Weight Initialization Techniques

1. Key points in weight initialization
2. Weight initialization techniques

Different types of weights initializing techniques are available in deep learning. Each technique follows a unique way to initialize the weights. The accuracy of the NN is also depends on the type of the weight initialization technique.

1. Key Points in Weights Initialization

The below key point should remember before Initialize the weights:

1. Weight should be small but not similar to zero ($w \not\approx 0$).
2. Weight should not be same.
3. Weights should have good variance.

1.1. Weight should be small:

- Faster Training: Smaller weights can lead to faster convergence during training.
- Regularization: Small weights act as implicit regularization, preventing overfitting. Regularization techniques like L1 or L2 are explicitly penalize the large weights, encouraging the network to learn simpler patterns that generalize better on unseen data.
- Numerical Stability: Smaller weights give numerical stability in math calculations and prevent overflow issues.
- Memory Efficiency: Small weights enhance memory efficiency, crucial for resource-constrained environments.

1.2. Weight should not be same:

If all weights are same then all neurons performance will be same, and the outputs of each neuron also same. In this case ANN can't separate the data in classification. That's why weights should not be same.

1.3. Weights should have good variance:

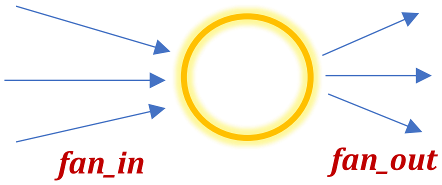
The learning rate is bad when the difference between weights is less. Show weights should have good variance.

Terms involved in weight initialization:

Mainly there are two terms involved in every weight initialization technique formula those are:

Inputs → fan_in

Outputs → fan_out



2. Weights Initialization Techniques:

Mainly there are four mostly used weight initialization techniques are available in deep learning:

1. Uniform distribution
2. Xavier / Gorat distribution
3. He init distribution

2.1. Uniform Distribution:

Here weight will take in between subrange depending on the formula

Normal Uniform
$\boldsymbol{w_{ij} \approx Uniform\left[\frac{-1}{\sqrt{fan\_in}}, \quad \frac{1}{\sqrt{fan\_in}}\right]}$

The distribution performed between these two points only.

2.2. Xavier / Gorat Distribution:

Xavier Normal	Xavier Uniform
$\boldsymbol{w_{ij} \approx N(0, \sigma)}$ $\boldsymbol{\sigma = \sqrt{\frac{2}{[fan\_in + fan\_out]}}}$	$w_{ij} \approx U\left[\frac{-\sqrt{6}}{\sqrt{fan\_in + fan\_out}}, \quad \frac{\sqrt{6}}{\sqrt{fan\_in + fan\_out}}\right]$

2.3. He Init Distribution:

He Normal	He Uniform
$\boldsymbol{w_{ij} \approx N(0, \sigma)}$	

$\boldsymbol{\sigma} = \sqrt{\dfrac{\mathbf{2}}{\boldsymbol{fan\_in}}}$	$w_{ij} \approx U\left[-\sqrt{\dfrac{6}{fan\_in}}, \quad \sqrt{\dfrac{6}{fan\_in}}\right]$