Machine Learning & Deep Learning

Lecture 16

Gradient Decent (경사 하강법)

Multiple Linear Regression (다중 선형 회귀)

김희숙 (H.S.Kim)

1. 경사 하강법 (Gradient Decent)

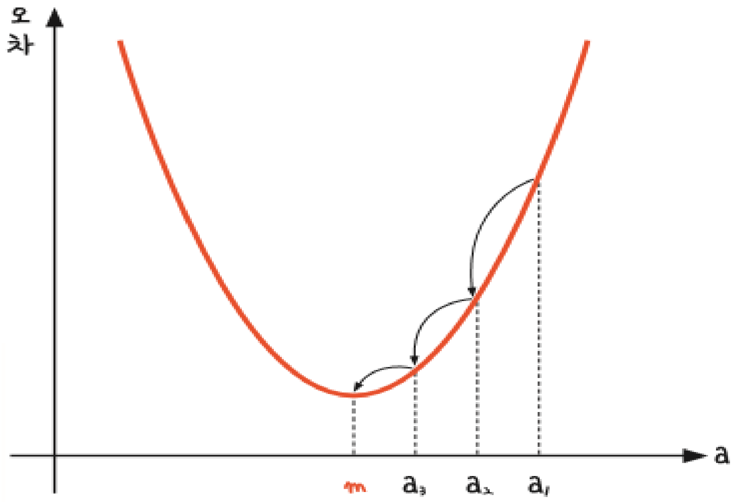
2. 학습률 (Learning Rate)

3. 코딩으로 확인하는 경사 하강법

4. 다중 선형 회귀 (Multiple Linear Regression)

5. 코딩으로 확인하는 다중 선형 회귀

- 기울기 a를 무한대로 키우면 오차도 무한대로 커지고, a를 무한대로 작게 해도 역시 오차가 무한대로 커지는 이러한 관계는 “이차 함수 그래프”로 표현할 수 있음



기울기 a와 오차와의 관계 : 적절한 기울기를 찾았을 때 오차가 최소화된다.

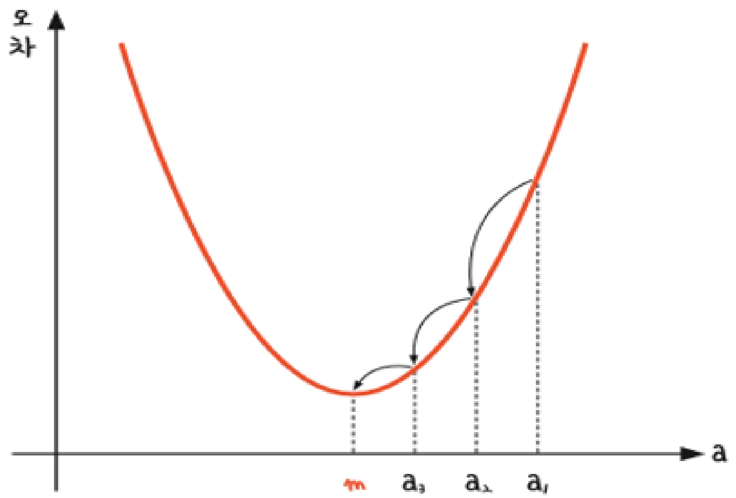
1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

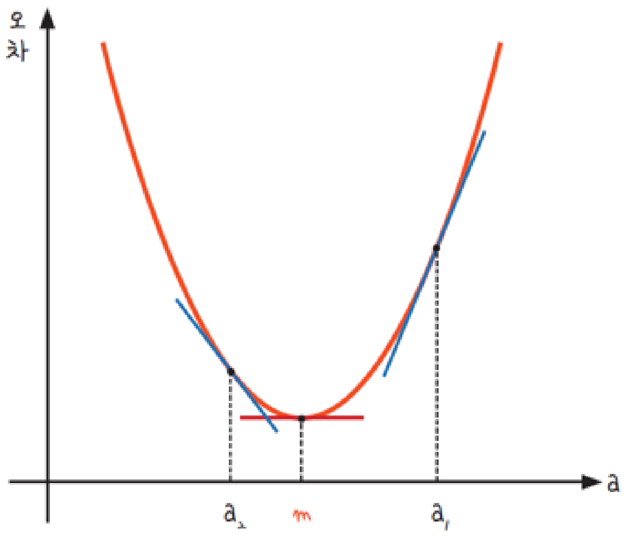
5. 코딩으로 확인하는
다중 선형 회귀



- 컴퓨터를 이용해 m 의 값을 구하려면 임의의 한 점(a_1)을 찍고 이 점을 m 에 가까운 쪽으로 점점 이동($a_1 \rightarrow a_2 \rightarrow a_3$)시키는 과정이 필요함
- **경사 하강법(Gradient Descent)**
  그래프에서 오차를 비교하여 가장 작은 방향으로 이동시키는 방법이 있는데 바로 미분 기울기를 이용

1. 경사 하강법 (Gradient Decent)
2. 학습률 (Learning Rate)
3. 코딩으로 확인하는 경사 하강법
4. 다중 선형 회귀 (Multiple Linear Regression)
5. 코딩으로 확인하는 다중 선형 회귀

- $y = x^2$ 그래프에서 x 에 다음과 같이 a_1 , m 그리고 a_2 을 대입하여 그 자리에서 미분하면 아래 그림처럼 각 점에서의 순간 기울기가 그려짐



순간 기울기가 0인 점이 곧 우리가 찾는 최소값 m이다.

- 여기서 눈 여겨 봐야 할 것은 우리가 찾는 최소값 m 에서의 순간 기울기임
- 그래프가 이차 함수 포물선이므로 꼭지점의 기울기는 x 축과 평행한 선이 됨
- 즉, 기울기가 0임
- 우리가 할 일은 '미분 값이 0인 지점'을 찾는 것이 됨

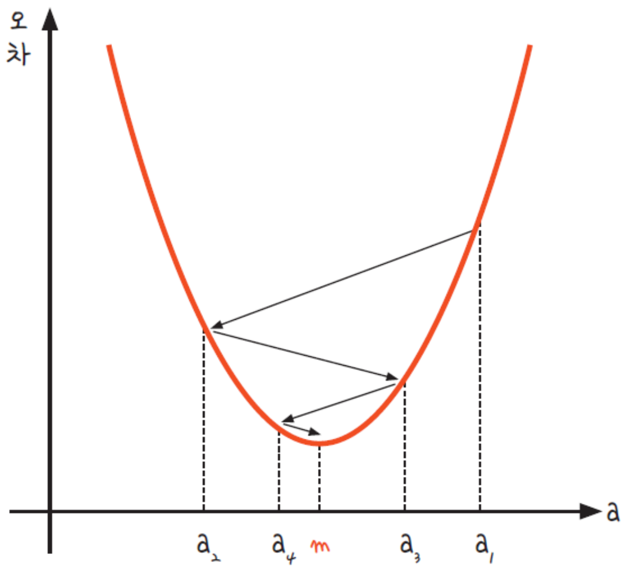
1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀



기울기가 0인 한 점(m)으로 수렴함

- 최소점 m을 찾아가는 과정

  1. a_1 에서 미분을 구함

  2. 구해진 기울기의 반대 방향(기울기가 +이면 음의 방향, - 이면 양의 방향)으로 얼마간 이동시킨 a_2 에서 미분을 구함 (그림 참조)

  3. 위에서 구한 미분 값이 0 이 아니면 위 과정을 반복함

1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

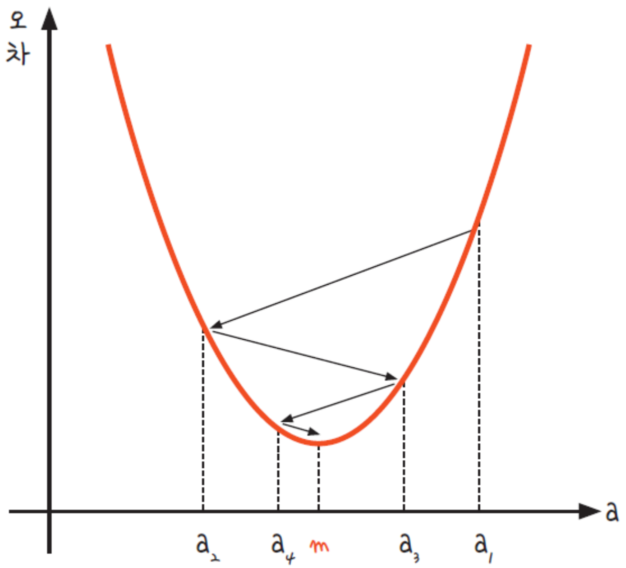
3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

- **경사 하강법**

이렇게 반복적으로 기울기 a 를 변화시켜서 m 의 값을 찾아내는 방법을 말함



1. 경사 하강법
(Gradient Decent)

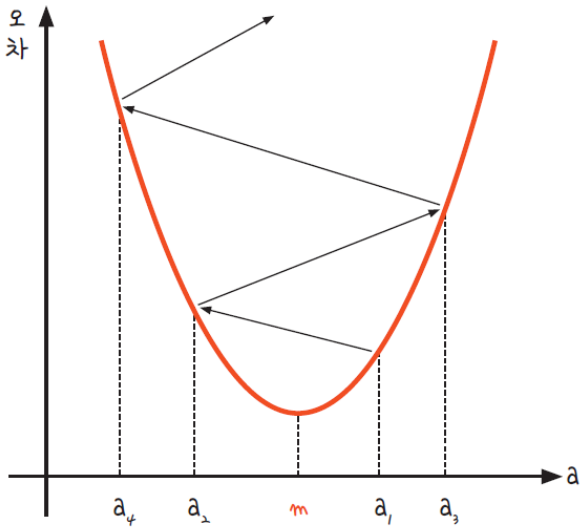
2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

- 기울기의 부호를 바꿔 이동시킬 때 적절한 거리를 찾지 못해 너무 멀리 이동시키면 a값이 한 점으로 모이지 않고 그림 처럼 위로 치솟아 버림
- **학습률(learning rate)** : 어느 만큼 이동시킬지를 신중히 결정해야 하는데, 이때 이동 거리를 정해주는 것



학습률을 너무 크게 잡으면 한 점으로 수렴하지 않고 발산한다.

1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

- 딥러닝에서 **학습률**의 값을 적절히 바꾸면서 최적의 학습률을 찾는 것은 중요한 최적화 과정 중 하나임

- **경사 하강법**
  - 오차의 변화에 따라 이차 함수 그래프를 만들고 적절한 학습률을 설정해 미분 값이 **0** 인 지점을 구하는 것
    - ✓ y 절편 b 의 값도 이와 같은 성질을 가지고 있음
    - ✓ b 값이 너무 크면 오차도 함께 커지고, 너무 작아도 오차가 커짐
    - ✓ 최적의 b 값을 구할 때도 역시 경사 하강법을 사용함

1. 경사 하강법
(Gradient Decent)
2. 학습률
(Learning Rate)
3. 코딩으로 확인하는
경사 하강법
4. 다중 선형 회귀
(Multiple Linear Regression)
5. 코딩으로 확인하는
다중 선형 회귀

- 최소값을 구하기 위해서는 **이차 함수**에서 미분을 해야 함
- **이차 함수는 평균 제곱 오차(MSE: Mean Squared Error)**를 통해 나온다는 것임
- 평균 제곱 오차의 식을 다시 옮겨 보면 다음과 같음

$$\text{평균 제곱 오차(MSE)} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2$$

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

- 여기서 $\hat{y}_i$ 은 x_i 를 넣었을 때의 값이므로 $y_i = ax_i + b$ 를 대입하면 다음과 같이 바뀜

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

1. 경사 하강법
   (Gradient Decent)
2. 학습률
   (Learning Rate)
3. 코딩으로 확인하는
   경사 하강법
4. 다중 선형 회귀
   (Multiple Linear Regression)
5. 코딩으로 확인하는
   다중 선형 회귀

- 이 값을 미분할 때 우리가 궁금한 것은 a 와 b 라는 것에 주의해야 함
- 식 전체를 미분하는 것이 아니라 **필요한 값을 중심으로 미분**해야 하기 때문임

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

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

미분을 이용한 기울기 계산

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

a로 편미분한 결과 유도 과정

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

b로 편미분한 결과 유도 과정

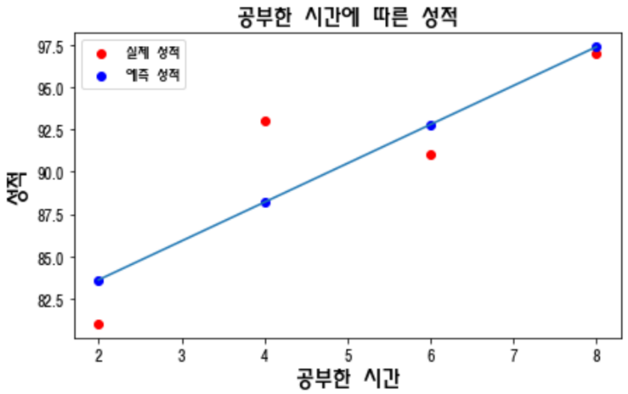
$$\begin{aligned}\frac{\partial}{\partial b}MSE(a,b) &= \frac{1}{n}\sum[(ax_i+b-y_i)^2]' \\ &= \frac{2}{n}(ax_i+b-y_i)[(ax_i+b-y_i)]' \\ &= \frac{2}{n}\sum(ax_i+b-y_i)\end{aligned}$$

1. 경사 하강법
(Gradient Decent)
2. 학습률
(Learning Rate)
3. 코딩으로 확인하는
경사 하강법
4. 다중 선형 회귀
(Multiple Linear Regression)
5. 코딩으로 확인하는
다중 선형 회귀

단순 선형 회귀 경사 하강법(Gradient Decent)을 코딩으로 구현해 보자.

16_(11 page) 강의용 단순 선형 회귀 경사 하강법.ipynb

epoch=2000, a=2.3000, b=79.0000



15_(15 page) 강의용 최소 제곱법 (Method of Least Squares).ipynb

기울기 a = 2.3, y 절편 b = 79.0 과 같음

4. 다중 선형 회귀 (Multiple Linear Regression)

1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

단순 선형 회귀(Simple Linear Regression) | 다중 선형 회귀(Multiple Linear Regression)

- 입력 x 값이 **공부한 시간** 하나가 아니라 여러 개일 경우도 발생한다.
  입력한 x 값이 여러 개인 경우는 다음 장에서 학습하고,
  우선 계속해서 **공부한 시간** 하나로 평균 제곱근 오차에 대입해보자.

이제 여기서 학습해보자.

1. 더 정확한 예측을 하려면 추가 정보를 입력해야 하며, 정보를 추가해 새로운 예측값을 계산하려면 변수의 개수를 늘려 **다중 선형 회귀**(Multiple Linear Regression)로 해결해야 함

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

공부한 시간, 과외 수업 횟수에 따른 성적 데이터

Dr. Heesuk Kim

1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

- 그럼 지금부터 두 개의 독립 변수 x_1 과 x_2 가 생긴 것임
- 이를 사용해 **종속 변수** y 를 만들 경우 **기울기를 2 개** 구해야 하므로 다음과 같은 식이 나옴

$$y = a_1 x_1 + a_2 x_2 + b$$

1. 경사 하강법
(Gradient Decent)

2. 학습률
(Learning Rate)

3. 코딩으로 확인하는
경사 하강법

4. 다중 선형 회귀
(Multiple Linear Regression)

5. 코딩으로 확인하는
다중 선형 회귀

다중 선형 회귀 경사 하강법(Gradient Decent)을 코딩으로 구현해 보자.

16_(14 page) 강의용 다중 선형 회귀 경사 하강법.ipynb