

Gradient Descent Intuition

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

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

$y^{(i)}$ = The label of i'th 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

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

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

w = The weight

b = The bias

Gradient Descent Intuition

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

m = The number of data

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

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

w = The weight

b = The bias

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

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

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

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

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

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

w = The weight

b = The bias

Hypothesis Function

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

Loss Function

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

Cost Function

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

Gradient Descent Intuition

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

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

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

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

1) Hypothesis Function과 Loss Function, Cost 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

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

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

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

2) Loss Function을 편미분한다.

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

1) Hypothesis Function과 Loss Function, Cost 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

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

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

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

1) Hypothesis Function과 Loss Function, Cost Function을 정의한다.

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

2) Loss Function을 편미분한다.

$$\frac{\partial}{\partial w} J(w, b) = \frac{\partial}{\partial w} \left(\frac{1}{m} \sum_{i=1}^m L(y, h(x)) \right)$$

3) Loss Function의 편미분을 바탕으로
Cost 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

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

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

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

1) Hypothesis Function과 Loss Function, Cost Function을 정의한다.

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

2) Loss Function을 편미분한다.

$$\frac{\partial}{\partial w} J(w, b) = \frac{\partial}{\partial w} \left(\frac{1}{m} \sum_{i=1}^m L(y, h(x)) \right)$$

3) Loss Function의 편미분을 바탕으로
Cost Function의 편미분을 구한다.

$$w = w - \lambda \frac{\partial}{\partial w} J(w, b)$$

4) 편미분한 결과를 w에 반복해서 업데이트해준다.

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

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

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

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)}))$$

1) Hypothesis Function과 Loss Function, Cost Function을 정의한다.

?

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

2) Loss Function을 편미분한다.

$$\frac{\partial}{\partial w} J(w, b) = \frac{\partial}{\partial w} \left(\frac{1}{m} \sum_{i=1}^m L(y, h(x)) \right)$$

3) Loss Function의 편미분을 바탕으로
Cost Function의 편미분을 구한다.

λ - learning_rate

$$w = w - \lambda \frac{\partial}{\partial w} J(w, b)$$

4) 편미분한 결과를 w에 반복해서 업데이트해준다.

Gradient Descent Intuition

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

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

**Loss Function의 편미분을 간단한 미분 공식과
합성함수의 미분(chain rule)을 이용하면 간단하다**

Gradient Descent Intuition

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

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

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

Gradient Descent Intuition

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

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

편미분 기호는 상수 안으로 들어갈 수 있다

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

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

Gradient Descent Intuition

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

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

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

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

Gradient Descent Intuition

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

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

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

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

Gradient Descent Intuition

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

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

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)

$$y = f(g(x)) \quad \text{일때}$$

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

참고자료 <https://goo.gl/P7kFWW>

그러므로

$$f(x) = (h(x) - y)^2 \quad \text{일때}$$

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

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

Gradient Descent Intuition

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

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

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

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

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

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)

$$y = f(g(x)) \quad \text{일때}$$

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

참고자료 <https://goo.gl/P7kFWW>

그러므로

$$f(x) = (h(x) - y)^2 \quad \text{일때}$$

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

$$g(x) = (h(x) - y) \quad \text{라고 가정하면}$$

합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

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

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

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

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

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

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

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

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

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

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

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

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

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

합성함수의 뒤 부분은 $h(x)$ 를 $w x + b$ 로 풀어주면
간단하게 편미분할 수 있다.

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

합성함수의 미분(chain rule)

$$y = f(g(x)) \quad \text{일때}$$

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

참고자료 <https://goo.gl/P7kFWW>

그러므로

$$f(x) = (h(x) - y)^2 \quad \text{일때}$$

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

$$g(x) = (h(x) - y) \quad \text{라고 가정하면}$$

합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

편미분은 자기 자신을 제외한 나머지는 상수로 가정한다.

그러므로 w 를 제외한 나머지는 상수이며, $wx + b - y$ 를 w 로 편미분하면 x 만 남는다.

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

합성함수의 미분(chain rule)

$$y = f(g(x)) \quad \text{일때}$$

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

참고자료 <https://goo.gl/P7kFWW>

그러므로

$$f(x) = (h(x) - y)^2 \quad \text{일때}$$

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

$$g(x) = (h(x) - y) \quad \text{라고 가정하면}$$

합성함수의 미분을 사용할 수 있다.

Gradient Descent Intuition

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

결론

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

$$\frac{\partial}{\partial w} L(y, h(x)) = (h(x) - y)x$$

$$\frac{\partial}{\partial w} J(w, b) = \frac{\partial}{\partial w} \left(\frac{1}{m} \sum_{i=1}^m L(y^{(i)}, h(x^{(i)})) \right)$$

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

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

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

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()
```


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()
```

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

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()
```

$$\frac{\partial}{\partial w} L(y, h(x)) = (h(x) - y)x$$

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()
```

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

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()
```

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

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()
```

$$w = w - \lambda \frac{\partial}{\partial w} J(w, b)$$

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()
```