2조 경사하강법

17학번 - 강신현
17학번 - 구범준
19학번 - 오승현
20학번 - 이은지

SMARCLE

목차

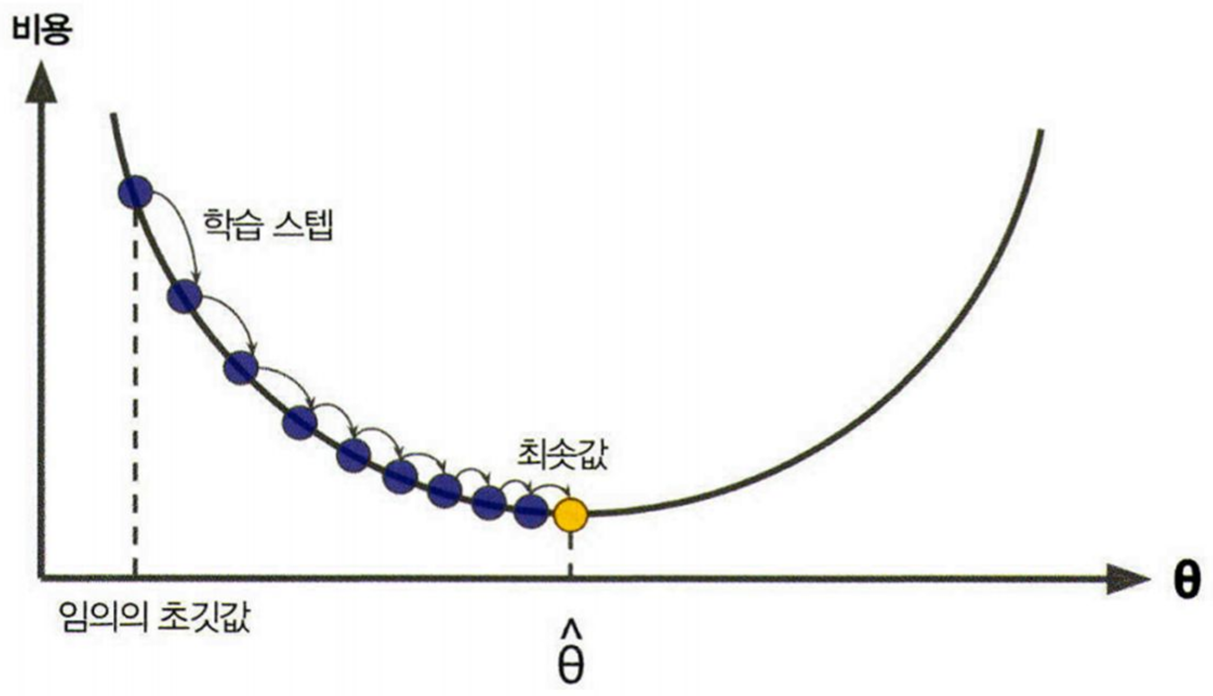
1. 경사하강법이란?

2. 경사하강법 실습

3. 다중 선형 회귀란?

4. 다중 선형 회귀 실습

경사하강법



경사하강법의 정의

“기울기가 0인 최소값을 찾는 방법 "

SMARCLE

경사 하강법의 공식이란?

1. 특정 지점에서 미분을 구한다.

2. 학습률에 따라 곱한 값에 따른 이동한 부분에 미분을 구한다.

3. 0이 될때 까지 반복한다.

경사하강법의 정의

"기울기가 0인 최소값을 찾는 방법 "

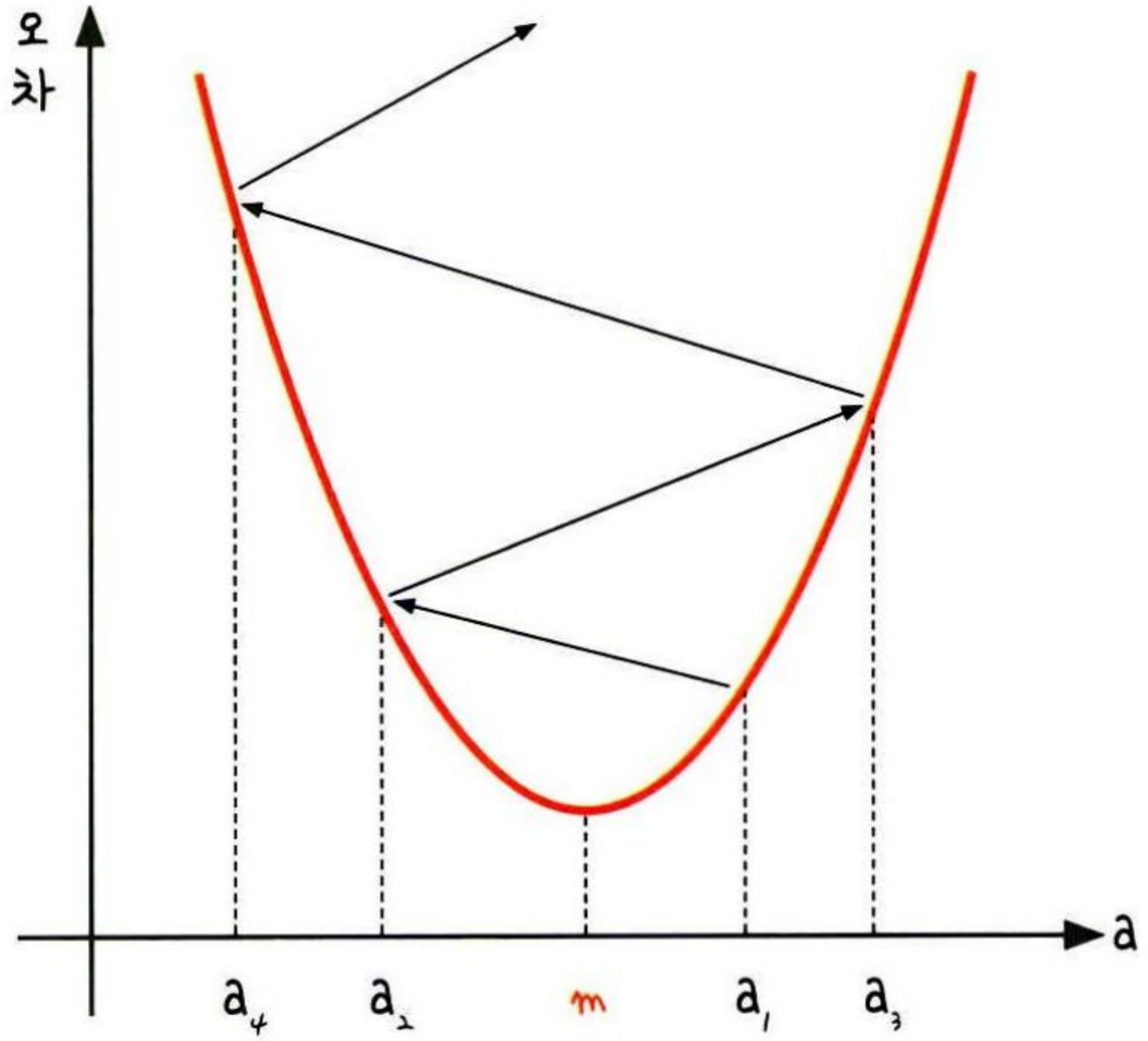
"학습률"

학습률이란?

데이터를 학습할 때, 한번의 학습에서 얼만큼 학습하는 양

➡ **"오버슈팅"**

➡ **"극소 최저치"**


오차
a
a4
a2
m
a1
a3

경사 하강법의 공식이란?

$\frac{1}{n}\sum(y_i - \hat{y}_i)^2 \Rightarrow \frac{1}{n}\sum(y_i - (ax_i + b))^2$

$$a\text{로 편미분 한 결과} = \frac{2}{n}\sum(ax_i + b - y_i)x_i$$

$$b\text{로 편미분 한 결과} = \frac{2}{n}\sum(ax_i + b - y_i)$$

SMARCLE

유도과정

$$\frac{1}{n}\sum\left(y_i-(ax_i+b)\right)^2$$

$$\int \frac{d}{dx}F(g(x))dx = \int f(g(x))g'(x)dx$$

TIP

a로 편미분한 결과 유도 과정

$$\begin{aligned}\frac{\partial}{\partial a}MSE(a,b) &= \frac{1}{n}\sum\left[(ax_i+b-y_i)^2\right]' \\ &= \frac{2}{n}\sum(ax_i+b-y_i)\left[(ax_i+b-y_i)\right]' \\ &= \frac{2}{n}\sum(ax_i+b-y_i)x_i\end{aligned}$$

b로 편미분한 결과 유도 과정

$$\begin{aligned}\frac{\partial}{\partial b}MSE(a,b) &= \frac{1}{n}\sum\left[(ax_i+b-y_i)^2\right]' \\ &= \frac{2}{n}\sum(ax_i+b-y_i)\left[(ax_i+b-y_i)\right]' \\ &= \frac{2}{n}\sum(ax_i+b-y_i)\end{aligned}$$

SMARCLE

https://www.google.com/intl/ko_kr/slides/about/

코딩으로 확인하기

코딩으로 나타내면?

```
y_pred = a * x_data + b    # 오차 함수인 y = ax + b를 정의한 부분
error = y_data - y_pred    # 실제값 - 예측값, 즉 오차를 구하는 식


# 평균 제곱 오차를 a로 미분한 결과
a_diff = -(2 / len(x_data)) * sum(x_data * (error))
# 평균 제곱 오차를 b로 미분한 결과
b_diff = -(2 / len(x_data)) * sum(error)
```

```
a = a - lr * a_diff    # 미분 결과에 학습률을 곱한 후 기존의 a값을 업데이트
b = b - lr * b_diff    # 미분 결과에 학습률을 곱한 후 기존의 b값을 업데이트
```

쉬는시간

< 다중 선형 회귀 >

공부한 시간=2, 실제 점수=81, 예측 점수=82

공부한 시간=4, 실제 점수=93, 예측 점수=88

공부한 시간=6, 실제 점수=91, 예측 점수=94

공부한 시간=8, 실제 점수=97, 예측 점수=100

mse 최종값: 11.0

차이가 생기는 이유?

공부한 시간 이외의 다른 요소가
성적에 영향

다중선형회귀

더 정확한 예측 -> 추가정보 입력

정보 추가 -> 새로운 예측값

변수의 개수 추가 -> **다중선형회귀**

표 4-1
공부한 시간, 과외 수업 횟수에 따른 성적 데이터

공부한 시간(x_1)	2	4	6	8
과외 수업 횟수(x_2)	**0**	**4**	**2**	**3**
성적(y)	81	93	91	97

$$y = a_1x_1 + a_2x_2 + b$$

SMARCLE

더 더 더 정확한 예측 ➡ X : 1000개, a : 1000개 ➡ H(x) = a1x1 + a2x2 + ⋯ + a1000x1000 + b ?

?

문제점 : 변수만 2000개 선언

비효율적

내적 이용

1×7 + 2×9 + 3×11 = 58

내적 이용

w:기울기, x:변수

Quiz 1 (x1)	Quiz 2 (x2)	Quiz 3 (x3)	Final (y)
73	80	75	152
93	88	93	185
89	91	80	180
96	98	100	196
73	66	70	142

H(X) = w1x1 + w2x2 + w3x3

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} x_1w_1 + x_2w_2 + x_3w_3 \end{pmatrix}$$

x의 개수가 3개였음에도
이제는 X와 W라는 두 개의 변수로 표현

X W = H(X)

H(X) = XW

SMARCLE

행렬 이용

Quiz 1 (x1)	Quiz 2 (x2)	Quiz 3 (x3)	Final (y)
73	80	75	152
93	88	93	185
89	91	80	180
96	98	100	196
73	66	70	142

H(X) = w1x1 + w2x2 + w3x3

변수가 아무리 많아져도 문제 없다!

$$\underbrace{\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}}_{W} = \underbrace{\begin{pmatrix} x_{11}w_1 + x_{12}w_2 + x_{13}w_3 \\ x_{21}w_1 + x_{22}w_2 + x_{23}w_3 \\ x_{31}w_1 + x_{32}w_2 + x_{33}w_3 \\ x_{41}w_1 + x_{42}w_2 + x_{43}w_3 \\ x_{51}w_1 + x_{52}w_2 + x_{53}w_3 \end{pmatrix}}_{H(X)}$$

H(X)=XW

SMARCLE

https://www.google.com/intl/ko_kr/sl
ides/about/

코딩으로 확인하기