Gradient descent 알고리즘 구현하기

Gradient descent 알고리즘 구현하기

- Gradient descent 알고리즘은 손실 함수(loss function)의 미분값인 gradient를 이용해 모델에게 맞는 최적의 가중치(weight), 즉 손실 함수의 값을 최소화 하는 가중치를 구할 수 있는 알고리즘입니다.
- 이번 실습에서는 Gradient descent 알고리즘을 직접 구현한 후, 이를 이용해 데이터를 가장 잘 설명하는 선형 회귀 직선의 기울기와 y절편, 즉 선형 회귀 모델에게 맞는 최적의 가중치를 찾아보겠습니다.
- 선형 회귀 직선의 수식은 다음과 같은 1차 함수 형태이며, 우리가 Gradient descent 알고리즘을 사용해 찾을 값,
- 즉 가중치는 $w0$과 $w1$입니다.

$$f(x) = w_0 + w_1 x$$

손실함수(loss function)

- 손실 함수(loss function)는 실제값과 모델이 예측한 값 간의 차이를 계산해주는 함수입니다. 손실 함수의 값은 가중치와 편향을 업데이트하는 데에 사용됩니다. 여기서는 손실 함수로 MSE (Mean Squared Error)를 사용합니다.
- MSE는 평균 제곱 오차 함수로, 수식은 다음과 같습니다.

$$
\begin{aligned}
Loss(w_0, w_1) &= \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(y_i - (w_0 + w_1 x_i))^2
\end{aligned}
$$

여기서 y_i는 실제값, $f(x_i) = w_0 + w_1 x_i$는 모델 $f(x)$에 x_i를 넣어서 나온 예측값입니다.

편미분

- gradient에는 **편미분**이라는 개념이 들어갑니다.
- 편미분이란 2개 이상의 변수를 가진 함수에서 우리가 미분할 하나의 변수를 제외한 나머지 변수들을 상수로 보고, 미분 대상인 그 변수로 미분하는 것입니다.

예를 들어 $f(x,y) = 2x^2 + y$ 라는 수식이 있을 때, x에 대해서만 편미분한다면 $(x,y) = \frac{\partial f(x,y)}{\partial x} = 4x$ 가 되는 것입니다.

Gradient

- gradient는 곧 기울기 벡터를 의미하며, 선형 함수의 각 파라미터들의 편미분으로 구성된 열벡터로 정의합니다.
- 학습률(learning rate)을 나타내는 α가 있고, gradient를 나타내는 수식인 $\nabla Loss(W)$가 있습니다.
- 즉 이를 풀어서 쓰면 다음과 같은 열벡터 형태입니다.

$$gradient = \nabla Loss(W) = \begin{pmatrix} \frac{\partial Loss}{\partial w_0} \\ \frac{\partial Loss}{\partial w_1} \\ \vdots \\ \frac{\partial Loss}{\partial w_p} \end{pmatrix}$$

따라서 우리가 구해야 할 w_0와 w_1에 대한 gradient는 다음과 같습니다.

$$gradient = [\frac{\partial Loss}{\partial w_0}, \frac{\partial Loss}{\partial w_1}]$$

- $w0$과 $w1$에 대한 gradient를 구하기 위해 *Loss*를 각각에 대해 편미분하면 다음과 같습니다.

$$\frac{\partial Loss}{\partial w_0} = \frac{2}{N}\sum_{i=1}^{N}(y_i - (w_0 + w_1 x_i))(-1)$$

$$\frac{\partial Loss}{\partial w_1} = \frac{2}{N}\sum_{i=1}^{N}(y_i - (w_0 + w_1 x_i))(-x_i)$$

실습

- 설명 중 '손실 함수' 파트의 수식을 참고해 MSE 손실 함수를 완성하세요.
- 설명 중 'Gradient' 파트의 마지막 두 수식을 참고해 w0와 w1에 대한 gradient인 gradient0과 gradient1을 반환하는 함수를 완성하세요.
- 설명 중 '가중치 업데이트' 파트의 두 수식을 참고해 gradient descent를 통한 가중치 업데이트 코드를 작성하세요.

MSE (Mean Squared Error)	예측값과 실젯값 차이를 제곱하여 평균 낸 값으로, 0에 가까울수록 예측값과 실제값의 차이가 없으므로 성능이 우수함.	$(MSE) = \frac{1}{n}\sum_{i=1}^{n}(H(x_i) - y_i)^2$
RMSE (Root Mean Squared Error)	MSE를 Root처리한 값으로 0에 가까울수록 성능이 우수함.	$(\mathrm{SE}) = \sqrt{\mathrm{MSE}}$
MAE (Mean Absolute Error)	예측값과 실젯값 차이의 절댓값 평균 낸 값으로, 0에 가까울수록 성능이 우수함.	$(MAE) = \frac{1}{n}\sum_{i=1}^{n} \mid H(x_i) - y_i \mid$
R2 (R Squared, R2, 결정계수)	1에 가까울수록 성능이 우수함.	$(R^2) = \frac{\sum_{i=1}^{n}(H(x_i) - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$