

# Weight Initialization

01 September 2023 10:35 AM

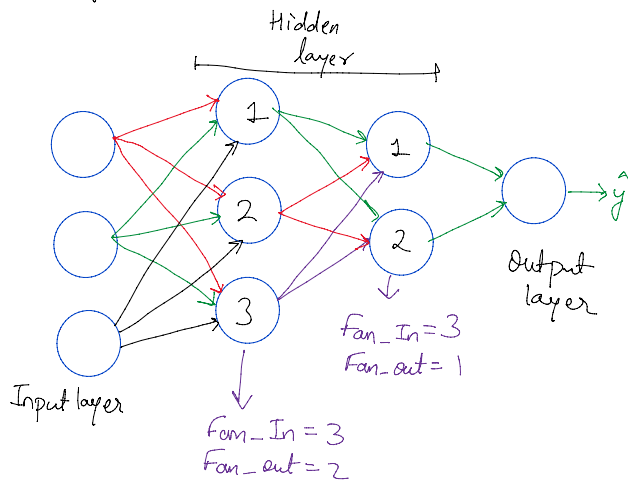
## Resources:

- Weight Initialization
- Uniform Distribution
- HeUniform
- Glorot
- HeUniform: [https://www.tensorflow.org/api\\_docs/python/tf/keras/initializers/HeUniform](https://www.tensorflow.org/api_docs/python/tf/keras/initializers/HeUniform)
- GlorotUniform: [https://www.tensorflow.org/api\\_docs/python/tf/keras/initializers/GlorotUniform](https://www.tensorflow.org/api_docs/python/tf/keras/initializers/GlorotUniform)
- Initialization Visualizer: <https://www.deeplearning.ai/ai-notes/initialization/index.html>

## What not to do during weight Initialization?

1. Zero Initialization → 1. Model does not train
2. Non-zero Constant Initialization → 2. Multiple weights update at same time. (Problem of Symmetry)
3. Random Initialization with small weights → 3. Maybe No training and Vanishing Gradient Problem.
4. Random Initialization with large weights → 4. Saturation happens ( $\sum w_i$  is very high for every neuron), can also have Vanishing Gradient as well as exploding gradient. (ReLU)

## Uniform Distribution



Uniform Distribution weight initialization is generally used with sigmoid activation function.

In the weights are initialized between the range of values:

$$\left[ \frac{-1}{\sqrt{\text{Fan-In}}}, \frac{1}{\sqrt{\text{Fan-In}}} \right]$$

## Xavier/Glorot Initialization

1. Xavier Normal → Weights are selected from a normally distributed range of values with ( $\mu=0$ ) and standard deviation ( $\sigma$ ) =  $\sqrt{\frac{2}{\text{Fan-In} + \text{Fan-Out}}}$

2. Xavier Uniform → Weights are selected from a uniform distribution ranging between

$$\left[ -\sqrt{\frac{6}{\text{Fan-In} + \text{Fan-Out}}}, \sqrt{\frac{6}{\text{Fan-In} + \text{Fan-Out}}} \right]$$

When tanh activation is used, prefer Xavier Weight Initialization

## He Weight Initialization

1. He Uniform: The initial weights are selected within the range  $\left[ -\sqrt{\frac{6}{\text{Fan-In}}}, \sqrt{\frac{6}{\text{Fan-In}}} \right]$

2. He Normal: The initial weights are selected from Normally distributed values with  $\mu=0$  and  $\sigma = \sqrt{\frac{2}{\text{Fan-In}}}$

When you have ReLU activation function prefer He Initialization.