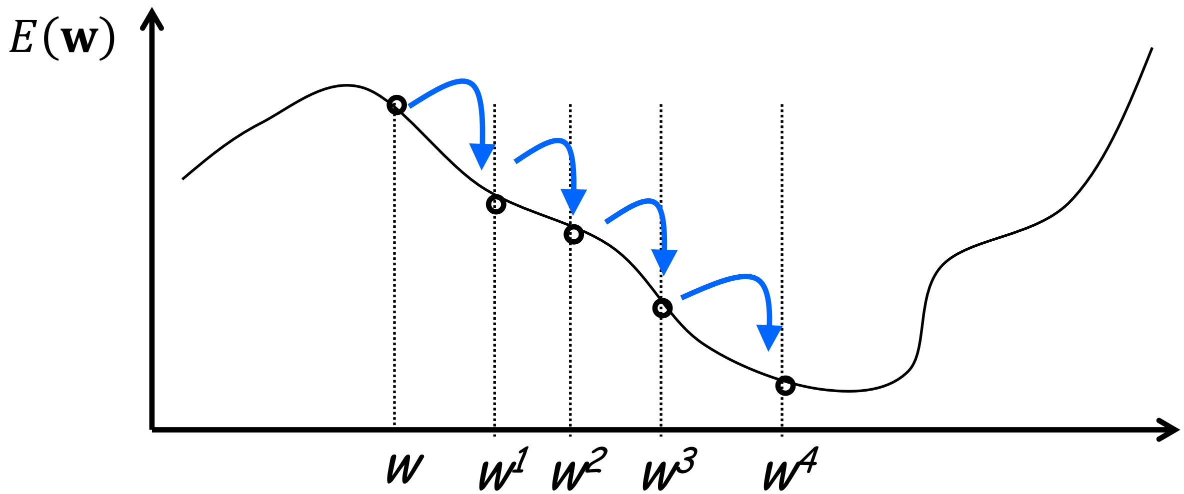
$$w^1 = w^0 - \eta \left.\frac{\partial E}{\partial w}\right|_{w=w^0}$$



경사면 따라 내려가기

- **4. 기울기가 0인 곳에 도달할 때까지 계속하기**

$$w^{t+1} = w^{t} - \eta \left.\frac{\partial E}{\partial w}\right|_{w=w^{t}}$$



Gradient Descent Method

- **Steps**

learning rate

Randomly choose an initial solution, w^0

Repeat

$$w^{t+1} = w^t - \eta \left.\frac{dE}{dw}\right|_{w=w^t}$$

Until stopping condition is satisfied

- $|w^{t+1} - w^t|$ is very small
- f(w) little moves
- fixed number of iterations

Gradient Descent Method

- 다차원인 경우 각 변수에 대해서 진행

Randomly choose an initial solution, $w_0^0\ w_1^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \left.\frac{\partial E}{\partial w_0}\right|_{w_0=w_0^t, w_1=w_1^t}$$

$$w_1^{t+1} = w_1^t - \eta \left.\frac{\partial E}{\partial w_1}\right|_{w_0=w_0^t, w_1=w_1^t}$$

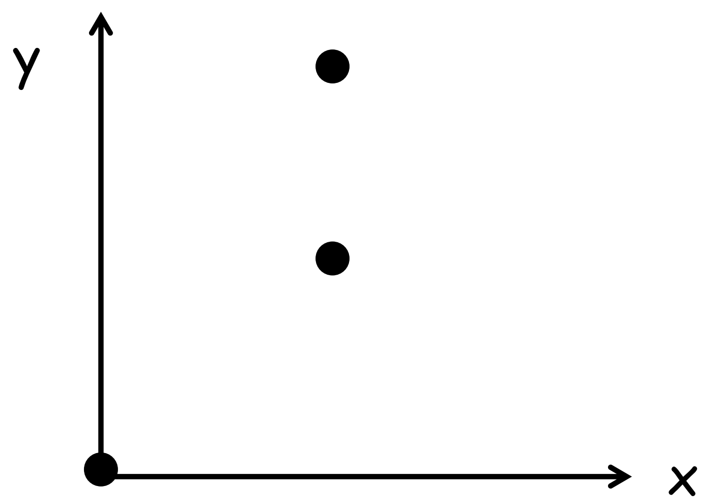
Until stopping condition is satisfied

지금까지 말한 것을 합하면

Goal of Machine Learning

- 다음 데이터를 가장 잘 설명하는 함수를 찾아라

$$(0.0, 0.0)\ (1.0, 1.0)\ (1.0, 2.0)$$

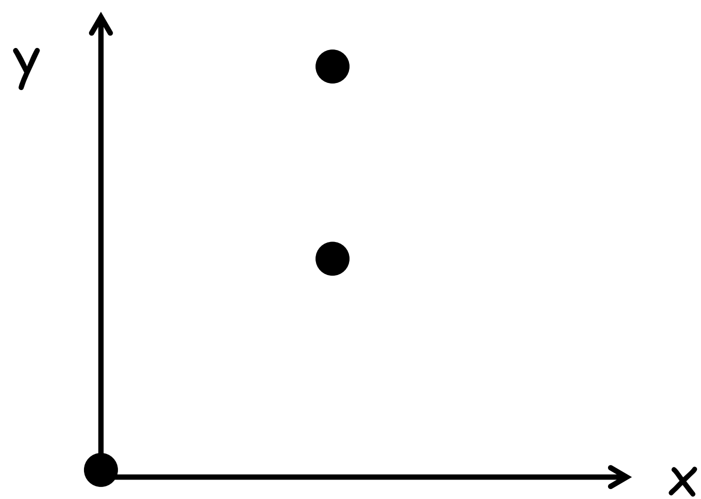


성균관대학교

Steps of Machine Learning

- **Step 1: 여러 종류의 Approximator 중에서 하나를 선택**
  - Approximator로 직선을 선택

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

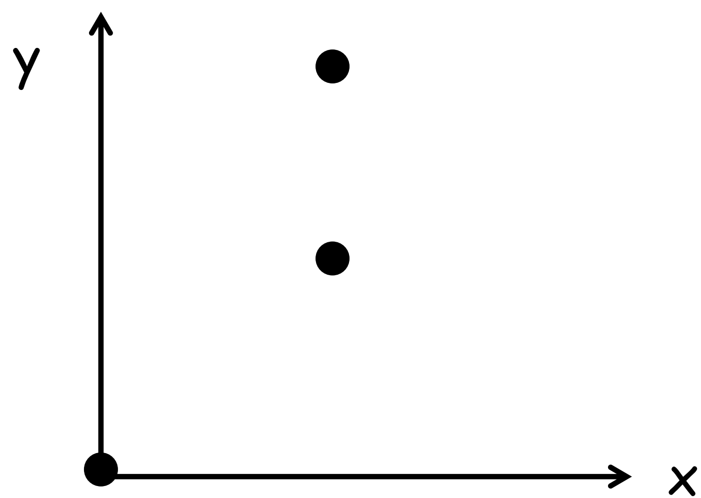


성균관대학교

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**

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



Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - Error를 정의한다 $$E(w_0, w_1) = \sum_{(\mathbf{x},y)\in Data} \left(y - f(\mathbf{x}; w_1, w_0)\right)^2$$
  - 주어지 데이터와 함수 f를 Error에 대입한다

$$(0.0, 0.0)\ (1.0, 1.0)\ (1.0, 2.0) \qquad f(x; w_0, w_1) = w_1 x + w_0$$

  - 잘 정리한다

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

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

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - Error의 Gradient의 일반식을 구한다

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

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

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

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - Error의 Gradient의 일반식을 GDM에 대입한다

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

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

Randomly choose an initial solution, $w_0^0\ w_1^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \left.\frac{\partial E}{\partial w_0}\right|_{w_0 = w_0^t, w_1 = w_1^t}$$

$$w_1^{t+1} = w_1^t - \eta \left.\frac{\partial E}{\partial w_1}\right|_{w_0 = w_0^t, w_1 = w_1^t}$$

Until stopping condition is satisfied

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - Gradient Descent Method로 Error를 최소하는 w_1, w_2를 찾는다

Randomly choose an initial solution, w_0^0 w_1^0

Repeat

$$w_0^{t+1} = w_0^t - \eta(4w_1^t + 6w_0^t - 6)$$

$$w_1^{t+1} = w_1^t - \eta(4w_1^t + 4w_0^t - 6)$$

Until stopping condition is satisfied

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - **Gradient Descent Method로 Error를 최소하는 w_1, w_2를 찾는다**
    - η 를 작은 값으로 선택한다. $\eta = 0.1$
    - Random하게 w_0^0 w_1^0 를 선택한다. $w_0^0 = 1$ $w_1^0 = 1$

Randomly choose an initial solution, w_0^0 w_1^0

Repeat

$$w_0^{t+1} = w_0^t - \eta(4w_1^t + 6w_0^t - 6)$$

$$w_1^{t+1} = w_1^t - \eta(4w_1^t + 4w_0^t - 6)$$

Until stopping condition is satisfied

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - **Gradient Descent Method로 Error를 최소하는 w_1, w_2를 찾는다**
    - 루프를 반복한다

Randomly choose an initial solution, w_0^0 w_1^0

Repeat

$$w_0^{t+1} = w_0^t - \eta(4w_1^t + 6w_0^t - 6)$$

$$w_1^{t+1} = w_1^t - \eta(4w_1^t + 4w_0^t - 6)$$

Until stopping condition is satisfied

$$w_0^0 = \mathbf{1}$$
$$w_1^0 = \mathbf{1}$$

$$w_0^1 = \mathbf{1} - 0.1(4\times\mathbf{1} + 6\times\mathbf{1} - 6) = \mathbf{0.6}$$
$$w_1^1 = \mathbf{1} - 0.1(4\times\mathbf{1} + 4\times\mathbf{1} - 6) = \mathbf{0.8}$$

$$w_0^2 = \mathbf{0.6} - 0.1(4\times\mathbf{0.8} + 6\times\mathbf{0.6} - 6) = \mathbf{0.54}$$
$$w_1^2 = \mathbf{0.8} - 0.1(4\times\mathbf{0.8} + 4\times\mathbf{0.6} - 6) = \mathbf{0.84}$$

$$w_0^3 = \mathbf{0.54} - 0.1(4\times\mathbf{0.84} + 6\times\mathbf{0.54} - 6) = \mathbf{0.480}$$
$$w_1^3 = \mathbf{0.84} - 0.1(4\times\mathbf{0.84} + 4\times\mathbf{0.54} - 6) = \mathbf{0.888}$$

Steps of Machine Learning

- **Step 2: f가 주어진 데이터에 가장 잘 부합도록 w를 잘 조정한다**
  - Gradient Descent Method로 Error를 최소하는 w_0, w_1를 찾는다
    - 루프를 정해진 횟수만큼 수행한 후 마지막 값을 사용한다

$$w_0^3 = \mathbf{0.54} - 0.1(4 \times \mathbf{0.84} + 6 \times \mathbf{0.54} - 6) = \mathbf{0.480}$$
$$w_1^3 = \mathbf{0.84} - 0.1(4 \times \mathbf{0.84} + 4 \times \mathbf{0.54} - 6) = \mathbf{0.888}$$

$$w_0^4 = \mathbf{0.480} - 0.1(4 \times \mathbf{0.888} + 6 \times \mathbf{0.480} - 6) = \mathbf{0.4368}$$
$$w_1^4 = \mathbf{0.888} - 0.1(4 \times \mathbf{0.888} + 4 \times \mathbf{0.480} - 6) = \mathbf{0.9408}$$

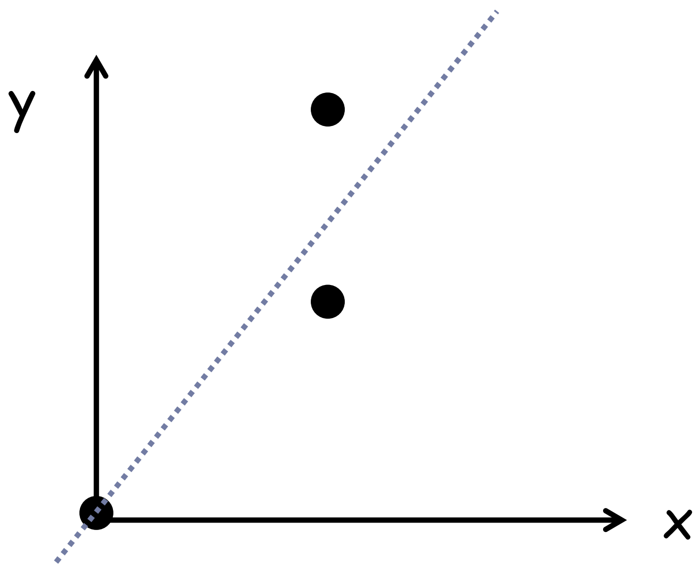
$$\dots$$

$$w_0^{100} = \mathbf{0.00007713}$$
$$w_1^{100} = \mathbf{1.49989171}$$

Final Results

- **최종 결과**

$$f(x) = 1.49989171x + 0.00007713$$



그러나

- **실제로 구현하는 방법**

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

$$= \sum_{(\mathbf{x},y)\in Data} E_i(x_i, y_i; w_0, w_1)$$

where $E_i(x_i, y_i; w_0, w_1) = \left(y_i - f(x_i; w_0, w_1)\right)^2$

$$\frac{\partial}{\partial w_j} E(w_0, w_1) = \frac{\partial}{\partial w_j} \sum_{(\mathbf{x},y)\in Data} E_i(x_i, y_i; w_0, w_1)$$

$$= \sum_{(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_j} E_i(x_i, y_i; w_0, w_1)$$

그러나

Randomly choose an initial solution, w_0^0 w_1^0

Repeat

$$w_0^{t+1} = w_0^t - \eta \left.\frac{\partial E}{\partial w_0}\right|_{w_0=w_0^t, w_1=w_1^t}$$

$$w_1^{t+1} = w_1^t - \eta \left.\frac{\partial E}{\partial w_1}\right|_{w_0=w_0^t, w_1=w_1^t}$$

Until stopping condition is satisfied

그러나

Randomly choose an initial solution, $w_0^0\ w_1^0\ w_2^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_0} E_d(x, y; w_0, w_1)\Bigg|_{w_0=w_0^t, w_1=w_1^t}$$

$$w_1^{t+1} = w_1^t - \eta \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_1} E_d(x, y; w_0, w_1)\Bigg|_{w_0=w_0^t, w_1=w_1^t}$$

Until stopping condition is satisfied

그러나

Randomly choose an initial solution, $w_0^0\ w_1^0$

Repeat

$$g_0^t = \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_0} E_d(x,y;w_0,w_1)\Big|_{w_0=w_0^t, w_1=w_1^t,}$$

$$g_1^t = \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_1} E_d(x,y;w_0,w_1)\Big|_{w_0=w_0^t, w_1=w_1^t,}$$

$$w_0^{t+1} = w_0^t - \eta g_0^t$$

$$w_1^{t+1} = w_1^t - \eta g_1^t$$

Until stopping condition is satisfied

그러나

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

$$\begin{aligned}\frac{\partial}{\partial w_0} E_d(x, y; w_0, w_1) &= \frac{\partial}{\partial w_0} \big(y - f(x; w_0, w_1)\big)^2 \\ &= -2\big(y - f(x; w_0, w_1)\big) \frac{\partial}{\partial w_0} f(x; w_0, w_1) \\ &= -2\big(y - f(x; w_0, w_1)\big)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial w_1} E_d(x, y; w_0, w_1) &= \frac{\partial}{\partial w_1} \big(y - f(x; w_0, w_1)\big)^2 \\ &= -2\big(y - f(x; w_0, w_1)\big) \frac{\partial}{\partial w_1} f(x; w_0, w_1) \\ &= -2\big(y - f(x; w_0, w_1)\big)x\end{aligned}$$

그러나

Randomly choose an initial solution, w_0^0 w_1^0

Repeat

$$g_0^t = \sum_{d=(\mathbf{x},y)\in Data} -2\big(y - f(x; w_0, w_1)\big)\Big|_{w_0=w_0^t, w_1=w_1^t}$$

$$g_1^t = \sum_{d=(\mathbf{x},y)\in Data} -2\big(y - f(x; w_0, w_1)\big)x\Big|_{w_0=w_0^t, w_1=w_1^t}$$

$$w_0^{t+1} = w_0^t - \eta g_0^t$$

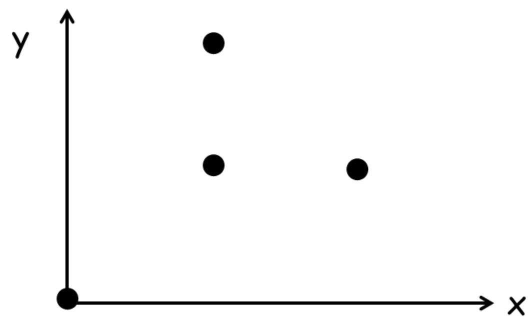
$$w_1^{t+1} = w_1^t - \eta g_1^t$$

Until stopping condition is satisfied

숙제 2

- 주어진 데이터에 가장 부합하는 모델을 찾아라

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



- 기본 모델로 2차 함수를 사용하자

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

  - w_0, w_1, w_2를 GDM으로 결정하라

연습

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

$f(x; w_0, w_1, w_2) = w_2x^2 + w_1x + w_0$

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

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

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

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

연습

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

$$\frac{\partial}{\partial w_0} E(w_0, w_1, w_2) = 4w_0 + 12w_2 + 8w_1 - 8$$

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

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

연습

$$\frac{\partial}{\partial w_0}E(w_0,w_1,w_2) = 4w_0 + 12w_2 + 8w_1 - 8$$

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

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

Randomly choose an initial solution, $w_0^0\ w_1^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \left.\frac{\partial E}{\partial w_0}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

$$w_1^{t+1} = w_1^t - \eta \left.\frac{\partial E}{\partial w_1}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

$$w_2^{t+1} = w_2^t - \eta \left.\frac{\partial E}{\partial w_2}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

Until stopping condition is satisfied

연습

Randomly choose an initial solution, $w_0^0\ w_1^0\ w_2^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta(4w_0^t + 12w_2^t + 8w_1^t - 8)$$

$$w_1^{t+1} = w_1^t - \eta(12w_1^t + 20w_2^t + 8w_0^t - 10)$$

$$w_2^{t+1} = w_2^t - \eta(36w_2^t + 20w_1^t + 12w_0^t - 14)$$

Until stopping condition is satisfied

연습

- 실제로 구현하는 방법

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

$$= \sum_{(\mathbf{x},y)\in Data} E_i(x_i, y_i; w_0, w_1, w_2)$$

where $E_i(x_i, y_i; w_0, w_1, w_2) = \left(y_i - f(x_i; w_0, w_1, w_2)\right)^2$

$$\frac{\partial}{\partial w_j} E(w_0, w_1, w_2) = \frac{\partial}{\partial w_j} \sum_{(\mathbf{x},y)\in Data} E_i(x_i, y_i; w_0, w_1, w_2)$$

$$= \sum_{(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_j} E_i(x_i, y_i; w_0, w_1, w_2)$$

연습

Randomly choose an initial solution, $w_0^0\ w_1^0\ w_2^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \left.\frac{\partial E}{\partial w_0}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

$$w_1^{t+1} = w_1^t - \eta \left.\frac{\partial E}{\partial w_1}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

$$w_2^{t+1} = w_2^t - \eta \left.\frac{\partial E}{\partial w_2}\right|_{w_0=w_0^t, w_1=w_1^t, w_2=w_2^t}$$

Until stopping condition is satisfied

연습

Randomly choose an initial solution, $w_0^0\ w_1^0\ w_2^0$

Repeat

$$w_0^{t+1} = w_0^t - \eta \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_0} E_d(x,y;w_0,w_1,w_2)\Big|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

$$w_1^{t+1} = w_1^t - \eta \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_1} E_d(x,y;w_0,w_1,w_2)\Big|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

$$w_2^{t+1} = w_2^t - \eta \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_2} E_d(x,y;w_0,w_1,w_2)\Big|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

Until stopping condition is satisfied

연습

Randomly choose an initial solution, $w_0^0\ w_1^0\ w_2^0$

Repeat

$$g_0^t = \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_0} E_d(x,y;w_0,w_1,w_2)\bigg|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

$$g_1^t = \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_1} E_d(x,y;w_0,w_1,w_2)\bigg|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

$$g_2^t = \sum_{d=(\mathbf{x},y)\in Data} \frac{\partial}{\partial w_2} E_d(x,y;w_0,w_1,w_2)\bigg|_{w_0=w_0^t,w_1=w_1^t,w_2=w_2^t}$$

$$w_0^{t+1} = w_0^t - \eta g_0^t$$

$$w_1^{t+1} = w_1^t - \eta g_1^t$$

$$w_2^{t+1} = w_2^t - \eta g_2^t$$

Until stopping condition is satisfied

연습

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

$$\frac{\partial}{\partial w_0} E_d(x, y; w_0, w_1, w_2) = \frac{\partial}{\partial w_0}\big(y - f(x; w_0, w_1, w_2)\big)^2$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)\frac{\partial}{\partial w_0}\ f(x; w_0, w_1, w_2)$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)$$

$$\frac{\partial}{\partial w_1} E_d(x, y; w_0, w_1, w_2) = \frac{\partial}{\partial w_1}\big(y - f(x; w_0, w_1, w_2)\big)^2$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)\frac{\partial}{\partial w_1}\ f(x; w_0, w_1, w_2)$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)x$$

$$\frac{\partial}{\partial w_2} E_d(x, y; w_0, w_1, w_2) = \frac{\partial}{\partial w_0}\big(y - f(x; w_0, w_1, w_2)\big)^2$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)\frac{\partial}{\partial w_2}\ f(x; w_0, w_1, w_2)$$
$$= -2\big(y - f(x; w_0, w_1, w_2)\big)x^2$$

Question and Answer