



deeplearning.ai

One hidden layer  
Neural Network

---

Gradient descent for  
neural networks

# Gradient descent for neural networks

Parameters:  $W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}$   
 $(n^{[1]}, n^{[0]})$   $(n^{[1]}, 1)$   $(n^{[2]}, n^{[1]})$   $(n^{[2]}, 1)$

$$n_x = n^{[0]}, \quad n^{[1]}, \quad \underline{n^{[2]} = 1}$$

Cost function:  $J(W^{[1]}, b^{[1]}, \underline{W^{[2]}}, \underline{b^{[2]}}) = \frac{1}{m} \sum_{i=1}^m \ell(\hat{y}, y)$   
 $\uparrow \quad \uparrow \quad \uparrow a^{[2]}$

Gradient descent:

→ Repeat {

→ Compute predicts  $(\hat{y}^{(i)}, i=1, \dots, m)$

$$\underline{dW^{[1]}} = \frac{\partial J}{\partial W^{[1]}}, \quad \underline{db^{[1]}} = \frac{\partial J}{\partial b^{[1]}}, \dots$$

$$W^{[1]} := W^{[1]} - \alpha dW^{[1]}$$

$$b^{[1]} := b^{[1]} - \alpha db^{[1]}$$

$$\downarrow \quad W^{[2]} := \dots \quad b^{[2]} := \dots$$

# Formulas for computing derivatives

Forward propagation:

$$z^{[1]} = w^{[1]}x + b^{[1]}$$

$$A^{[1]} = g^{[1]}(z^{[1]}) \leftarrow$$

$$z^{[2]} = w^{[2]}A^{[1]} + b^{[2]}$$

$$A^{[2]} = g^{[2]}(z^{[2]}) = \underline{\sigma}(z^{[2]})$$

Back propagation:

$$dz^{[2]} = A^{[2]} - Y \leftarrow$$

$$dw^{[2]} = \frac{1}{m} dz^{[2]} A^{[1]T}$$

$$db^{[2]} = \frac{1}{m} \text{np.sum}(dz^{[2]}, \text{axis}=1, \text{keepdims}=\text{True})$$

$$dz^{[1]} = \underbrace{w^{[2]T} dz^{[2]}}_{(n^{[1]}, m)} \star \underbrace{g^{[1]'}(z^{[1]})}_{\text{element-wise product}} \quad (n^{[1]}, m)$$

$$dw^{[1]} = \frac{1}{m} dz^{[1]} x^T$$

$$db^{[1]} = \frac{1}{m} \text{np.sum}(dz^{[1]}, \text{axis}=1, \text{keepdims}=\text{True})$$

$\underline{(n^{[1]}, 1)}$        $(n^{[1]}, )$        $\text{reshape} \uparrow$

$$Y = [y^{(1)} \ y^{(2)} \ \dots \ y^{(m)}]$$

$$(n^{[1]}) \leftarrow$$

$$\downarrow (n^{[2]}, 1) \leftarrow$$