

Weight Normalization

Reparameterization

$$w = g \cdot (v / \|v\|)$$

g

Magnitude
Scale

v / ||v||

Direction
Normalized



Weight Norm vs Batch Norm

Key Features

Decouples magnitude (**g**) and direction (**v**) of weights

Normalizes **weights** rather than activations

Reduces **parameters** compared to Batch Norm

Faster training than Batch Norm in some cases

✓ Works Well With

RNNs and **Reinforcement Learning** applications

Weight Norm

Normalizes weights

Batch Norm

Normalizes activations

Fewer params

g and v only

More params

γ , β + statistics

Less commonly used in modern architectures