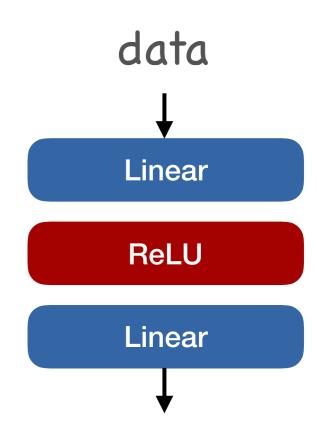
Network initialization

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

Network initialization

 What do we set the initial parameters to?



Idea 1: All zero

$$\frac{\partial \mathcal{E}(\mathbf{o})}{\partial \mathbf{W}_{1}} = \mathbf{v}_{1}\mathbf{x}^{\mathsf{T}} \qquad \mathbf{v}_{0} = \mathbf{W}_{1}^{\mathsf{T}}\mathbf{v}_{1} \qquad \mathbf{z}_{1} = \mathbf{W}_{1}\mathbf{x}$$

$$\mathbf{v}_{1} = \mathbf{v}_{2}[\mathbf{z}_{1} > 0] \qquad \mathbf{z}_{2} = \max(\mathbf{z}_{1}, 0)$$

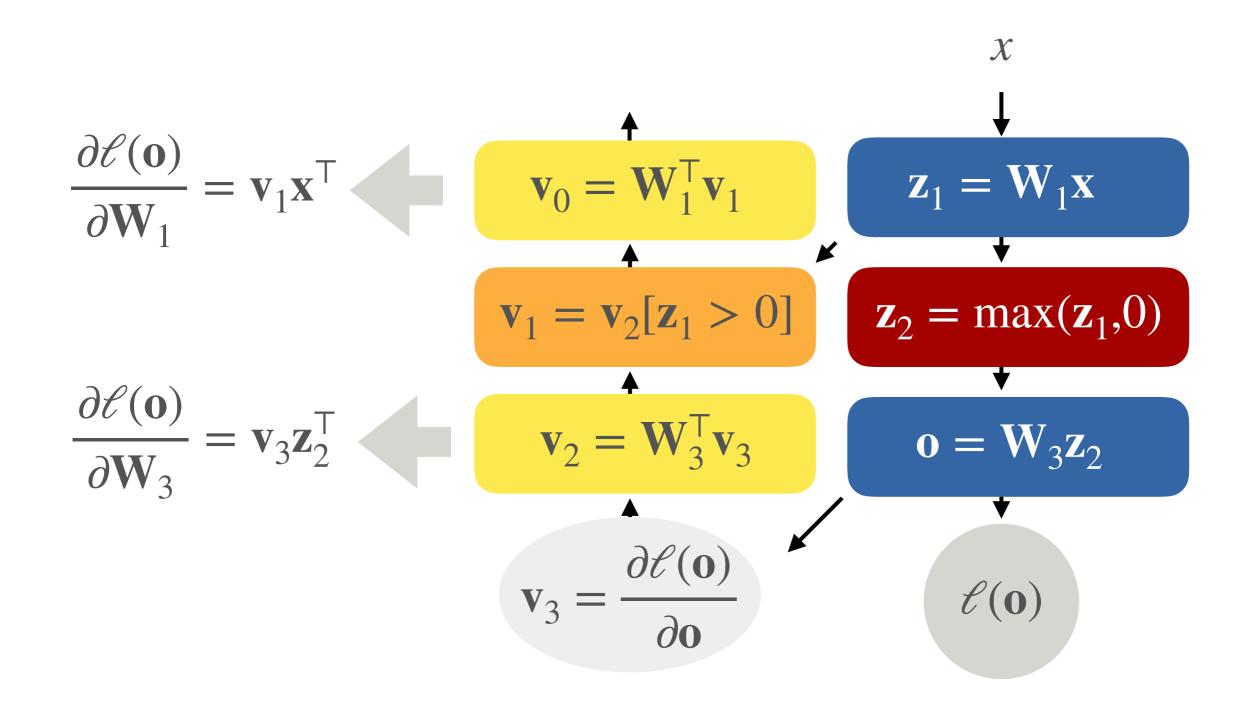
$$\frac{\partial \mathcal{E}(\mathbf{o})}{\partial \mathbf{W}_{3}} = \mathbf{v}_{3}\mathbf{z}_{2}^{\mathsf{T}} \qquad \mathbf{v}_{2} = \mathbf{W}_{3}^{\mathsf{T}}\mathbf{v}_{3} \qquad \mathbf{o} = \mathbf{W}_{3}\mathbf{z}_{2}$$

$$\mathbf{v}_{3} = \frac{\partial \mathcal{E}(\mathbf{o})}{\partial \mathbf{o}} \qquad \mathcal{E}(\mathbf{o})$$

Idea 1: All zero

- Does not work
 - No gradient
- Saddle point

Idea 2: constant



Idea 2: constant

 Does not break symmetries

Solution

Random initialization

$$\mathbf{W}_i = \mathcal{N}(\mu_i, \sigma_i^2 \mathbf{I})$$