Weight Intialization -

The first step that comes in consideration while building a neural network is the intialization of parameters, If done correctly then optimization will be achieved in the least time otherwise converging to a minima using gradient descent will be impossible.

Some Intialization technique are

ⓐ Zero Initialization -

~~Initi~~ * Initialized all weights to 0)

If we ~~init~~ initialized all weight with 0, then what happens is that the derivative wrt loss function is same for every weight, thus all weights have the ~~same~~ (same) value in subsequent iteration. This make hidden layers symmetric and this process continues for all the n iteration. Thus initialized weights with zero make your network no better than linear model.

*

Fan_in = Number of input paths ~~two~~ towards the neuron

Fan_out = Number of outputpaths towards the neuron

No.
Date : / /

2. Random Initialization

In an attempt to overcome the shortcoming of zero or constant initialization assign random values except for zeros as weights to neuro paths. ~~these~~

Random Initialization can be of two kinds

- Random Normal
- Random Uniform

Random Normal: The weight are in initialized from values in a normal distribution

$w_i \sim N(0,1)$

Uniform → The weights are initialized from values in uniform distribution

$w_i \sim$ is selected from uniform ~~distributed~~ distributed b/w range lower (a) to highest (b)

$$\left[\underset{a}{\frac{-1}{\sqrt{Fan\_in}}} \quad \underset{b}{\frac{1}{\sqrt{Fan\_in}}} \right]$$

$$\boxed{w_1 \sim [a, b]}$$

er

L
ayers

ed
k

ds

③ Xavier / Glorot ~~Norm~~ Initialization

Xavier / Glorot Initialization ~~often~~ is suitable for layers where the activation function used is sigmoid

There are two formulas:

ⓐ Uniform Xavier / Glorot Initialization draw each weight w_i from a random ~~number~~ uniform distribution in [a, b]

$$w_i \approx \left[\underbrace{-\sqrt{\frac{6}{fan\_in + fan\_out}}}_{a}, \underbrace{\sqrt{\frac{6}{fan\_in + fan\_out}}}_{b} \right]$$

$$\boxed{w_i \sim [a, b]}$$

ⓑ Normal Xavier / Glorot Initialization draw each weigh w_i from a normal distribution with a mean 0 & a standard deviation σ:

$$\sigma = \sqrt{\frac{2}{fan\_in + fan\_out}}$$

$$\boxed{w_i \approx (0, \sigma)}$$

~~[illegible]~~

(4) He-Initialization

He Initialization is suitable for layers where ReLU activation function is used.

He uniform Initialization

In He uniform weight Initialization, the weights are assigned from values of uniform distribution in $[-a, b]$

$$W \approx UD\left[\underbrace{-\sqrt{\frac{6}{fan\_in}}}_{a} ; \underbrace{\sqrt{\frac{6}{fan\_in}}}_{b}\right].$$

$$w_i \sim N[-a, b]$$

He Normal Initialization

In He normal Initialization, the weight are assigned from values of a normal distribution as follow.

$$w_i \sim N(0, \sigma) \qquad \boxed{w_i \sim N(0, \sigma)}$$

$$\sigma = \sqrt{\frac{2}{fan\_in}}$$