

# BATCH NORMALIZATION

In neural networks, Batch Normalization (BatchNorm) is a technique used to normalize the output of each activation layer, using the mean and standard deviation of each mini-batch during training.

This helps to standardize the activations, which avoids internal covariate shift — the problem where the input distribution to each layer keeps changing during training.

## WHY IS THIS USEFUL?

It keeps the activations in a stable range

Less sensitive to initialization

Able to use higher learning rates

Less prone to overfitting

## STEPS IN BATCH NORMALIZATION

Normalize the batch:

$$\hat{a}_i = \frac{a_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

Where:

- $\mu_B$  = mean of the batch
- $\sigma_B^2$  = variance of the batch
- $\epsilon$  = small constant to prevent divide-by-zero

Scale and shift

$$y_i = \gamma \hat{a}_i + \beta$$

Where:

- $\gamma$  = scaling factor (learned)
- $\beta$  = shifting factor (learned)