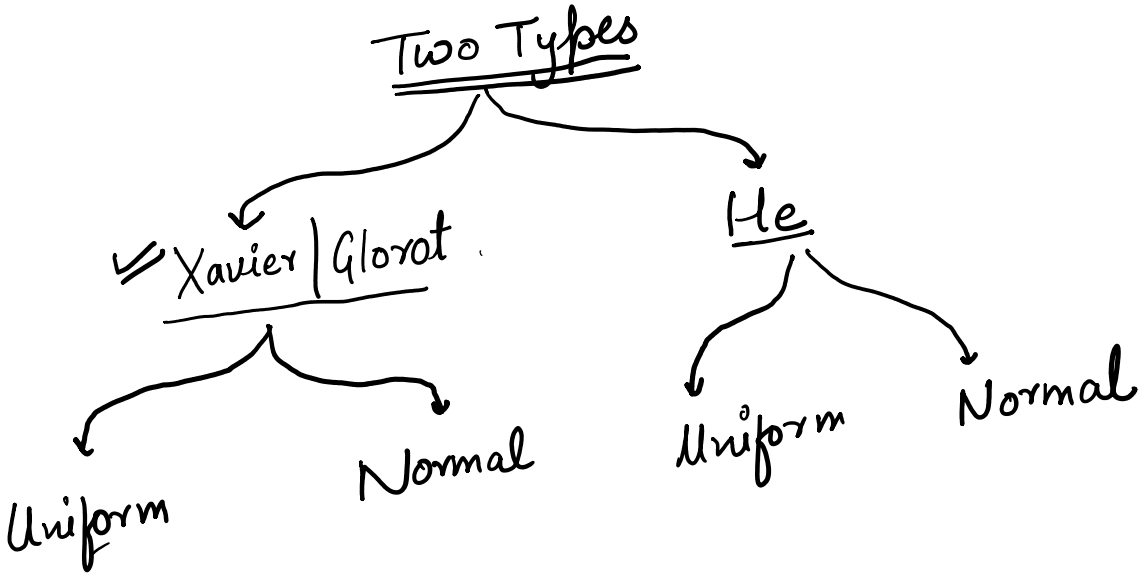
WHAT YOU WILL STUDY IN TODAY VIDEO ?

- ► Weight Initialization Techniques ✓
- ► Weight Initialization Practical Implementation ✓

Weight Initialization

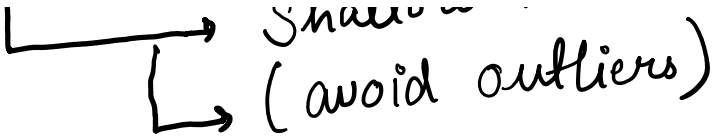
* **Uses**
  1. Prevents Vanishing / Exploding gradient
  2. Model convergence ↑



* **Uniform Vs Normal**

1. **Normal** ⇒ weights spread across zero
   → deep neural networks (stable grad flow) ✓
   ResNet

2. **Uniform** :- Weights spread across a range $(-a, a)$
   → Shallow neural networks (avoid outliers)


(avoid outliers)

Xavier / Glorot

n_{in} → no of input neurons (4)

n_{out} → no of output neurons (3)

* Neural Network

4 input neuron

3 output neuron

* formula

① $W \sim U\left(\frac{-\sqrt{6}}{\sqrt{n_{in}+n_{out}}}, \frac{\sqrt{6}}{\sqrt{n_{in}+n_{out}}}\right)$

$=$ ② $W \sim N\left(0, \frac{1}{n_{in}+n_{out}}\right)$

* Example

$n_{in} = 4$

$n_{out} = 3$

limit (Uniform) $= \left(\frac{-\sqrt{6}}{\sqrt{4+3}}, \frac{\sqrt{6}}{\sqrt{4+3}}\right)$

$= (-0.925, 0.925)$

$$= \left(-0.925, 0.925\right)$$

Now weights will be picked randomly from above range.

+ Main Use

= Vanishing Gradient

He initialization

formula

① $W \sim U\left(-\sqrt{\frac{6}{n_{in}}}, \sqrt{\frac{6}{n_{in}}}\right)$

② $W \sim N\left(0, \frac{2}{n_{in}}\right)$

Example

$n_{in} = 4$

~~$n_{out} = 3$~~

Limit (uniform) $= \left(-\sqrt{\frac{6}{4}}, \sqrt{\frac{6}{4}}\right)$

$= (-1.22, 1.22)$

Weight randomly picked from above range

pickle

* Main Use

Exploding Gradient