Backprop.

Goal: compute $\frac{\partial J}{\partial \theta_i}$ for all θ_i

Key thing: chain rule for derivatives

$$(f \circ g)'(x) = g'(x) \times f' \circ g(x)$$

$$(f \circ g \circ h)'(x) = h'(x) \times g' \circ h(x) \times f' \circ g \circ h(x)$$

Illustration with a simple example

Let's introduce $z_{i+1} = w_i u_i + b_i$; $u_{i+1} = \sigma(z_{i+1})$

Loss function: $J(w_1, w_2, w_3, b_1, b_2, b_3) = \frac{1}{2}(\hat{y} - y)^2$ y: true target

Let's compute the gradients

$\Rightarrow \hat{y} = u_4 = w_3 u_3 + b_3$

$$\frac{\partial J}{\partial w_3} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_3} = (\hat{y} - y)\, u_3$$

$$\frac{\partial J}{\partial b_3} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b_3} = (\hat{y} - y) \times 1$$

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_2} = (\hat{y} - y) \underbrace{\frac{\partial \hat{y}}{\partial u_3}}_{w_3} \underbrace{\frac{\partial u_3}{\partial w_2}}$$

$$\underbrace{\frac{\partial u_3}{\partial z_3}}_{\sigma'(z_3)} \underbrace{\frac{\partial z_3}{\partial w_2}}_{u_2}$$

$$u_3 = \sigma(z_3)$$

$$z_3 = w_2 u_2 + b_2$$

$$\frac{\partial J}{\partial w_2} = \underbrace{(\hat{y} - y) w_3 \sigma'(z_3)} u_2$$

$$\frac{\partial J}{\partial b_2} = \underbrace{(\hat{y} - y) w_3 \sigma'(z_3)} \times 1$$

$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial u_3} \frac{\partial u_3}{\partial z_3} \frac{\partial z_3}{\partial u_2} \frac{\partial u_2}{\partial z_2} \frac{\partial z_2}{\partial w_1}$$

$$\underbrace{(\hat{y}-y)\, w_3\, \sigma'(z_3)}\; w_2\; \sigma'(z_2)\; u_1$$

$$(x)$$

$$\frac{\partial J}{\partial b_1} = \underbrace{(\hat{y}-y)\, w_3\, \sigma'(z_3)}\; w_2\; \sigma'(z_2) * 1$$