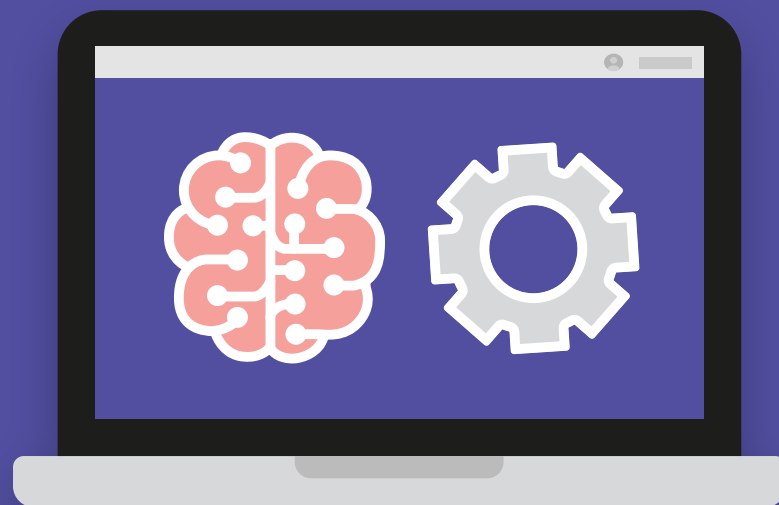


```
/* elice */
```

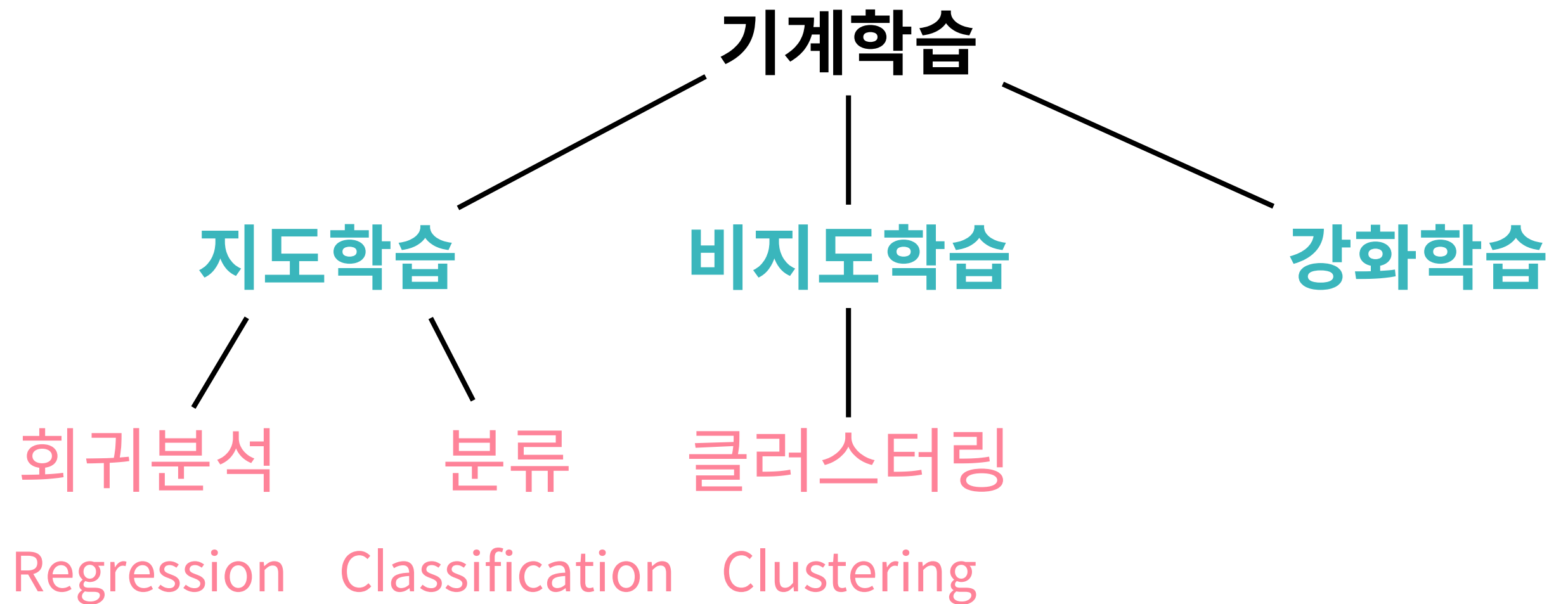
인공지능/머신러닝 기초

Module 3: Linear Regression



2월 7일 ~ 3월 7일

기계학습



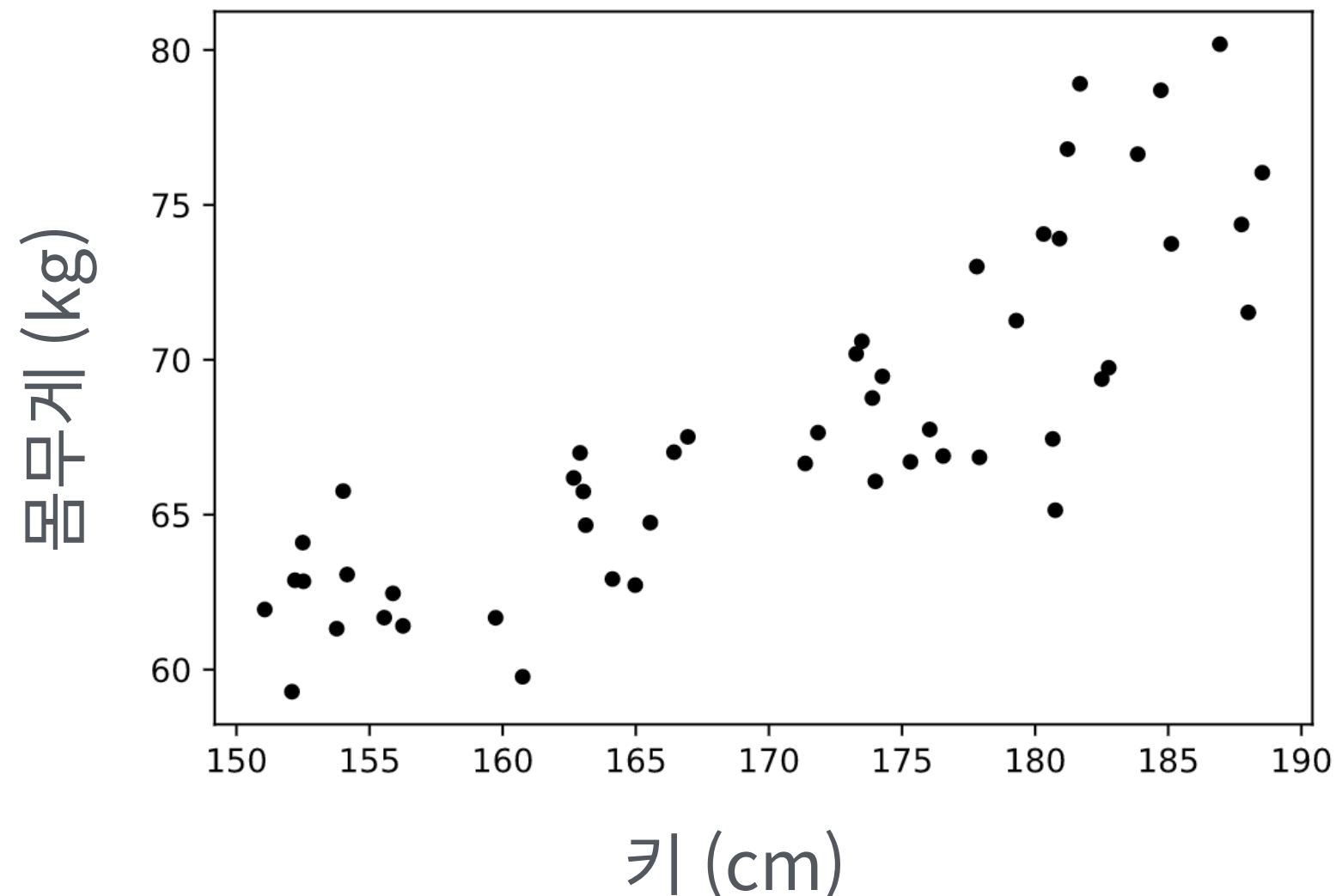
단순선형회귀분석

Simple Linear Regression

회귀분석이란?

대학 운동부 학생들의 신체검사 자료

신입생 **A**가 들어왔다.(키는 175cm이다) 예상 몸무게는 얼마인가?

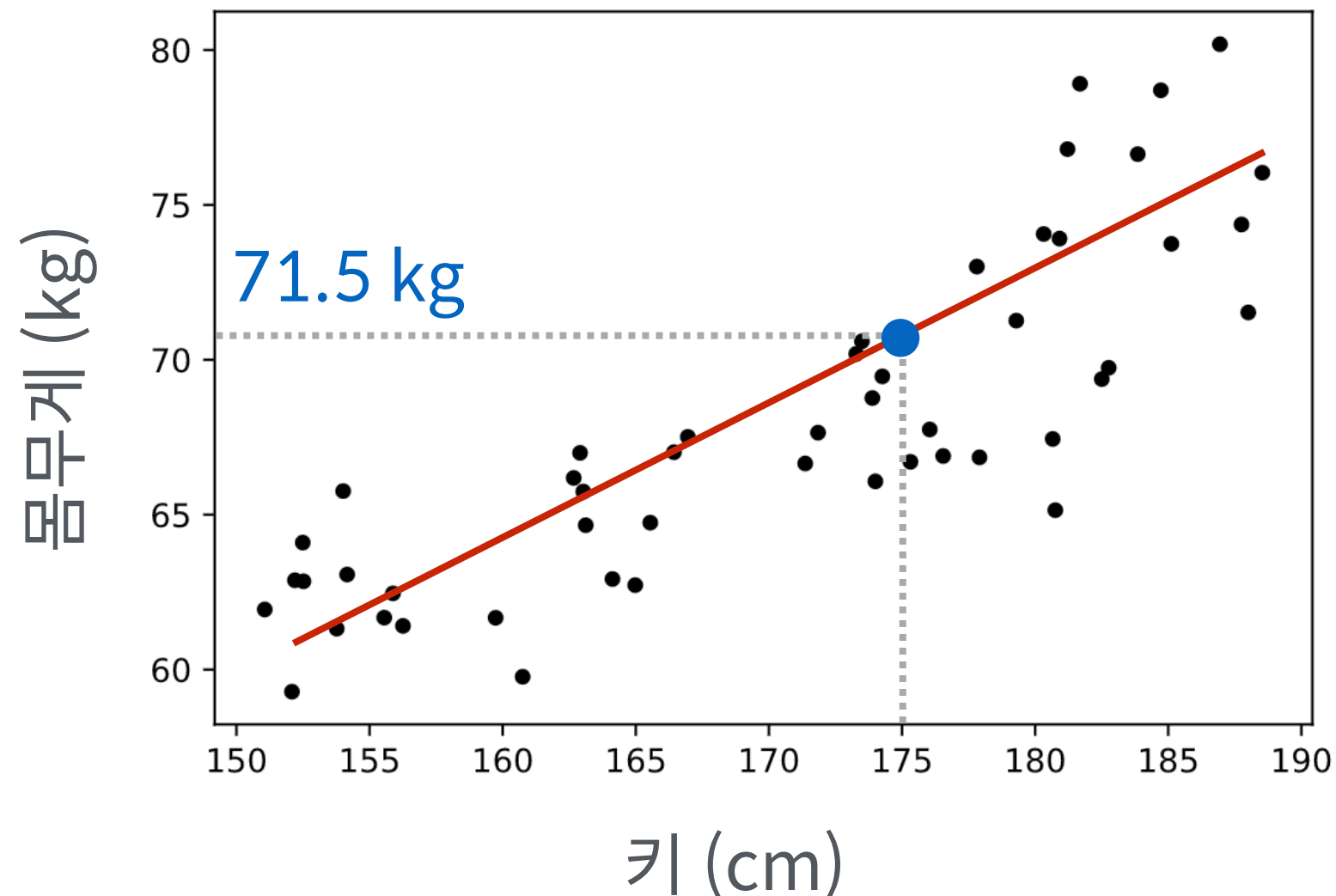


/* elice */

회귀분석이란?

대학 운동부 학생들의 신체검사 자료

신입생 **A**가 들어왔다.(키는 175cm이다) 예상 몸무게는 얼마인가?



/* elice */

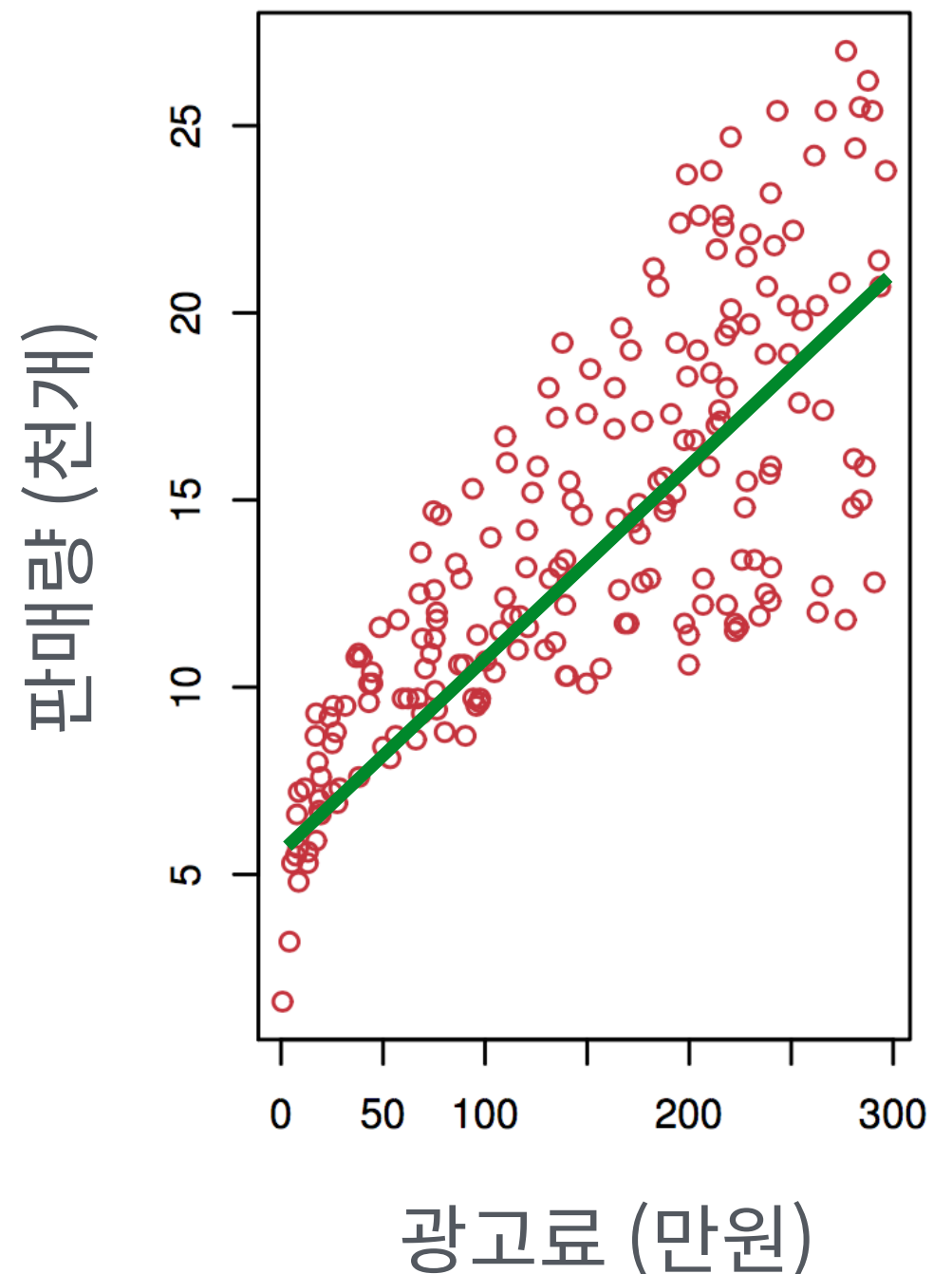
회귀분석법

당신은 엘리스의
데이터 사이언티스트다.

데이터: 광고 분석과 판매량

목표: FB 광고에 얼마를 투자하면...
상품이 얼마나 팔릴까?

방법: 데이터를 가장 잘 설명하는
어떤 선을 하나 찾는다.



변수 표기

N : 데이터의 개수

X : Input; 데이터/Feature

“광고료”

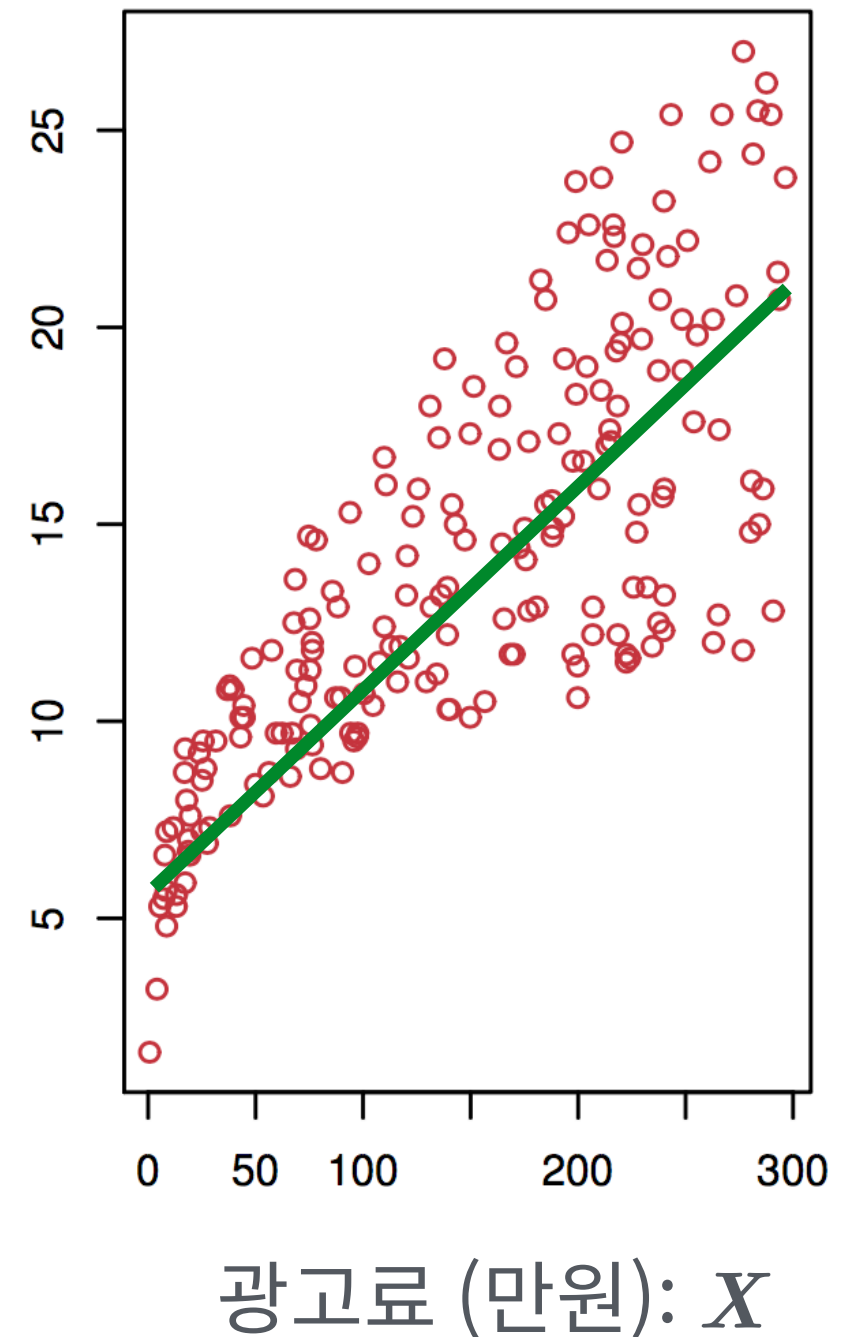
Y : Output; 해답/응답

“판매량”

$(x^{(i)}, y^{(i)})$: i 번째 데이터

X	Y
230.1	22.1
44.5	10.4
17.2	9.3
151.5	18.5
180.8	12.9
8.7	7.2
57.5	11.8
\vdots	\vdots

판매량 (천개): Y



문제 정의

X Y

데이터: N개의 FB 광고 예산과 판매량

목표: 광고에 얼마를 투자했을 때 얼마나 팔릴까?

X	Y
230.1	22.1
44.5	10.4
17.2	9.3
151.5	18.5
180.8	12.9
8.7	7.2
57.5	11.8
\vdots	\vdots

광고 예산 → 학습된 모델 → 판매량

문제 정의

X Y

데이터: N개의 FB 광고 예산과 판매량

목표: 광고에 얼마를 투자했을 때 얼마나 팔릴까?

광고 예산 → 학습된 모델 → 판매량

가정: TV 광고 예산과 판매량은 **선형적** 관계를 가진다

$$Y \sim \beta_0 X + \beta_1$$

문제: 어떤 β_0, β_1 이 좋은 것인가?

X	Y
230.1	22.1
44.5	10.4
17.2	9.3
151.5	18.5
180.8	12.9
8.7	7.2
57.5	11.8
\vdots	\vdots

실습 1: 기울기와 절편



`/* elice */`

모델의 학습 목표

$$Y \sim \beta_0 X + \beta_1$$

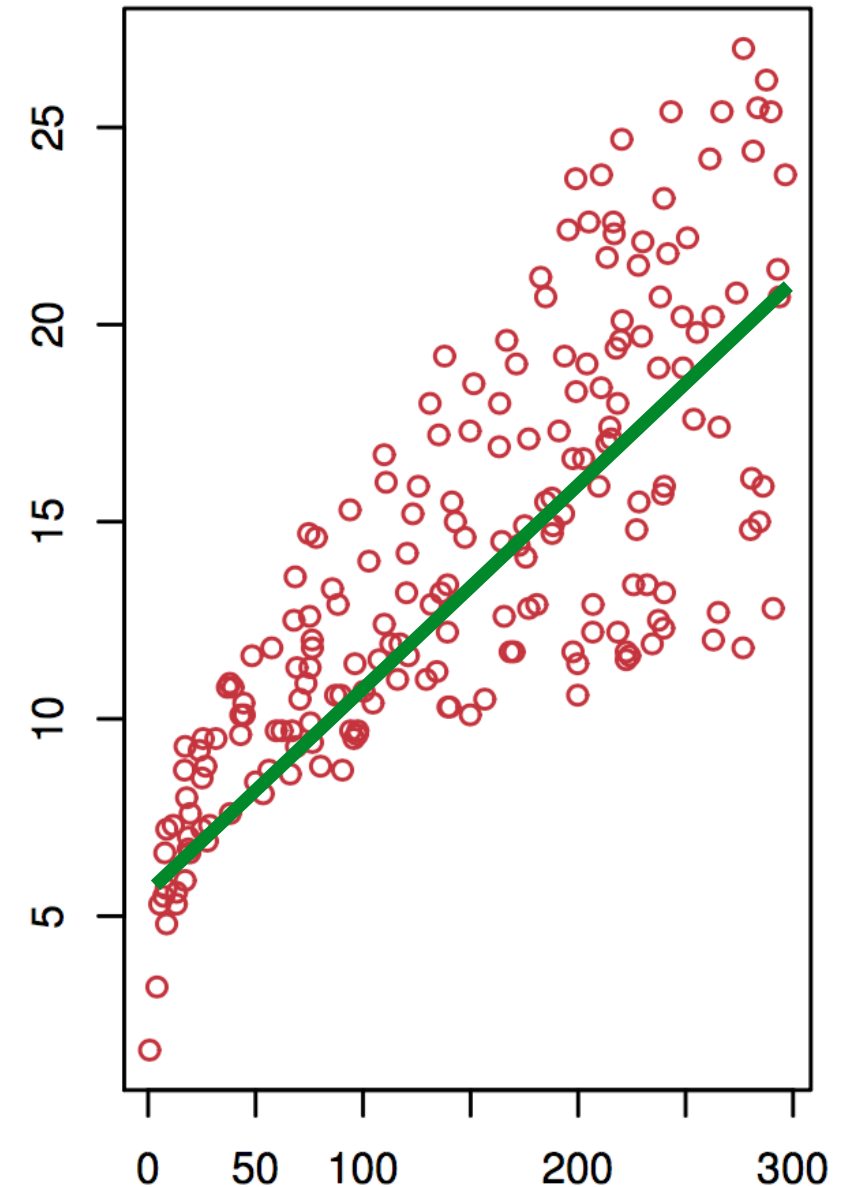
아이디어:

완벽한 예측은 불가능하다.

각 데이터 $(x^{(i)}, y^{(i)})$ 의 실제 값과

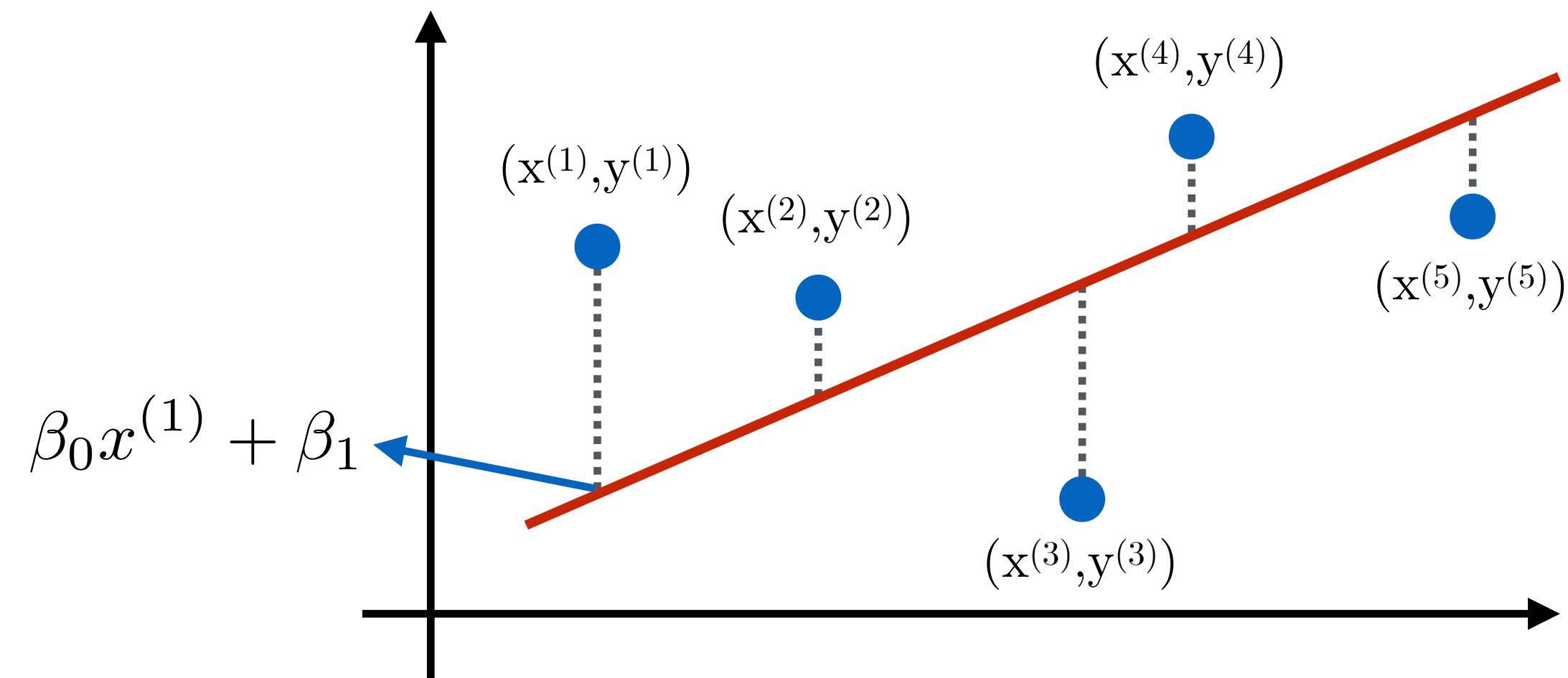
모델이 예측하는 값을 최소한으로 하자!

어떻게?



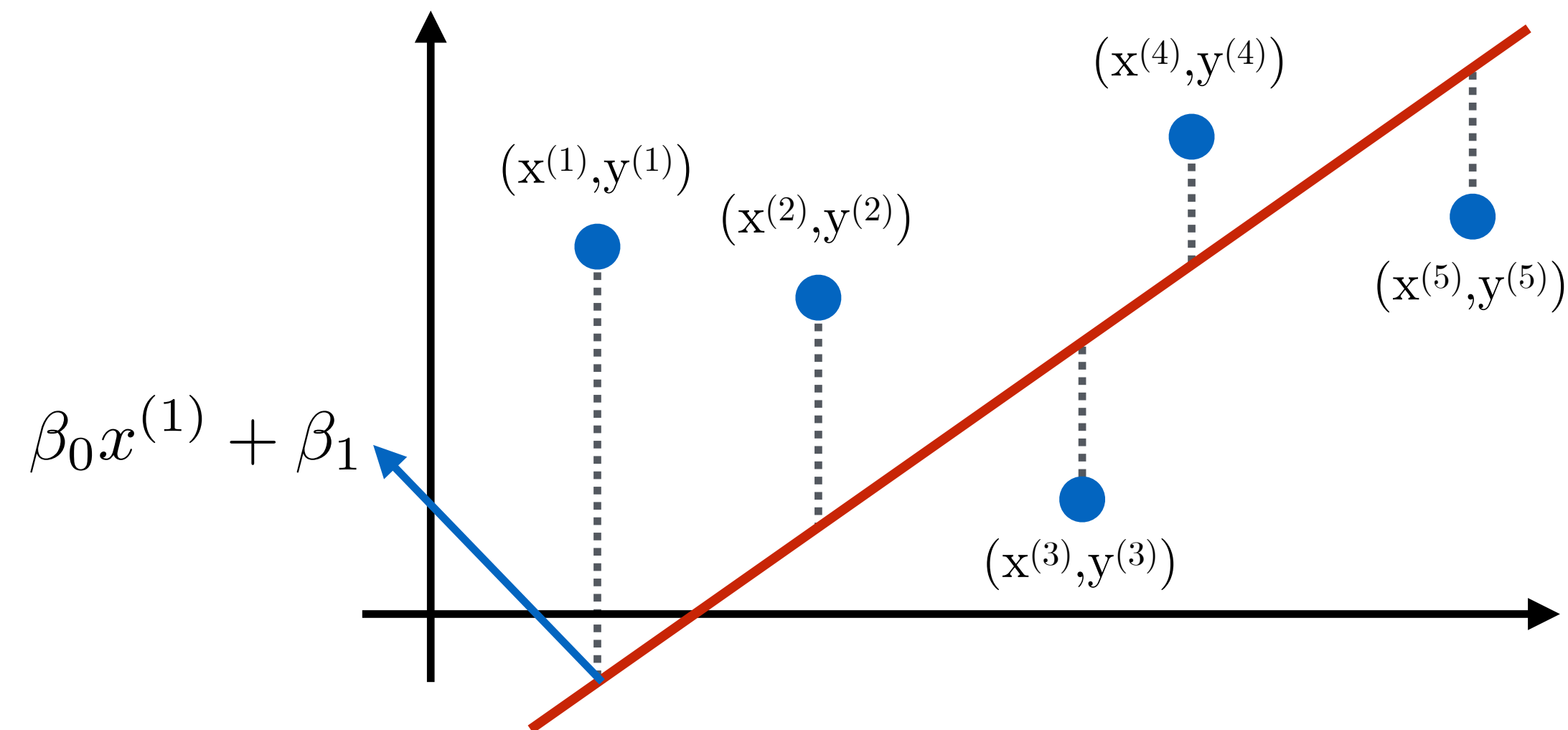
모델의 학습 목표

$$Y \sim \beta_0 X + \beta_1$$



모델의 학습 목표

$$Y \sim \beta_0 X + \beta_1$$



모델의 학습 목표

$$Y \sim \beta_0 X + \beta_1$$

i번째 데이터 $(x^{(i)}, y^{(i)})$ 에 대해:

실제 값: $y^{(i)}$

예측 값: $\beta_0 x^{(i)} + \beta_1$

차이:

전체 모델의 차이:

모델의 학습 목표

$$Y \sim \beta_0 X + \beta_1$$

i번째 데이터 $(x^{(i)}, y^{(i)})$ 에 대해:

실제 값: $y^{(i)}$

예측 값: $\beta_0 x^{(i)} + \beta_1$

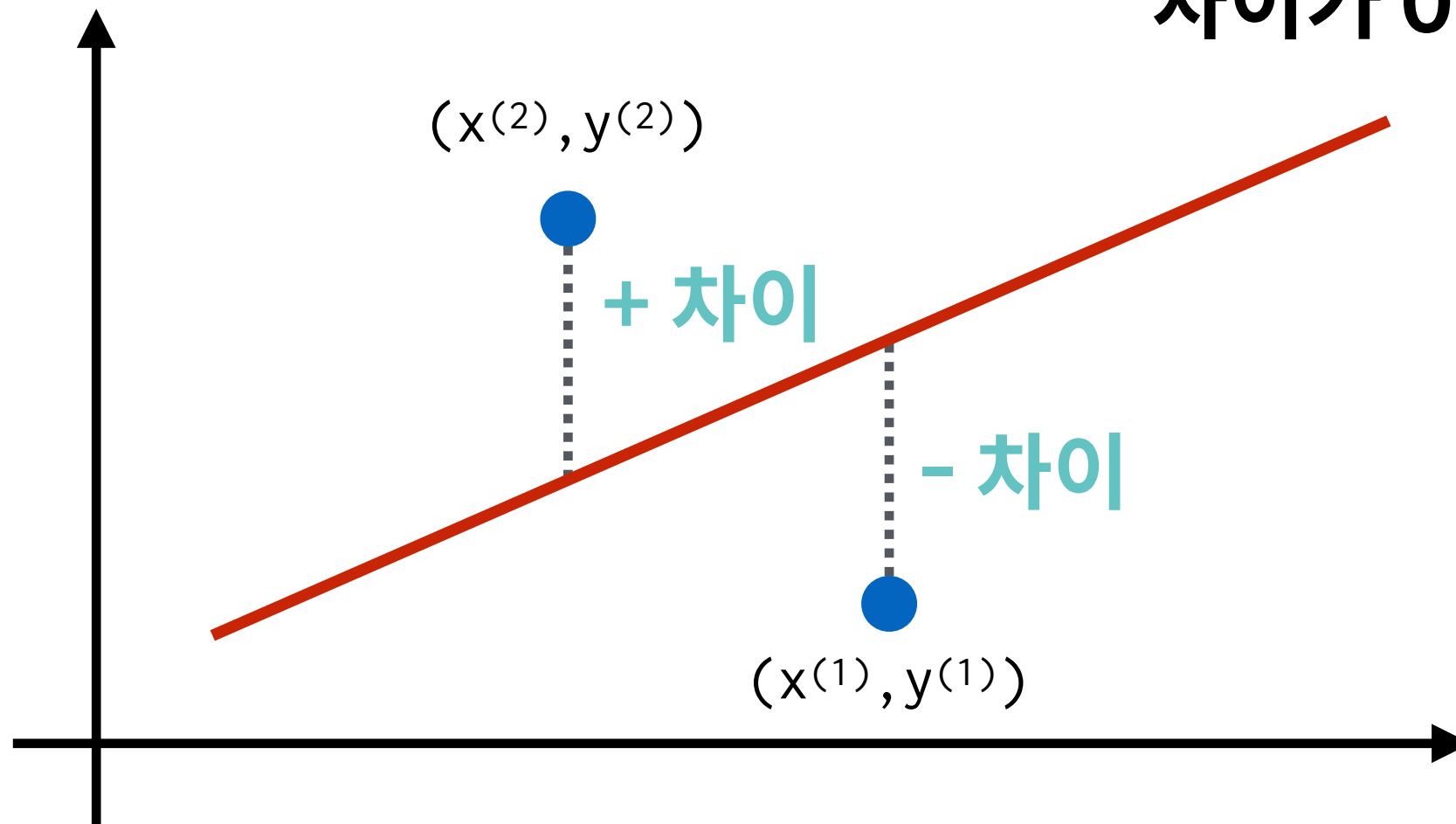
차이: $y^{(i)} - (\beta_0 x^{(i)} + \beta_1)$

전체 모델의 차이: $\sum_i^N y^{(i)} - (\beta_0 x^{(i)} + \beta_1)$

반례

$$\sum_i^N y^{(i)} - (\beta_0 x^{(i)} + \beta_1) = 0$$

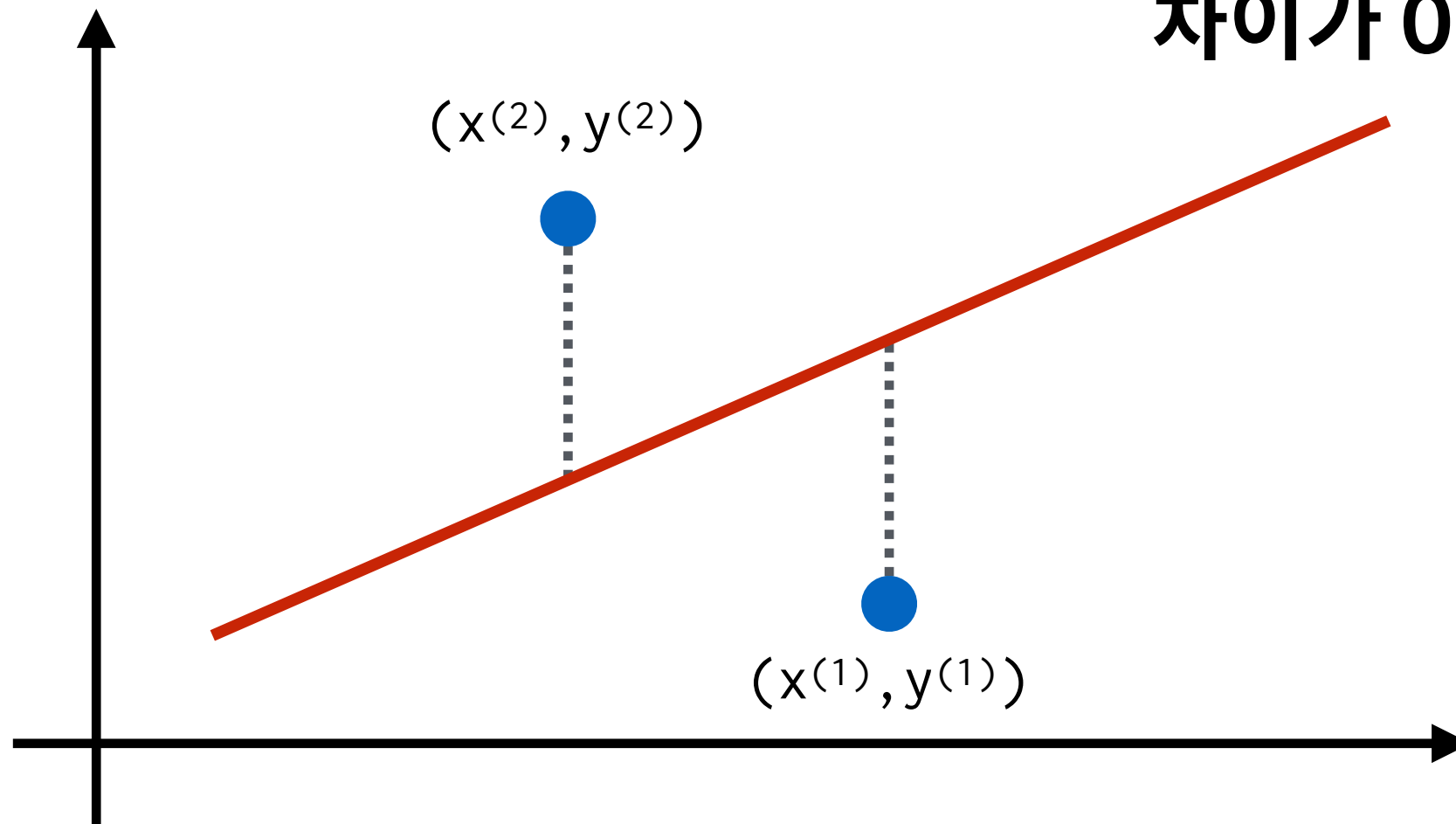
차이가 0이라고??



해결

$$\sum_i^N (y^{(i)} - (\beta_0 x^{(i)} + \beta_1))^2 > 0$$

차이가 0 이상이 됨



문제 재정의

$$Y \sim \beta_0 X + \beta_1$$

전체 모델의 차이: $\sum_i^N (y^{(i)} - (\beta_0 x^{(i)} + \beta_1))^2$
“Loss function”

이 차이를 최소로 하는 β_0, β_1 을 구하자.

$$\arg \min_{\beta_0, \beta_1} (y^{(i)} - (\beta_0 x^{(i)} + \beta_1))^2$$

실습 2: Loss function



`/* elice */`

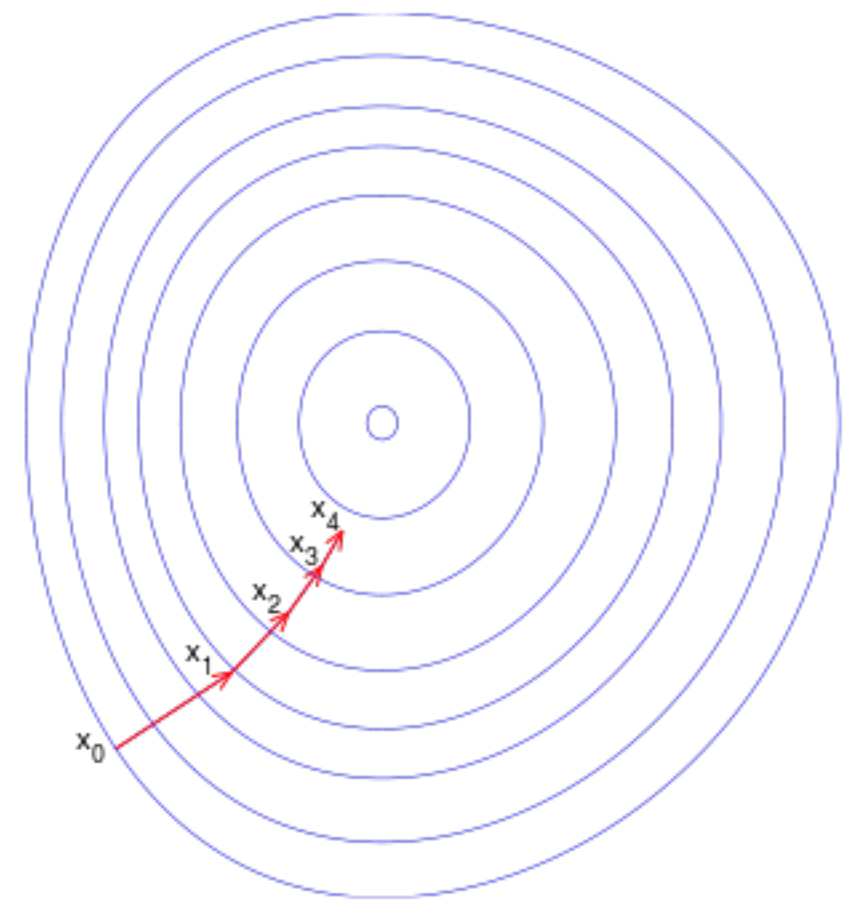
산 정상 오르기

산 정상이 되는 지점을 찾고 싶다.

아무 곳에서나 시작했을 때, 가장 정상을 빠르게 찾아가는 방법은?

가정

- **정상의 위치**는 알 수 없다.
- 현재 **나의 위치와 높이**를 알 수 있다.
- 내 위치에서 일정 수준 이동할 수 있다.



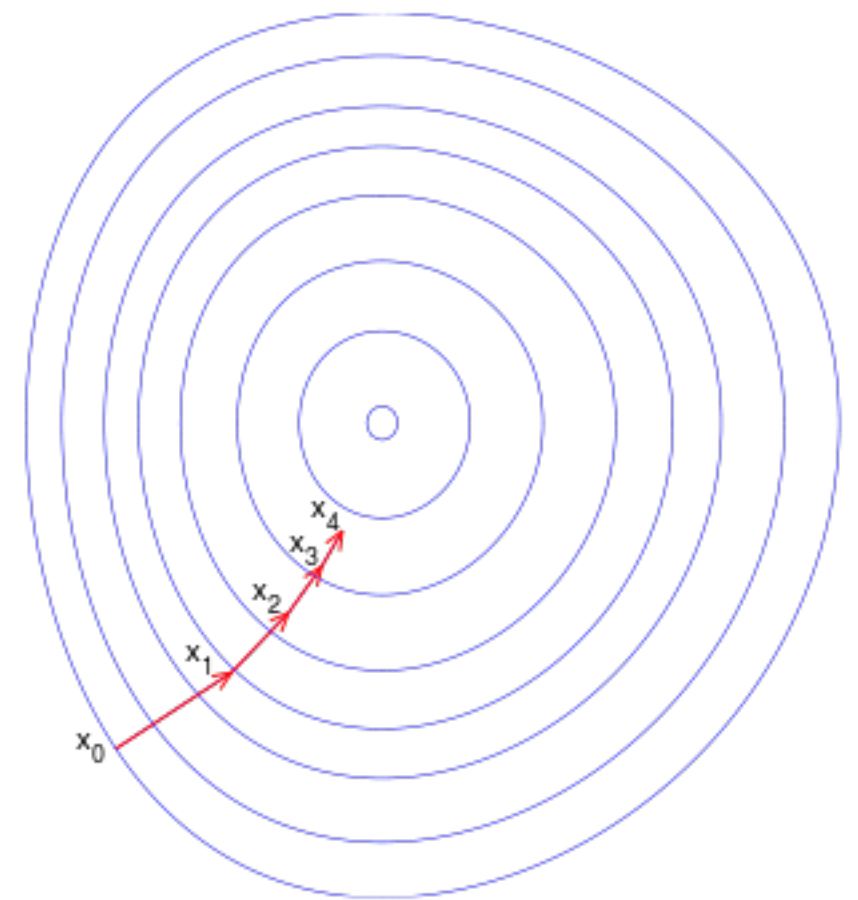
산 정상 오르기

산 정상이 되는 지점을 찾고 싶다.

아무 곳에서나 시작했을 때, 가장 정상을 빠르게 찾아가는 방법은?

방법

- 현재 위치에서 **가장 경사가 높은 쪽**을 찾는다.
- 오르막 방향으로 일정 수준 이동한다.
- 더 이상 높이의 변화가 없을 때까지 반복!



거꾸로 된 산을 내려가기

데이터를 가장 잘 설명하는 β_0, β_1 을 구하자

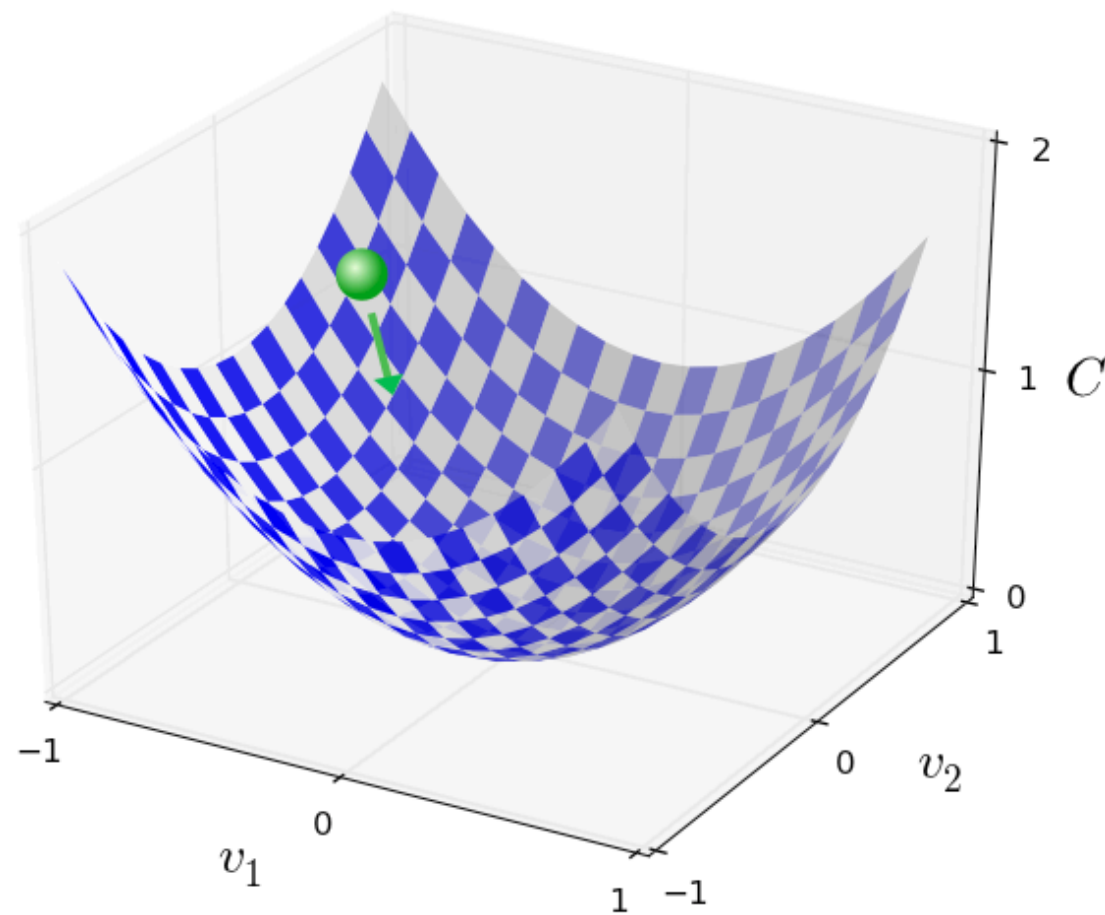
= 예측 값과 실제 값의 차이를 최소로 만드는 값을 구하자

= Loss function을 최소로 만드는 β_0, β_1 을 구하자

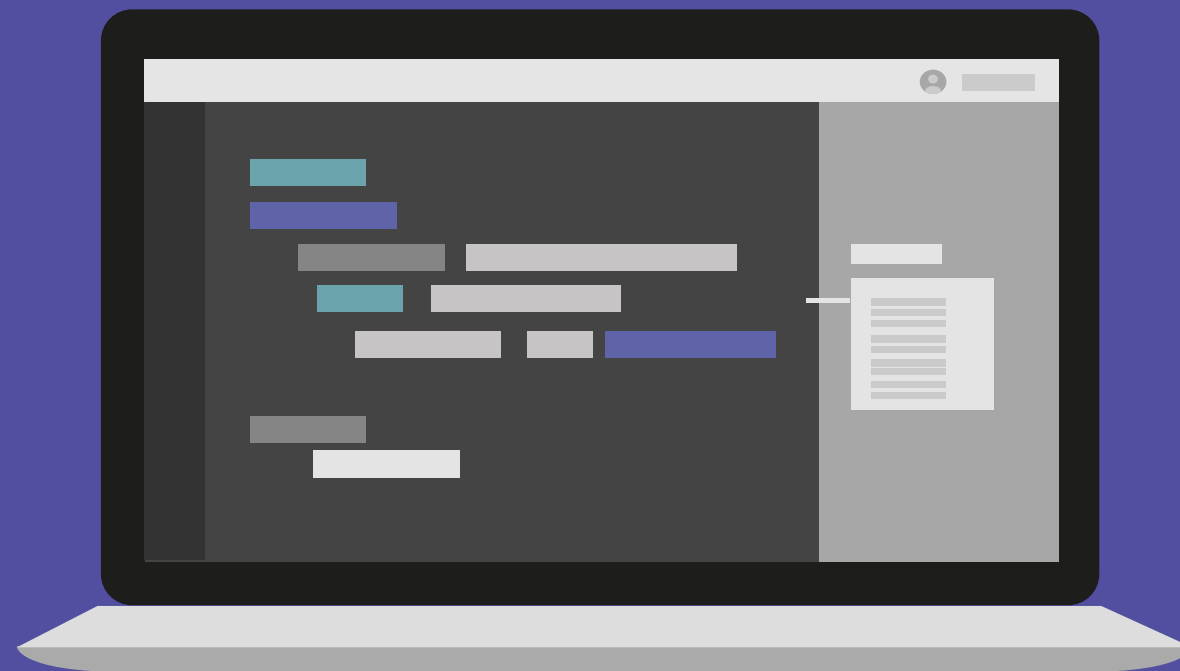
$$L(\beta_0, \beta_1) = \sum_i^N (y^{(i)} - (\beta_0 x^{(i)} + \beta_1))^2$$

거꾸로 된 산을 내려가기

$$L(\beta_0, \beta_1) = \sum_i^N (y^{(i)} - (\beta_0 x^{(i)} + \beta_1))^2$$



실습 3: Scikit-learn을 이용한 회귀분석



`/* elice */`

`/* elice */`

문의 및 연락처

academy.elice.io

contact@elice.io

facebook.com/elice.io

blog.naver.com/elicer