

PARAMETER NORM

Penalties for Regularization

Adds a parameter norm penalty to loss function:

$$\text{Loss}' = \text{Loss} + \lambda \Omega(\theta)$$

Diagram illustrating the components of the regularized loss function:

- Loss' : Total regularized loss
- Loss : Loss function
- λ : Hyperparameter (positive number)
- $\Omega(\theta)$: Norm Penalty
- θ : Parameters

- Most often not applied to parameters of biases.
- Technically each layer can have its own hyperparameter and/or norm penalty.

Chris Albon