

WEIGHT DECAY

A parameter norm penalty using the L2 norm:

$$\text{Loss} = \text{Loss} + \frac{1}{2} \|w\|_2^2$$

Annotations:

- Loss function (pointing to the first Loss)
- Loss function (pointing to the second Loss)
- weights (pointing to w)
- L2 norm penalty (pointing to the entire term $\frac{1}{2} \|w\|_2^2$)

In regression this use of the L2 norm penalty is called ridge regression.

Chris Albon