Neural Networks and Optimization III

Dr. Parlett-Pelleriti

All We Want in Life is Gradients

$$\begin{bmatrix} \frac{\partial h}{\partial b_1} \\ \frac{\partial h}{\partial b_2} \\ \frac{\partial h}{\partial w_1} \\ \frac{\partial h}{\partial w_2} \end{bmatrix}$$

Backpropagation

The Chain-Rule

$$f(x) = cos(x)$$

$$g(x) = x^2$$

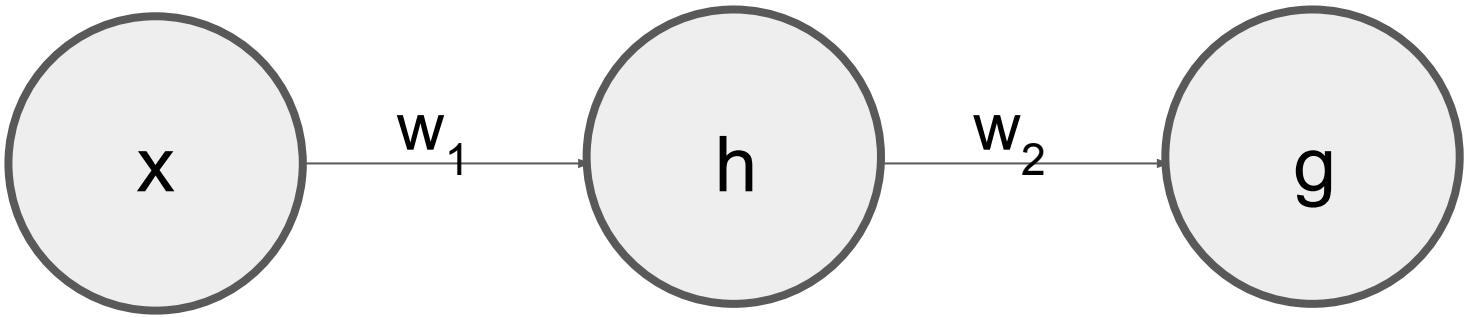
$$f(g(x)) = cos(x^2)$$

If we want to know how changing **x** affects **f(g(x))** we *first* need to think about how changing **x** affects **g(x)** *then* how changing **g(x)** affects **f(g(x))**

$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x}$$

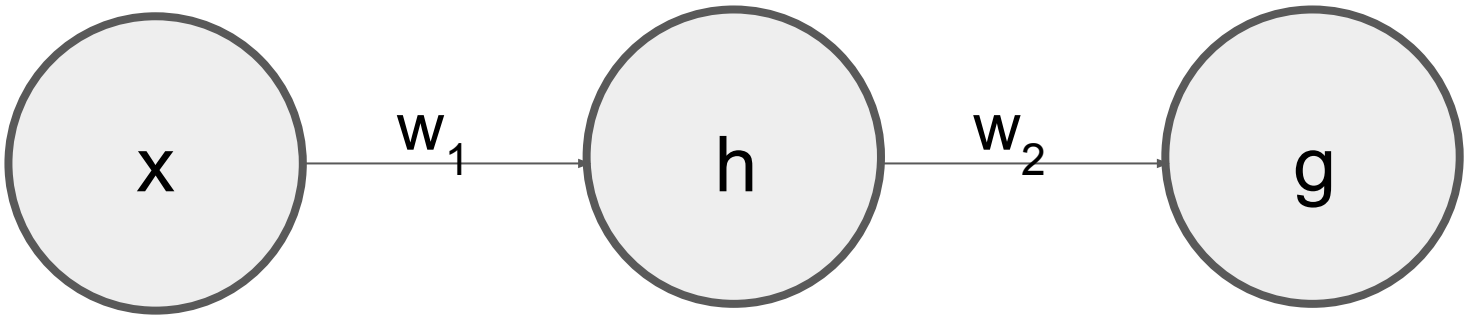
Backpropagation

$$h = w_1 * x + b_1$$
$$g = w_2 * h + b_2$$



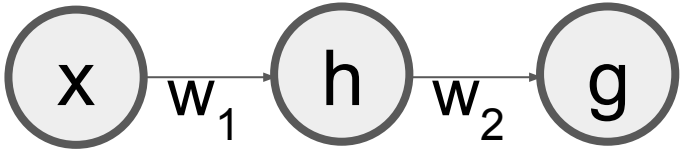
Backpropagation

$$h = w_1 * x + b_1$$
$$g = w_2 * h + b_2$$



$$\text{loss} = \sum_{i=0}^{N} (y_i - g_i)^2 \longrightarrow \text{loss} = \frac{1}{N} \sum_{i=0}^{N} (y_i - (w_2 * (w_1 * x_i + b_1) + b_2))^2$$

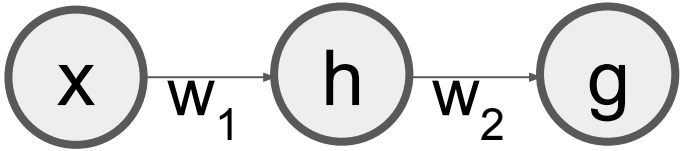
Backpropagation



$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

We don't have control over the actual values or predictor values, but we CAN change our **parameters**!

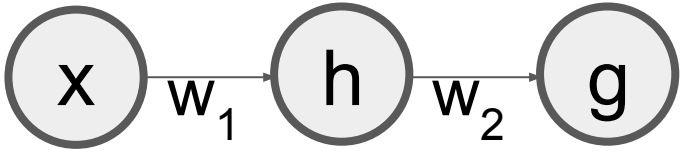
Backpropagation



$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

How should we change **w_1** to improve our loss?
How should we change **w_2** to improve our loss?
How should we change **b_1** to improve our loss?
How should we change **b_2** to improve our loss?

Backpropagation

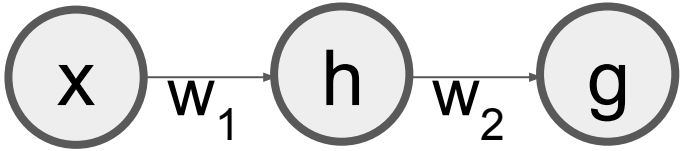


$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

This is what the gradient tells us

How should we change **w_1** to improve our loss?
How should we change **w_2** to improve our loss?
How should we change **b_1** to improve our loss?
How should we change **b_2** to improve our loss?

Backpropagation



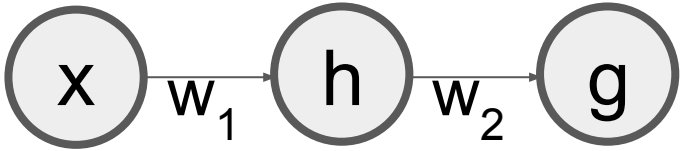
$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

How do we calculate the gradient??

- How should we change **w_1** to improve our loss?
- How should we change **w_2** to improve our loss?
- How should we change **b_1** to improve our loss?
- How should we change **b_2** to improve our loss?

Backpropagation

Backpropagation



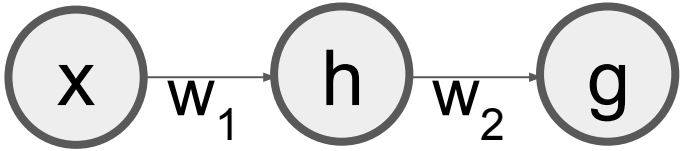
$$h = w_1 * x + b_1$$

$$g = w_2 * h + b_2$$

$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

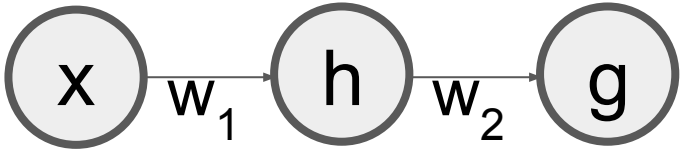
$$\frac{\partial \text{loss}}{\partial w_1} = \frac{\partial \text{loss}}{\partial g} * \frac{\partial g}{\partial h} * \frac{\partial h}{\partial w_1}$$

Backpropagation



$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

Backpropagation



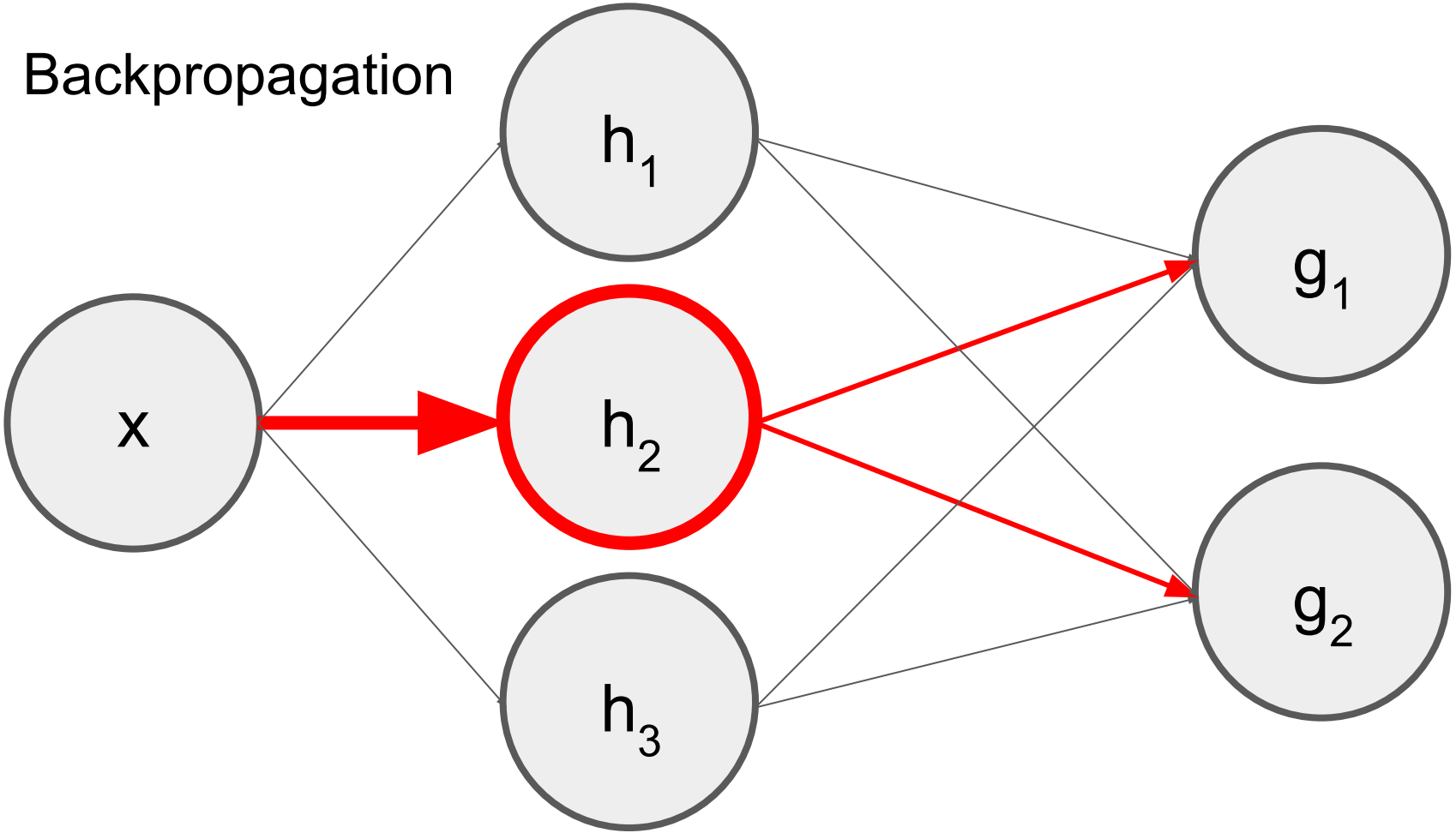
$$h = w_1 * x + b_1$$

$$g = w_2 * h + b_2$$

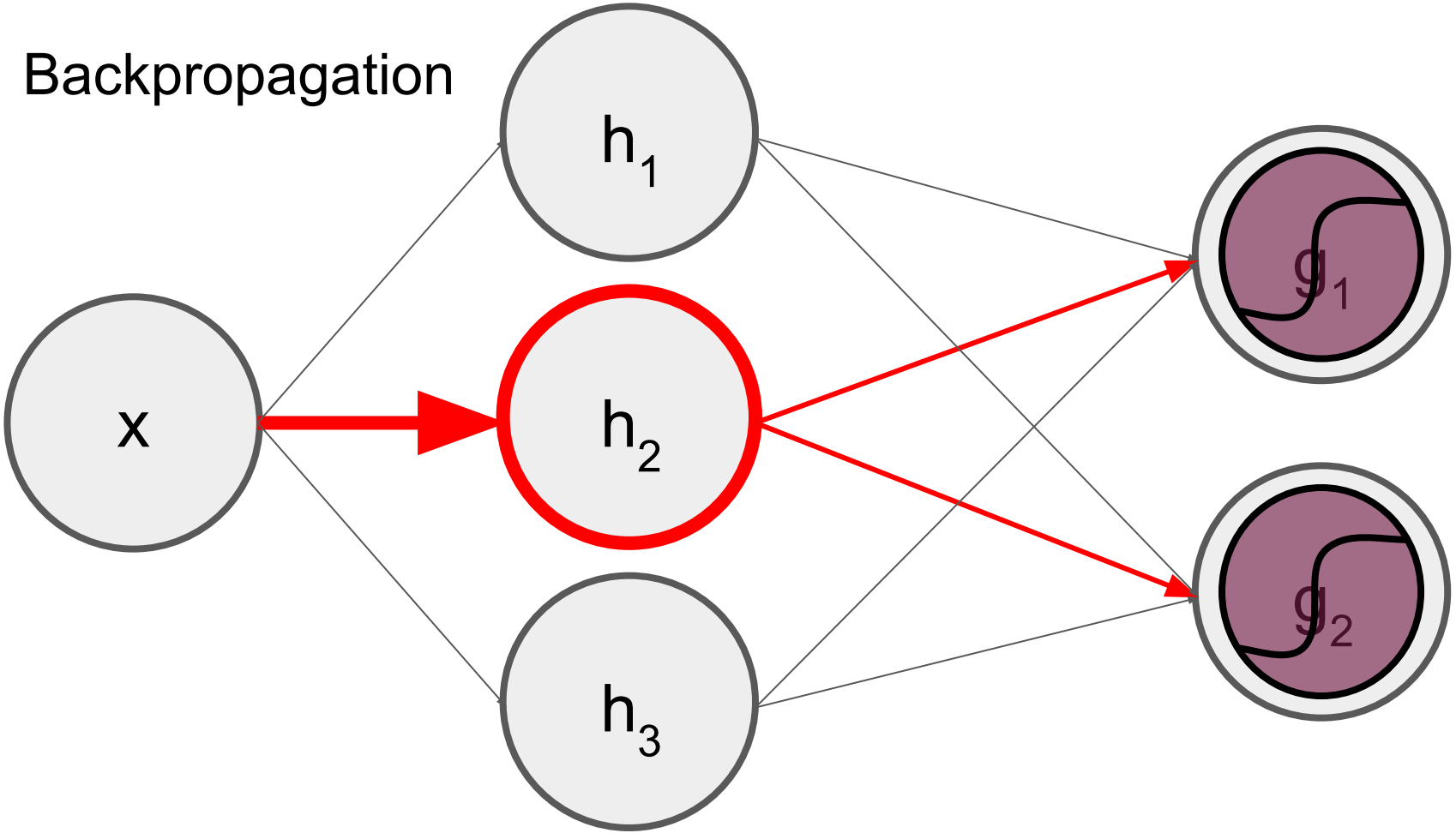
$$\frac{1}{N}\sum_{i=0}^{N}\left[\underbrace{y_i}_{\text{actual}} - \underbrace{(w_2 * (w_1 * x_{1i} + b1) + b_2)}_{\text{predicted}}\right]^2$$

$$\frac{\partial loss}{\partial b_1} = \frac{\partial loss}{\partial g} * \frac{\partial g}{\partial h} * \frac{\partial h}{\partial b_1}$$

Backpropagation

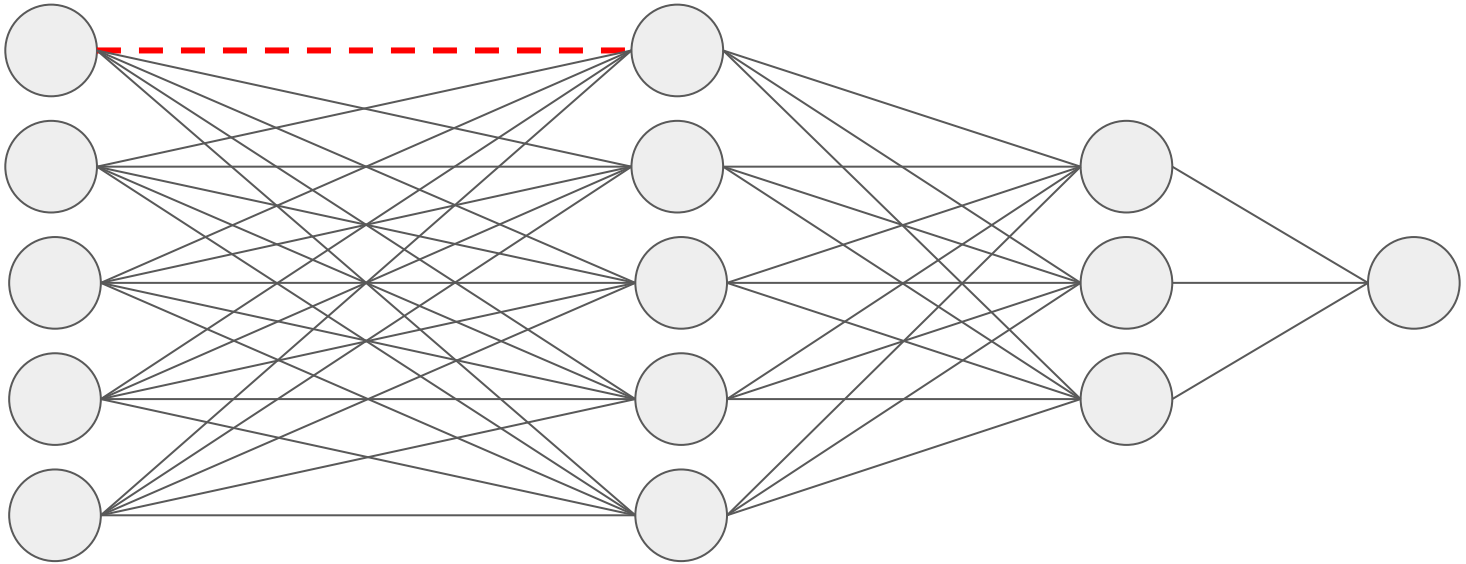


Backpropagation



Activity

When we change this weight, which other nodes does this affect?



Backpropagation (Vocab)

- Forward Pass
- Backward Pass
- Iteration
- Epoch

Back to Gradients

$$\begin{bmatrix} \frac{\partial h}{\partial b_1} \\ \frac{\partial h}{\partial b_2} \\ \frac{\partial h}{\partial w_1} \\ \frac{\partial h}{\partial w_2} \end{bmatrix}$$