· 'Norm'은 벡터, 행렬, 텐서의 '크기'를 의미하는 값이다.

· Machine learning model의 weight norm은 해당 "모델의 복잡도"를 의미한다.

② **L1 Regularization** (Lasso)

Regularization은 통상적으로 L1과 L2 regularization으로 나눠지게 된다.
앞서 살펴본 수식은 L2 regularization에 속하고,
L1 regularization은 2차항 대신에 1차항이 오며, 식은 아래와 같다.

$$C = \underbrace{C_0}_{\text{원래의}} + \underbrace{\frac{\lambda}{n}\sum_w |w|}_{\text{cost function}}$$

'w'에 대해 편미분 하면,

앞서 살펴본 것과 마찬가지로 가중치 w에 대해서 편미분을 수행하면,
결과적으로는 새로운 가중치는 아래와 같이 결정이 된다.

$$w \rightarrow w' = w - \frac{\overset{\text{learning rate}}{\eta}\lambda}{n}\mathrm{sgn}(w) - \underset{\text{learning rate}}{\eta}\frac{\partial C_0}{\partial w}$$

이에 따라, 원래 w 값의 "절대값"이 작아짐.

결과적으로 위 식을 보면,
(weight 값 자체를 줄이는 것이 아니라)
"w의 부호에 따라 상수 값을 빼주는 방식"으로 regularization을 수행한다.

음수면 '-1', 양수면 '1'을 뺀다.

① **L2 Regularization** (Ridge)

Regularization은 (정확하게 표현하면, L2 regularization은) 아래의 수식으로 표현할 수 있다.

$$C = C_0 + \frac{\lambda}{2n}\sum_w w^2$$

'w'에 대해 편미분 하면,

위 수식에서 C_0는 원래의 cost function이며,
n은 훈련 데이터의 개수, hyperparameter.
λ는 regularization 변수, w는 가중치를 나타낸다.
위 식처럼 regularization 항목이 들어가면,
학습의 방향이 단순하게 C_0값이 작아지는 방향으로만 진행되는 것이 아니라,
w값들 역시 최소가 되는 방향으로 진행을 하게 된다.
이렇게 정의된 cost function을 가중치 w에 대해서 편미분을 수행하면,
결과적으로는 새로운 가중치는 아래와 같이 결정이 된다.

$$w \rightarrow w - \eta\frac{\partial C_0}{\partial w} - \frac{\eta\lambda}{n}w = w - \frac{\eta\lambda}{n}\cdot w - \eta\frac{\partial C_0}{\partial w}$$

$$= \left(1 - \frac{\eta\lambda}{n}\right)w - \eta\frac{\partial C_0}{\partial w}$$

원래 w 값의 "절대값"이 작아짐.

위 식에서 $(1 - \eta\lambda/n)w$는
원래의 w값에 $(1 - \eta\lambda/n)$ 항목을 곱한 형태가 되기 때문에
값이 작아지는 방향으로 진행을 하게 된다.