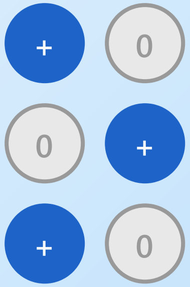
He Initialization (for ReLU)

ReLU Effect: Half Neurons Output 0



Approximately 50% neurons killed by ReLU

He Initialization Formula

$$W \sim N(0,\ 2/n_{in})$$

Variance factor of 2 compensates for ReLU's effect

Key Features

- Specifically designed for **ReLU activation** functions
- Accounts for ReLU **killing half** the neurons (output 0)
- **Maintains signal strength** in deep ReLU networks
- Significantly improves training of **very deep networks**

✓ Default Choice

Standard initialization for modern **CNN architectures** and deep learning models with ReLU

📊 Practical Calculation Example

Example 1: Layer with 4 inputs → 3 outputs

Step 1: Calculate Standard Deviation

Example 2: Layer with 128 inputs → 64 outputs

Step 1: Calculate Standard Deviation

$n_{in} = 4$

$\sigma = \sqrt{(2/n_{in})} = \sqrt{(2/4)} = \sqrt{0.5}$

$\sigma \approx 0.707$

Step 2: Generate Weight Matrix (3×4)

Sample from $N(0, 0.707^2)$

```
W = [[ 0.423, -0.891, 0.156, 0.734]
[-0.612, 0.289, -0.445, 0.821]
[ 0.534, -0.178, 0.693, -0.356]]
```

Verification

Variance ≈ 0.5 ✓

Compensates for ReLU killing ~50% neurons

$n_{in} = 128$

$\sigma = \sqrt{(2/n_{in})} = \sqrt{(2/128)}$

$\sigma = \sqrt{0.015625}$

$\sigma \approx 0.125$

Step 2: Generate Weight Matrix (64×128)

Sample from $N(0, 0.125^2)$

```
W[0,:5] = [ 0.089, -0.134, 0.112, -0.078, 0.156]
W[1,:5] = [-0.091, 0.145, -0.123, 0.067, -0.089]
W[2,:5] = [ 0.134, -0.056, 0.178, -0.145, 0.103]
...
```

(Much smaller values due to larger n_{in})

Key Insight

Larger n_{in} → Smaller weights

Prevents activation explosion 🚀

🔍 Comparison: Xavier vs He Initialization

Layer Size (n_{in})	Xavier: $\sigma = \sqrt{(1/n_{in})}$	He: $\sigma = \sqrt{(2/n_{in})}$	Ratio (He/Xavier)
4	0.500	0.707	√2 ≈ 1.41×

16	0.250	0.354	√2 ≈ 1.41×
64	0.125	0.177	√2 ≈ 1.41×
256	0.0625	0.088	√2 ≈ 1.41×

💡 Key Takeaway

He initialization uses **√2 times larger** weights than Xavier

This compensates for ReLU zeroing out half the neurons

Result: **Stable gradient flow** in deep ReLU networks