Calculus for Backpropagation

label

$$\text{Cost} = (a_L - y)^2$$

output activation

$$a_L = \sigma(w_L \cdot a_{L-1} + b_L)$$

bias of last node

Weight for last node

Activation of node before

main focus, to change cost

Chain Rule.

$$\frac{\partial \text{cost}}{\partial w_L} = \boxed{\frac{\partial (w_L \cdot a_{L-1} + b_L)}{\partial w_L} \cdot \frac{\partial a_L}{\partial (w_L \cdot a_{L-1} + b_L)} \cdot \frac{\partial \text{cost}}{\partial a_L}}$$

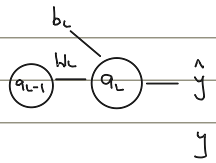
← change in numerator due to change in denominator.

change in cost due to change in weight

* w_L affects $(w_L \cdot a_{L-1} + b_L)$

$(w_L \cdot a_{L-1} + b_L)$ affects a_L

a_L affects Cost.



$$\text{cost} = (a_L - y)^2 \qquad u = a_L - y \qquad v = u^2$$

$$\frac{\partial \text{cost}}{\partial a_L} = 2(a_L - y) \cdot 1 \qquad \frac{du}{da_L} = 1 \qquad \frac{dv}{du} = 2u$$

$$\frac{\partial a_L}{\partial (w_L \cdot a_{L-1} + b_L)} = \sigma'(w_L \cdot a_{L-1} + b_L)$$

$$\frac{\partial (w_L \cdot a_{L-1} + b_L)}{\partial w_L} = a_{L-1}$$

change in $w_L \cdot a_{L-1} + b_L$ due to change in w_L depends strongly on a_{L-1}

$$\frac{\partial \text{cost}_0}{\partial w_L} = 2(a_L - y)\,\sigma'(w_L \cdot a_{L-1} + b_L) \cdot a_{L-1}$$

$$\frac{\partial C}{\partial w_L} = \frac{1}{n} \sum_{k=1}^{n} \frac{\partial C_k}{\partial w_L}$$

Full cost fn for all examples for that particular weight

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w_1} \\ \vdots \\ \frac{\partial C}{\partial w_L} \end{bmatrix}$$

* For Bias, perform same mathematics.