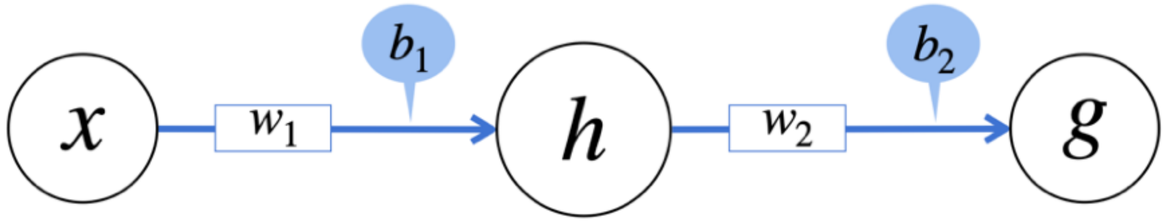
Backpropagation



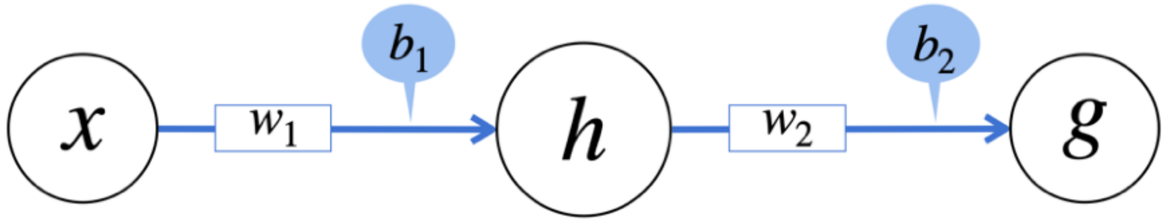
$$\frac{1}{N}\sum_{i}^{N}(y_i - g_i)^2 \implies \frac{1}{N}\sum_{i}^{N}(\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 \cdot (w_1 \cdot x_i + b_1) + b_2}_{\text{predicted}})^2$$

$$\frac{\partial \text{Loss}}{\partial w_1} = \frac{\text{Loss}}{\partial g} \cdot \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial w_1}$$

changing w_1 changes h, and changing h will change g, and changing g will change overall loss

$\implies$ we need the chain rule!

Backpropagation



$$\frac{1}{N}\sum_{i}^{N}(y_i - g_i)^2 \implies \frac{1}{N}\sum_{i}^{N}(\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 \cdot (w_1 \cdot x_i + b_1) + b_2)}_{\text{predicted}})^2$$

$$\frac{\partial \text{Loss}}{\partial w_1} = \underbrace{\underbrace{\frac{\text{Loss}}{\partial g}}_{-2(y_i - g_i)} \cdot \underbrace{\frac{\partial g}{\partial h}}_{w_2} \cdot \underbrace{\frac{\partial h}{\partial w_1}}_{x}}_{\text{sum over all observations}}$$

this is the first part of our gradient