Random initialization

Random initialization

- Initialize weights

  - Normal distribution

  - Uniform distribution

- What should μ_i and σ_i be?

  - For simplicity $\mu_i = 0$ and bias $= 0$

$$\mathbf{W}_1 = \mathcal{N}(\mu_1, \sigma_1^2 \mathbf{I})$$

$$\mathbf{W}_3 = \mathcal{N}(\mu_3, \sigma_3^2 \mathbf{I})$$

$\mathbf{X}$

Linear

ReLU

Linear

$\mathbf{0}$
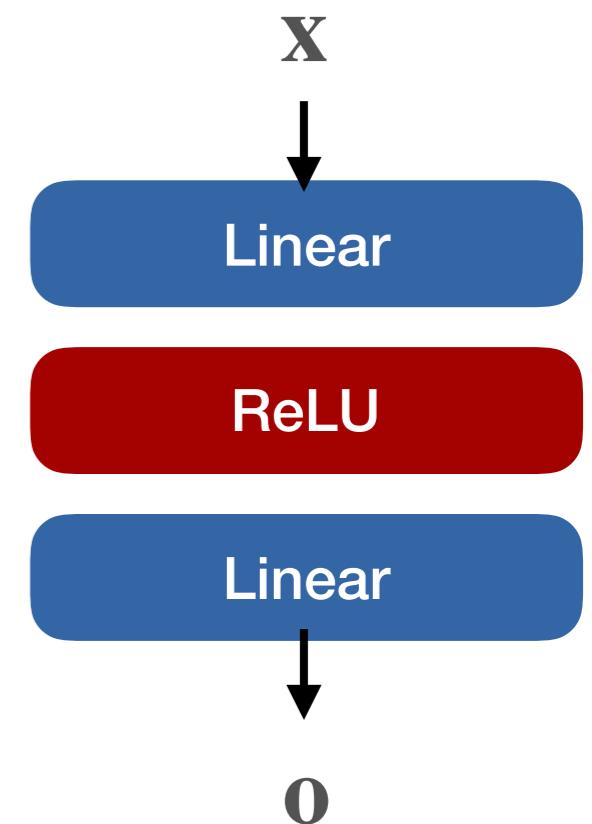
Scaling matters

$$\mathbf{o} = \mathbf{W}_2\mathbf{W}_1\mathbf{x}$$

$$\frac{\partial \ell(\mathbf{o})}{\partial \mathbf{W}_1} = \left(\mathbf{W}_2^\top \frac{\partial \ell(\mathbf{o})}{\partial \mathbf{o}} \right) \mathbf{x}^\top$$

$$\frac{\partial \ell(\mathbf{o})}{\partial \mathbf{W}_2} = \frac{\partial \ell(\mathbf{o})}{\partial \mathbf{o}} \left(\mathbf{W}_1\mathbf{x} \right)^\top$$

$\mathbf{x}$

$\mathbf{W}_1$ | Linear

$\mathbf{W}_2$ | Linear

$\mathbf{o}$
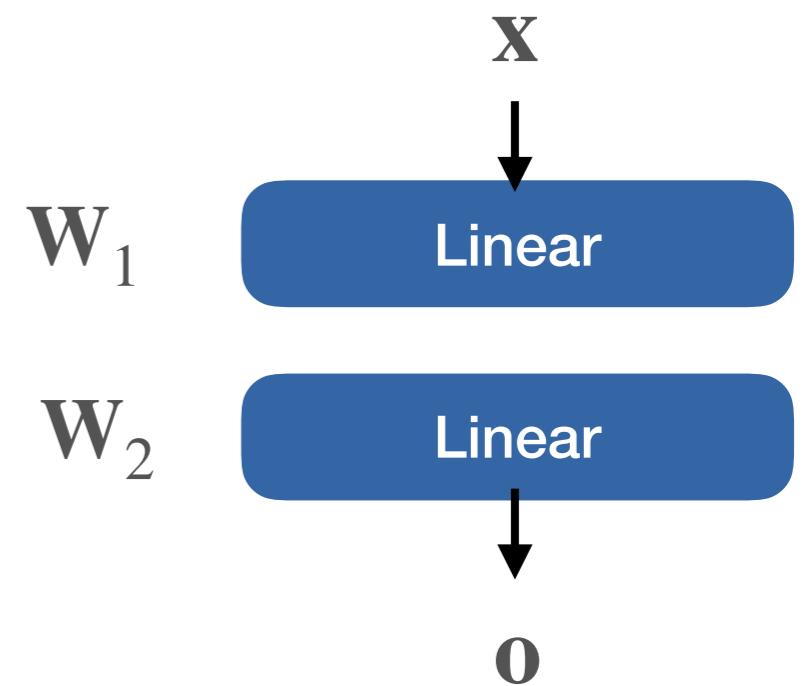
How do we scale the initialization?

- By hand

  - A lot of tuning

- Automatically

  - A lot of math

X

$\downarrow$

Linear

ReLU

Linear

ReLU

Linear

$\downarrow$

0