Gradient Descent Intuition

DS School

https://dsschool.co.kr

Gradient Descent Intuition

다음과 같은 값들이 주어졌을 때
우리는 Loss Function을 정의함으로써 주어진 문제를 잘 맞췄는지 판단할 수 있다

m - The Number of data

$x^{(i)}$ - i'th feature of data

$y^{(i)}$ - i'th label of data

Gradient Descent Intuition

다음과 같은 값들이 주어졌을 때
우리는 **Loss Function**을 정의함으로써 주어진 문제를 잘 맞췄는지 판단할 수 있다

m - The Number of data

$x^{(i)}$ - i'th feature of data

$y^{(i)}$ - i'th label of data

$w = Weight$
$b = Bias$

Gradient Descent Intuition

다음과 같은 값들이 주어졌을 때
우리는 Loss Function을 정의함으로써 주어진 문제를 잘 맞췄는지 판단할 수 있다

m - The Number of data

$x^{(i)}$ - i'th feature of data

$y^{(i)}$ - i'th label of data

$w = Weight$
$b = Bias$

$$h(x) = wx + b$$

$$L(x) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

Gradient Descent Intuition

다음과 같은 값들이 주어졌을 때
우리는 Loss Function을 정의함으로써 주어진 문제를 잘 맞췄는지 판단할 수 있다

m - The Number of data

$x^{(i)}$ - i'th feature of data

$y^{(i)}$ - i'th label of data

$w = Weight$
$b = Bias$

Hypothesis Function

$$h(x) = wx + b$$

Loss Function

(=Cost Function)
(=Objective Function)

$$L(x) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

Gradient Descent Intuition

그러므로 우리는 1) Loss Function을 편미분하고, 2) 편미분한 결과로 weight의 변화량을 찾는다.
그리고 2번을 w에 계속 업데이트 해주면 언젠가는 Loss Function이 0에 근접해지면서 적합한 weight를 찾을 수 있다

$$h(x) = wx + b$$

$$L(x) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

1) Hypothesis Function과 Loss Function을 정의한다.

Gradient Descent Intuition

그러므로 우리는 1) Loss Function을 편미분하고, 2) 편미분한 결과로 weight의 변화량을 찾는다.
그리고 2번을 w에 계속 업데이트 해주면 언젠가는 Loss Function이 0에 근접해지면서 적합한 weight를 찾을 수 있다

$$h(x) = wx + b$$

$$L(x) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

1) Hypothesis Function과 Loss Function을 정의한다.

$$\frac{\partial L(x)}{\partial w} = ?$$

2) Loss Function의 편미분을 구한다

Gradient Descent Intuition

그러므로 우리는 1) Loss Function을 편미분하고, 2) 편미분한 결과로 weight의 변화량을 찾는다.
그리고 2번을 w에 계속 업데이트 해주면 언젠가는 Loss Function이 0에 근접해지면서 적합한 weight를 찾을 수 있다

$$h(x) = wx + b$$

$$L(x) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$$

1) Hypothesis Function과 Loss Function을 정의한다.

$$\frac{\partial L(x)}{\partial w} = ?$$

2) Loss Function의 편미분을 구한다

$$w = w - \lambda \frac{\partial L(x)}{\partial w}$$

3) 편미분한 결과를 w에 반복해서 업데이트해준다.

Gradient Descent Intuition

먼저 Loss Function을 살펴보는데,
우리가 Loss Function의 편미분 버전을 이해하는데 평균에 해당하는 부분은 필요하지 않기 때문에 이를 제거하고 편미분을 유도한다

$$L(x) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$$

Gradient Descent Intuition

먼저 Loss Function을 살펴보는데,
우리가 Loss Function의 편미분 버전을 이해하는데 평균에 해당하는 부분은 필요하지 않기 때문에 이를 제거하고 편미분을 유도한다

이 부분은 평균(mean)에 해당하는데
우리가 이 Loss Fucntion의 편미분 유도를 이해하는데 중요하지 않으므로
제거하고 유도한다.

$$L(x) = \boxed{\frac{1}{2m}\sum_{i=1}^{m}}(h(x^{(i)}) - y^{(i)})^2$$

Gradient Descent Intuition

먼저 Loss Function을 살펴보는데,
우리가 Loss Function의 편미분 버전을 이해하는데 평균에 해당하는 부분은 필요하지 않기 때문에 이를 제거하고 편미분을 유도한다

이 부분은 평균(mean)에 해당하는데
우리가 이 Loss Fucntion의 편미분 유도를 이해하는데 중요하지 않으므로
제거하고 유도한다.

$$L(x) = \boxed{\frac{1}{2m} \sum_{i=1}^{m}} (h(x^{(i)}) - y^{(i)})^2$$

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

Gradient Descent Intuition

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} =$$

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

Gradient Descent Intuition

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

편미분 기호는 상수 안으로 들어갈 수 있다

$$\frac{\partial L(x)}{\partial w} = \boxed{\frac{\partial}{\partial w}}\left(\frac{1}{2}(h(x) - y)^2\right)$$

Gradient Descent Intuition

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\boxed{\frac{\partial}{\partial w}\left((h(x) - y)^2\right)}$$

얼핏 보면 하나의 함수 같이 보이지만,
실제로는 1) h(x) - y 와 2) 1)번의 제곱이 묶여있는 합성함수이다.

그러므로 합성함수의 미분(chain rule)을 사용해서 풀어줘야 한다.

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \boxed{\frac{1}{2}2(h(x) - y)}\frac{\partial}{\partial w}\left(h(x) - y\right)$$

일단 합성함수의 바깥 부분을 먼저 편미분해준다.
편미분 후에 나오는 2를 통해 1/2를 없앨 수 있다.

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\boxed{\frac{\partial}{\partial w}\left(h(x) - y\right)}$$

합성함수의 뒤 부분은 h(x)를 wx + b로 풀어주면
간단하게 편미분할 수 있다.

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}(wx + b - y)$$

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\boxed{\frac{\partial}{\partial w}(wx + b - y)}$$

편미분은 자기 자신을 제외한 나머지는 상수로 가정한다.
그러므로 w를 제외한 나머지는 상수이며, wx + b - y를 w로 편미분하면 x만 남는다.

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\frac{\partial L(x)}{\partial w} = \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right)$$

$$= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right)$$

$$= (h(x) - y)\frac{\partial}{\partial w}(wx + b - y)$$

$$= (h(x) - y)x$$

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

합성함수의 미분(chain rule)

$y = f(g(x))$ 일때

$y' = f'(g(x))g'(x)$

참고자료 https://goo.gl/P7kFWW

그러므로

$f(x) = (h(x) - y)^2$ 일때

$f(x) = g(x)^2$

$g(x) = (h(x) - y)$ 라고 가정하면
합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

이후에는 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다

$$L(x) = \frac{1}{2}(h(x) - y)^2$$

$$\begin{aligned}
\frac{\partial L(x)}{\partial w} &= \frac{\partial}{\partial w}\left(\frac{1}{2}(h(x) - y)^2\right) \\
&= \frac{1}{2}\frac{\partial}{\partial w}\left((h(x) - y)^2\right) \\
&= \frac{1}{2}2(h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right) \\
&= (h(x) - y)\frac{\partial}{\partial w}\left(h(x) - y\right) \\
&= (h(x) - y)\frac{\partial}{\partial w}(wx + b - y) \\
&= (h(x) - y)x
\end{aligned}$$

결론

$$\frac{\partial L(x)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})x^{(i)}$$

$$\frac{\partial L(x)}{\partial b} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})$$

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 Loss Function의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 **Loss Function**의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```

$$h(x) = wx + b$$

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 Loss Function의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```

$$h(x) = wx + b$$

$$\frac{\partial L(x)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})x^{(i)}$$

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 Loss Function의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```

$$h(x) = wx + b$$

$$\frac{\partial L(x)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})x^{(i)}$$

$$\frac{\partial L(x)}{\partial b} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})$$

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 **Loss Function**의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```

$$h(x) = wx + b$$

$$\frac{\partial L(x)}{\partial w} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})x^{(i)}$$

$$w = w - \lambda\frac{\partial L(x)}{\partial w}$$

$$\frac{\partial L(x)}{\partial b} = \frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})$$

Gradient Descent Intuition

우리가 지금까지 작성한 코드를 살펴보면
위 Loss Function의 편미분 버전이 포함되어 있다는 것을 알 수 있다

```
num_epoch = 100
learning_rate = 1.0

w1 = np.random.uniform(low=-1.0, high=1.0)
w2 = np.random.uniform(low=-1.0, high=1.0)
b = np.random.uniform(low=-1.0, high=1.0)

for epoch in range(num_epoch):
    y_predict = x1 * w1 + x2 * w2 + b

    w1 = w1 - learning_rate * ((y_predict - y) * x1).mean()
    w2 = w2 - learning_rate * ((y_predict - y) * x2).mean()
    b = b - learning_rate * (y_predict - y).mean()
```