Backpropagation: Complete Derivation

Step 1: Forward Pass

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$$

$$a^{(l)} = \sigma(z^{(l)})$$

Step 2: Output Error

$$\delta^{(L)} = \nabla_a L \odot \sigma'(z^{(L)})$$

For MSE: $\delta^{(L)} = (a^{(L)} - y) \odot \sigma'(z^{(L)})$

Step 3: Backpropagate Error

$$\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)}) \odot \sigma'(z^{(l)})$$

Error flows backward through weights

Step 4: Compute Gradients

$$\frac{\partial L}{\partial W^{(l)}} = \delta^{(l)}(a^{(l-1)})^T$$

$$\frac{\partial L}{\partial b^{(l)}} = \delta^{(l)}$$

Step 5: Update Weights

$$W^{(l)} \leftarrow W^{(l)} - \eta \frac{\partial L}{\partial W^{(l)}}$$

$$b^{(l)} \leftarrow b^{(l)} - \eta \frac{\partial L}{\partial b^{(l)}}$$

Key insight: The chain rule allows efficient computation of all gradients in one backward pass