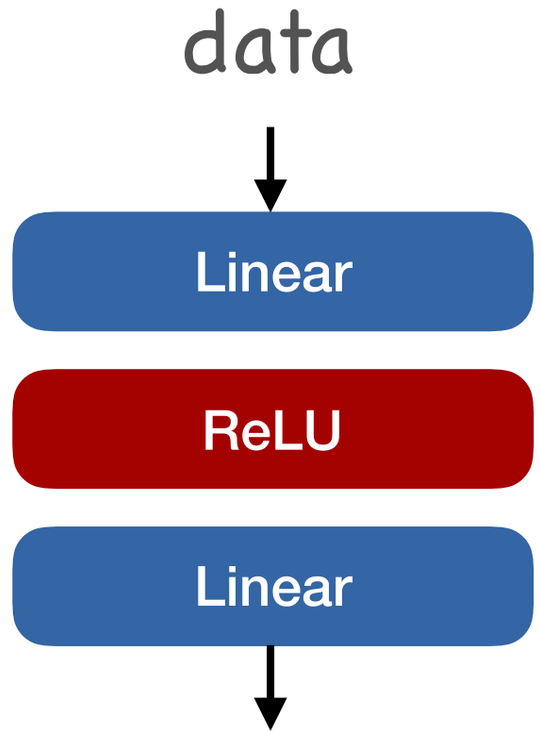
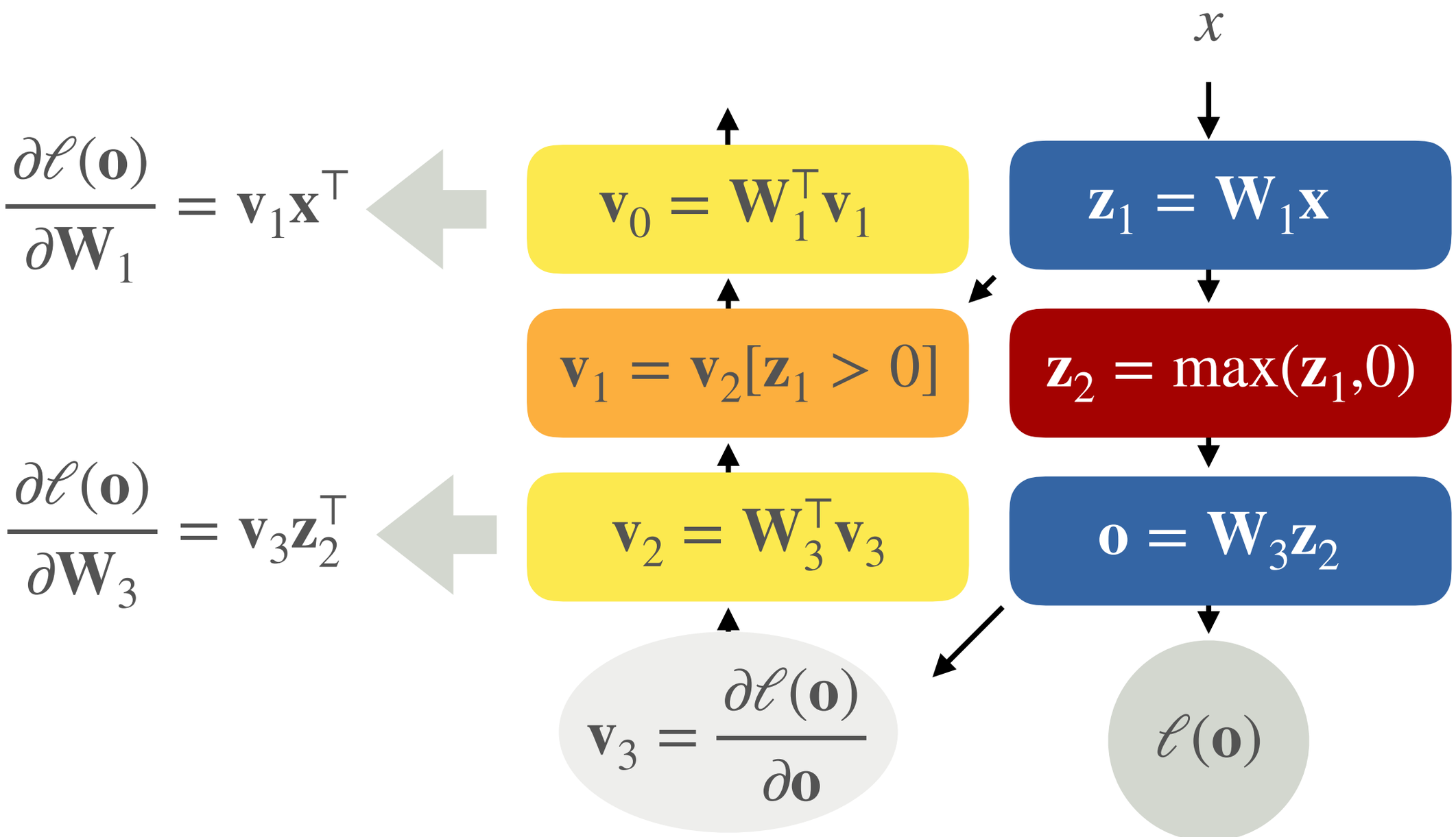
Network initialization

Network initialization

- What do we set the initial parameters to?



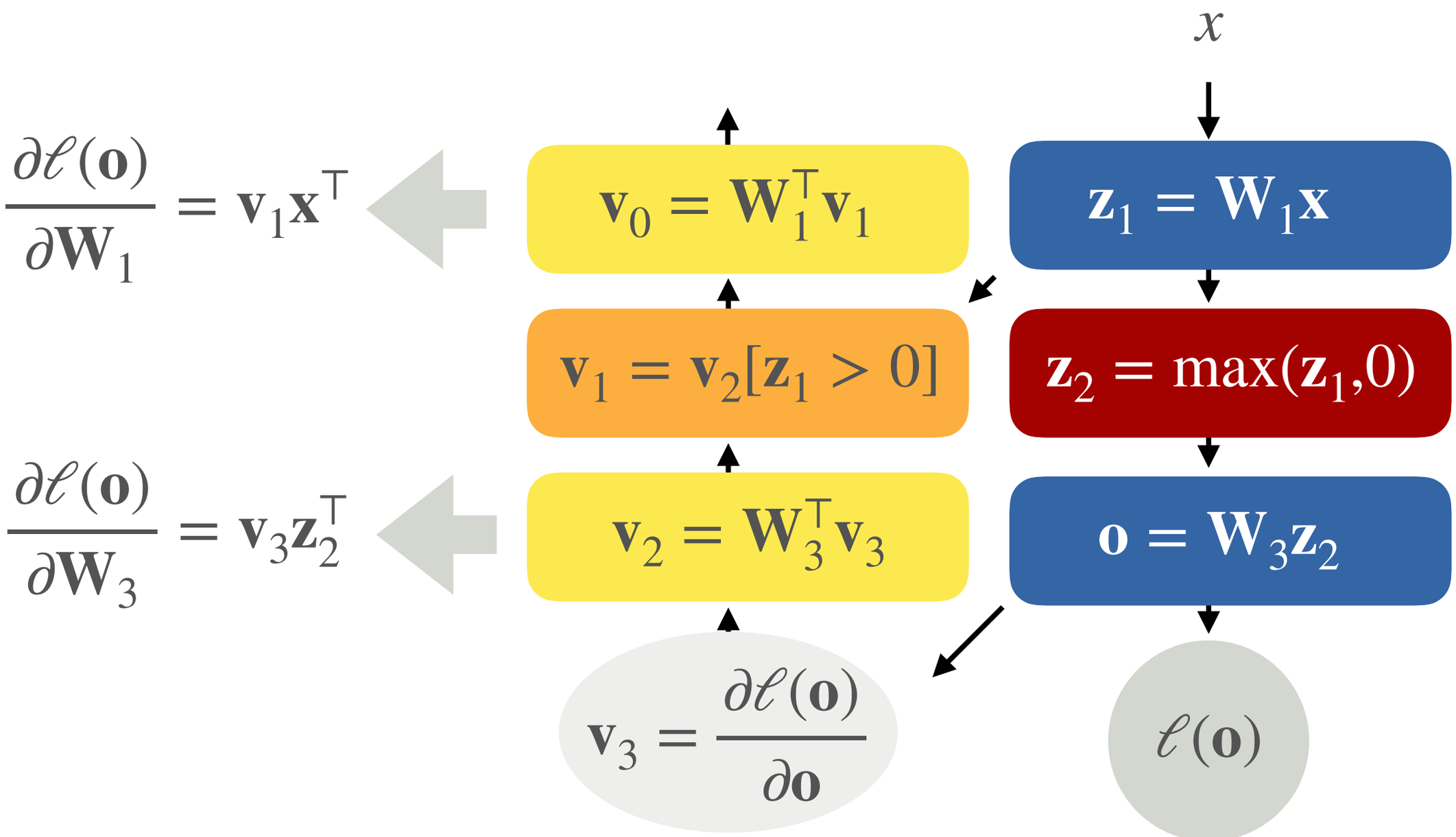
Idea 1: All zero



Idea 1: All zero

- Does not work
  - No gradient
- Saddle point

Idea 2: constant



Idea 2: constant

- Does not break symmetries

Solution

- Random initialization

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