

# Lab 1

## Linear Regression Using Gradient Descent Method

---

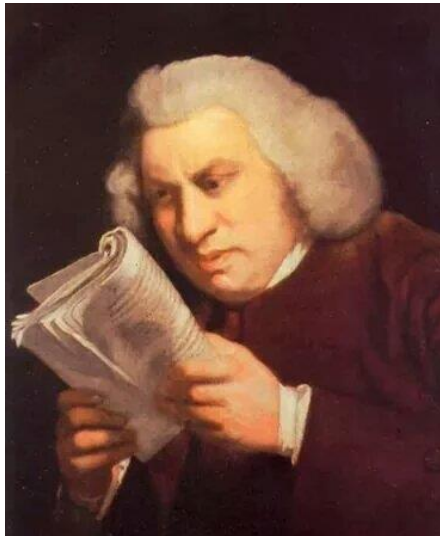
권용혜 ([robotmany@kw.ac.kr](mailto:robotmany@kw.ac.kr))

남건욱 ([ngotic@kw.ac.kr](mailto:ngotic@kw.ac.kr))

## 실습 과제 (1)

- 팽운대학교의 남건육 조교는 만사가 다 귀찮다. 어느 날, 남건육 조교는 수강생이 59명에 달하는 과목의 기말고사 채점을 맡게 됐다. 남건육 조교는 수강생 30명의 시험지를 채점하다가 귀찮음을 느끼며 산책을 나갔다. 사과 나무에 앉아 시간을 허비하던 중 떨어지는 사과를 보며 생각을 했다.

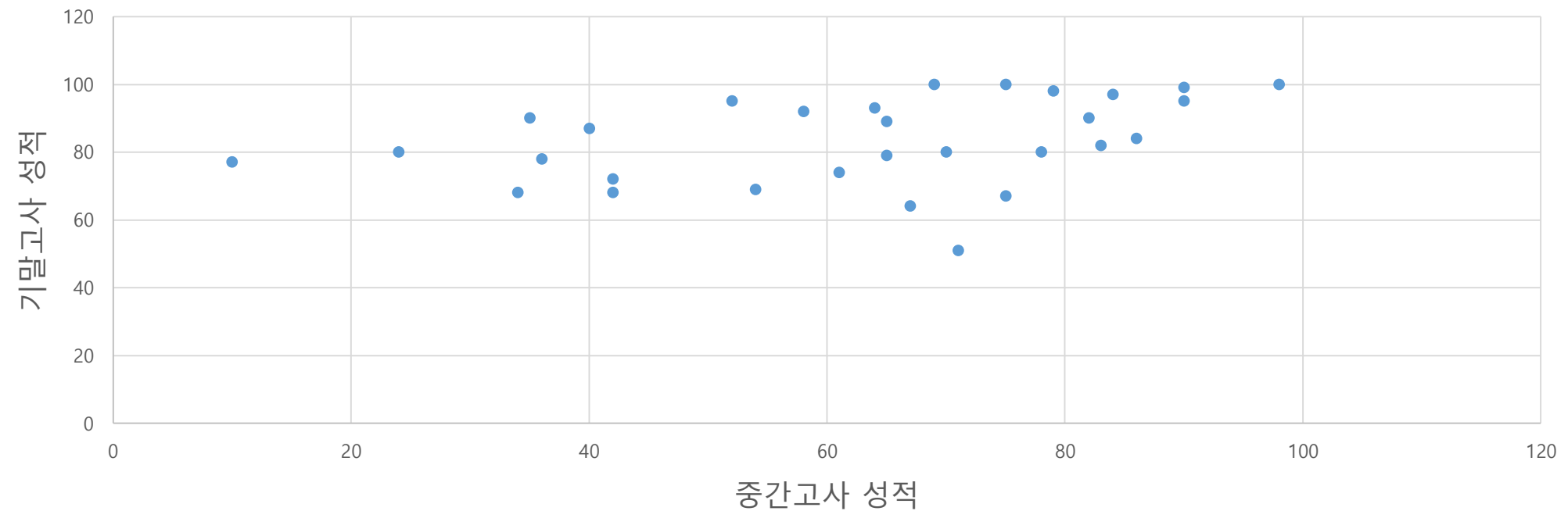
**“중간 고사 성적과 기말 고사 성적은 선형적인 관계가 있지 않을까?”**



Linear Regression 모델을 생성하고 지금까지 채점한 수강생 30명의 기말고사 성적과 중간고사 성적을 이용하여 나머지 수강생 29명의 기말고사 성적은 중간고사 성적을 이용하여 예측해보자!

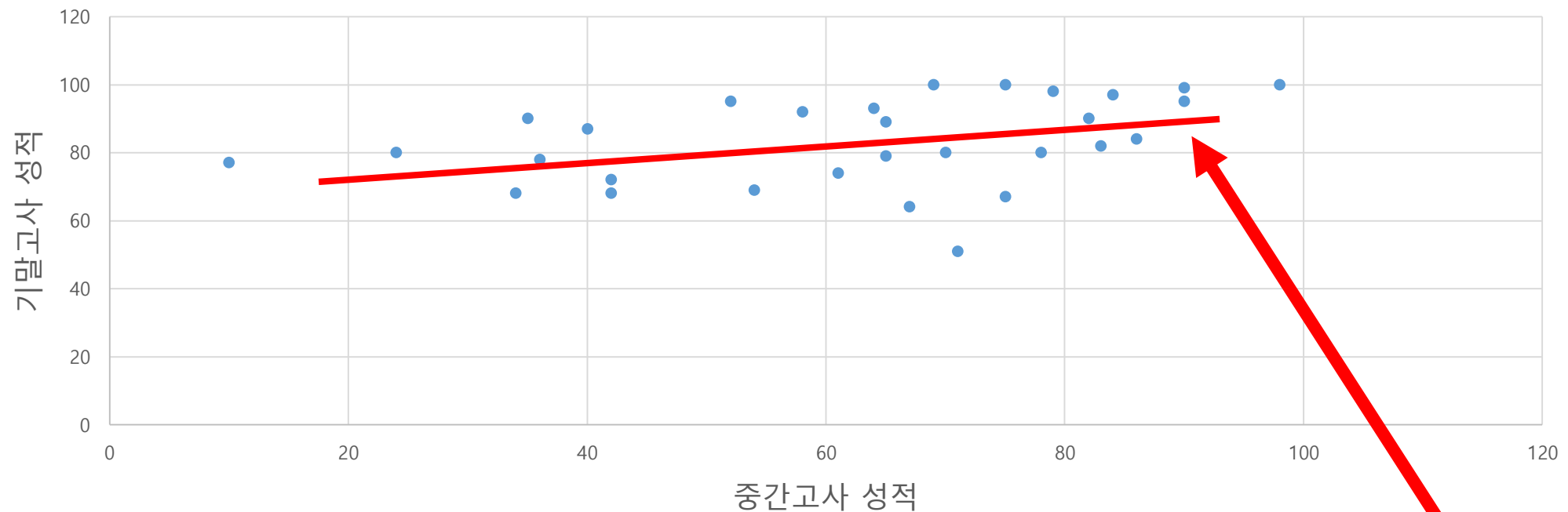
## 실습 과제 (2)

남건욱 조교가 채점한 수강생 30명의 중간고사 성적과 기말고사 성적



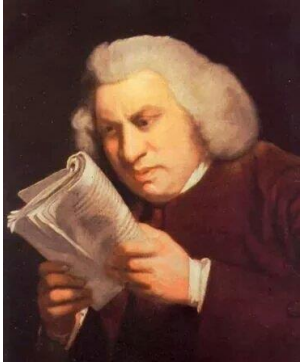
## 실습 과제 (3)

남건욱 조교가 채점한 수강생 30명의 중간고사 성적과 기말고사 성적



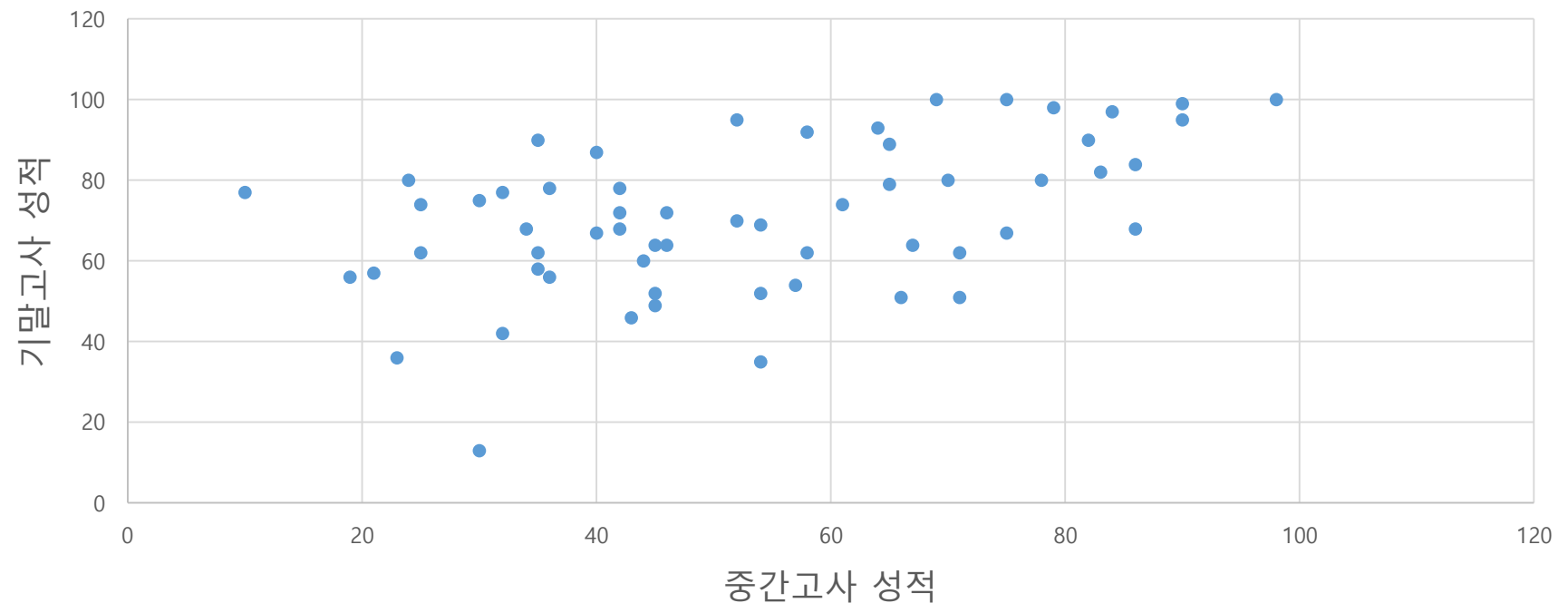
The linear regression model  
we want to fit!

## 실습 과제 (4)



나머지 29 명의 학생도 30명의 수강생들의 정보를 이용하여 생성한 linear regression 모델로 잘 예측됐으면 좋겠다!

수강생 59명의 중간고사 성적과 기말고사 성적



# 실습 과제 (5)

---

## ■ Materials

- train.txt
  - 남건욱 조교가 채점한 수강생 30명의 중간/기말고사 성적이 나열된 파일
- test.txt
  - 남건욱 조교가 채점하지 않은 나머지 수강생 29명의 중간/기말고사 성적이 나열된 파일
- [https://github.com/developer0hye/ML\\_Lab/tree/master/Lab01-LinearRegression](https://github.com/developer0hye/ML_Lab/tree/master/Lab01-LinearRegression)

## ■ Conditions

- C++ 이용
- Gradient Descent 기법 이용하여 linear regression 모델의 파라미터 학습
  - Cost function 은 MSE(Mean Squared Error)

# 실습 과제 (6)

---

## ■ Criteria

- train.txt 파일의 중간/기말고사 성적으로 학습된 linear regression 모델을 이용하여 test.txt 파일의 중간고사 성적을 입력 받아 기말고사 성적을 예측
- 예측된 기말고사 성적과 실제 기말고사 성적의 차이가 10 이하인 경우, 정답으로 판단
- 정답의 수를 기준으로 성능 평가

# 실습 과제 (7)

## ■ Linear Regression Model

–  $Y = mX + c$

- $X$ : 중간고사 성적
- $Y$ : 중간고사 성적을 통해 예측되는 기말고사 성적

## ■ Cost function

$$E = \frac{1}{n} \sum_{i=0}^n (y_i - (mx_i + c))^2$$

$$D_m = \frac{1}{n} \sum_{i=0}^n 2(y_i - (mx_i + c))(-x_i)$$

$$D_m = \frac{-2}{n} \sum_{i=0}^n x_i(y_i - \bar{y}_i)$$

$$D_c = \frac{-2}{n} \sum_{i=0}^n (y_i - \bar{y}_i)$$

Gradient Descent

$$m = m - L \times D_m$$

$$c = c - L \times D_c$$

E: Cost function

$x_i$ :  $i$ 번째 학생의 중간고사 성적

$y_i$ :  $i$ 번째 학생의 기말고사 성적

$mx_i + c$ :  $i$ 번째 학생의 중간고사 성적을 통해 예측되는 기말고사 성적

$D_m$ :  $E$  를  $m$  으로 미분한 식

$D_c$ :  $E$  를  $c$  로 미분한 식

$L$ : Learning Rate



## 실습 과제 (8)

---

- 결국 남건욱 조교는 귀찮음을 극복하고 성실히 채점을 마쳤다고 한다.