deeplearning.ai

One hidden layer Neural Network

Backpropagation intuition (Optional)

Computing gradients

Logistic regression

$$\begin{array}{l} x \\ w \\ b \end{array} \to \boxed{z = w^T x + b} \to \boxed{a = \sigma(z)} \to \boxed{\mathcal{L}(a, y)}$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} = da \cdot \frac{\partial}{\partial z}(\sigma(z))$$

$$= da \cdot \frac{\partial}{\partial z} g(z)$$

$$dz = da \cdot g'(z)$$

$$\frac{\partial L}{\partial a} = da = -\frac{y}{a} + \frac{1-y}{1-a}$$

$$\frac{d}{da} L(a,y) = \frac{\partial}{\partial a}[-y \log a - (1-y)\log(1-a)]$$

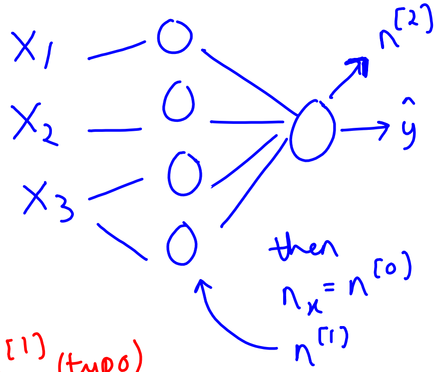
ie, $da = -\frac{y}{a} + \frac{1-y}{1-a}$

$$\partial w = dz \cdot x$$

$$\partial b = dz$$

$$\frac{\partial L}{\partial w} = \underbrace{\frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}}_{dz} \cdot \underbrace{\frac{\partial z}{\partial w}}_{= x}$$

Neural network gradients



$W^{[2]}$ $dw^{[2]}$

$b^{[2]}$ $db^{[2]}$

x

$W^{[1]}$ $dw^{[1]}$

$b^{[1]}$ $db^{[1]}$

$z^{[1]} = W^{[1]}x + b^{[1]}$ → $a^{[1]} = \sigma(z^{[1]})$ → $z^{[2]} = W^{[2]}\cancel{x} + b^{[2]}$ → $a^{[2]} = \sigma(z^{[2]})$ → $\mathcal{L}(a^{[2]}, y)$

$a^{[1]}$ (typo)

$dz^{[1]}$

$\hookrightarrow \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial z^{[1]}}$

$dz^{[2]} \cdot W^{[2]} \cdot g^{[1]'}(z^{[1]})$

$\boxed{dz^{[1]} = dz^{[2]} \cdot W^{[2]} \cdot g^{[1]'}(z^{[1]})}$

$da^{[1]}$

$\hookrightarrow \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz^{[2]}} \cdot \underbrace{\frac{\partial z^{[2]}}{\partial a^{[1]}}}_{W^{[2]}}$

$dz^{[2]} \cdot W^{[2]}$

$\boxed{da^{[1]} = dz^{[2]} \cdot W^{[2]}}$

✓ $\boxed{dz^{[2]} = a^{[2]} - y}$

✓ $\boxed{dw^{[2]} = dz^{[2]} \cdot a^{[1]T}}$ ☆

$\hookrightarrow \frac{\partial L}{\partial w^{[2]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz^{[2]}} \cdot \underbrace{\frac{\partial z^{[2]}}{\partial w^{[2]}}}_{a^{[1]}}$

✓ $\boxed{db^{[2]} = dz^{[2]}}$ ☆

$da^{[2]}$

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

$db^{[1]} = dz^{[1]}$ ☆

See how the derivatives of w_1 & b_1 closely match w_2 & b_2

Dim Analysis

$W^{[2]} = (n^{[2]}, n^{[1]})$

$z^{[2]}$ & $dz^{[2]} \rightarrow (n^{[2]}, 1)$

$z^{[1]}$ & $dz^{[1]} \rightarrow (n^{[1]}, 1)$

∴ Say we check for $dz^{[1]}$'s dims

what is dim of $dz^{[1]}$? Same as that of $z^{[1]}$

$dz^{[1]} = \underbrace{\underbrace{dz^{[2]}}_{(n^{[2]}, 1)} \cdot \underbrace{W^{[2]}}_{(n^{[2]} \times n^{[1]})}}_{} \cdot \underbrace{g^{[1]'}(z^{[1]})}_{\boxed{(n^{[1]} \times 1)}}$

element wise product

So we do $W^{[2]T} \cdot dz^{[2]} \rightarrow (n^{[1]} \times n^{[2]}) \times (n^{[2]} \times 1) = \boxed{(n^{[1]} \times 1)}$

Andrew Ng

Summary of gradient descent

$$dz^{[2]} = a^{[2]} - y$$

$$dW^{[2]} = dz^{[2]} a^{[1]^T}$$

$$db^{[2]} = dz^{[2]}$$

$$dz^{[1]} = W^{[2]T} dz^{[2]} * g^{[1]\prime}(\mathrm{z}^{[1]})$$

$$dW^{[1]} = dz^{[1]} x^T$$

$$db^{[1]} = dz^{[1]}$$

We saw earlier as well

The left denotes 1 example, lets vectorize

$$Z^{[1]} = \begin{bmatrix} | & | & | & & | \\ z^{[1](1)} & z^{[1](2)} & z^{[1](3)} & \cdots & z^{[1](m)} \\ | & | & | & & | \end{bmatrix}$$

1st hidden layer for all m examples

nodes in the first hidden layer

m training examples →

$$A^{[1]} = g^{[1]}(Z^{[1]})$$

Summary of gradient descent

$$dz^{[2]} = a^{[2]} - y$$

$$dW^{[2]} = dz^{[2]} a^{[1]^T}$$

$$db^{[2]} = dz^{[2]}$$

$$dz^{[1]} = W^{[2]T} dz^{[2]} * g^{[1]\prime}(\mathrm{z}^{[1]})$$

This we saw as a $(n^{[1]}, 1)$ dim vector

$$dW^{[1]} = dz^{[1]} x^T$$

$$db^{[1]} = dz^{[1]}$$

$$dZ^{[2]} = A^{[2]} - Y$$

$$dW^{[2]} = \frac{1}{m} dZ^{[2]} A^{[1]^T}$$

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

$$dZ^{[1]} = W^{[2]T} dZ^{[2]} * g^{[1]\prime}(\mathrm{Z}^{[1]})$$

→ element wise product

↳ since this is done for all m examples ⇒ $n^{[1]} \times m$ dim

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} X^T$$

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

Andrew Ng