$$J = \frac{1}{2m}\sum (y-a^L)^2$$

$$\Delta a^L \sim \frac{\partial J}{\partial a^L} = \frac{1}{2m}\sum \frac{\partial}{\partial a^L}(y-a^L)^2$$

$$= \frac{1}{\cancel{2}m}\sum \cancel{2}\cdot(y-a^L)\frac{\partial \cancel{a^L}}{\partial \cancel{a^L}}$$

$$= \frac{1}{m}\sum (y-a^L) = \text{mean}(y-a^L)$$

$= \text{L1 loss} \rightarrow (\text{scalar})$

$$\underset{(n^L,\, n^{L-1})}{\Delta w^L} \sim \Delta a^L \cdot a^{L'} \cdot a^{L-1} = \underset{(n^L,m)}{\Delta z^L} \times \underset{(n^{L-1},m)^T}{a^{L-1\,T}}$$

$$\Delta b^L \sim db^L = \underset{(n^L,m)}{\Delta z^L}$$

$(\Delta z^L, a^{L-1})$

$$\underset{(n^L,m)}{\Delta z^L} \sim \frac{\partial J}{\partial z^L} = \frac{\partial J}{\partial a^L}\cdot\frac{\partial a^L}{\partial z^L} = \Delta a^L \cdot g^{L'}(z^L)$$

$$= \underset{(\text{scalar})}{\Delta a^L} \cdot \underset{(n^L,m)}{a^{L'}} \rightarrow (n^L, m)$$

$$\Delta w^{L-1} \sim dw^{L-1} = \Delta z^{L-1} \times a^{L-2\,T}$$

$$\Delta b^{L-1} \sim \Delta z^{L-1}$$

$$\underset{(n^{L-1},m)}{\Delta a^{L-1}} \sim \frac{\partial J}{\partial a^{L-1}} = \frac{\partial J}{\partial a^L}\cdot\frac{\partial a^L}{\partial z^L}\cdot\frac{\partial z^L}{\partial a^{L-1}}$$

$$\frac{\partial z^L}{\partial a^{L-1}} = \frac{\partial}{\partial a^{L-1}}(w^L a^{L-1} + b^L) = w^L$$

$$= \underset{(n^L,m)}{\Delta z^L} \cdot \underset{(n^L,n^{L-1})}{w^L}$$

$$= \underset{(n^{L-1},n^L)}{w^{L\,T}} \times \underset{(n^L,m)}{\Delta z^L} \rightarrow (n^{L-1}, m)$$

$$\underset{(n^{L-1},m)}{\Delta z^{L-1}} \sim \frac{\partial J}{\partial z^{L-1}} = \frac{\partial J}{\partial a^L}\cdot\frac{\partial a^L}{\partial z^L}\cdot\frac{\partial z^L}{\partial a^{L-1}}\cdot\frac{\partial a^{L-1}}{\partial z^{L-1}} = \underset{(n^{L-1},m)}{w^{L\,T}\Delta z^L} \cdot \underset{(n^{L-1},m)}{a^{L-1\,'}} \rightarrow (n^{L-1}, m)$$

$$\Delta a^{L-2} \sim \frac{\partial J}{\partial a^{L-2}} = \Delta z^{L-1}\frac{\partial z^{L-1}}{\partial a^{L-2}} = \Delta z^{L-1} w^{L-1} = w^{L-1\,T} \times \Delta z^{L-1}$$

⋮

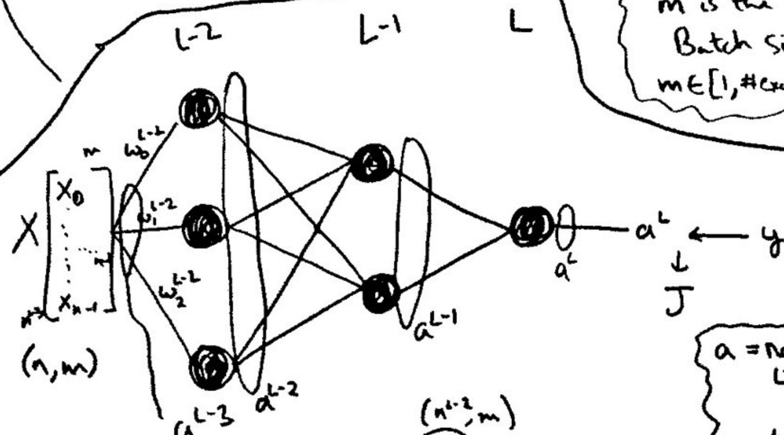
Backpropagation

Approach:

| Variable | Dictionary |
| --- | --- |
| $\Delta a^L$ | $y, a^L$ |
| | $\Delta a^L$ |
| $\Delta z^L$ | $a^{L'}$ |
| | $\Delta z^L$ |
| $\Delta b^L$ | $a^{L-1}$ |
| $\Delta w^L$ | ⊗ |
| $\Delta a^{L-1}$ | $\Delta z^L, w^L$ |
| | $\Delta a^{L-1}$ |
| $\Delta z^{L-1}$ | $a^{L-1\,'}$ |
| | $\Delta z^{L-1}$ |
| $\Delta b^{L-1}$ | $a^{L-2}$ |
| $\Delta w^{L-1}$ | ⊗ |

⋮

m is the Batch Size!
$m \in [1, \#\text{examples}]$




a = Non-Linearity
$g(z)$
$g'(z) = a'$

Gradient Dec.

$w^L = w^L - \eta\Delta w^L$
$b^L = b^L - \eta\Delta b^L$
$w^{L-1} = w^{L-1} - \eta\Delta w^{L-1}$
$b^{L-1} = b^{L-1} - \eta\Delta b^{L-1}$

⋮

$a^{L-3} = X \rightarrow (n^{L-3}, m)$
$w^{L-2} \rightarrow (n^{L-2}, n^{L-3})$
$b^{L-2} \rightarrow (n^{L-2}, m)$

$$z^{L-2} = \underbrace{w^{[L-2]}a^{[L-3]}}_{(n^{L-2},m)} + b^{[L-2]} \rightarrow (n^{L-2}, m)$$

$(n^{L-2}, n^{L-3}) \times (n^{L-3}, m) + (n^{L-2}, m)$

$$a^{L-2} = g^{L-2}(z^{L-2}) \rightarrow (n^{L-2}, m)$$

loop.

F-prop
↓
B-prop
↓
Grad Dec

output Loss/Cost

In general,

$dw^L \rightarrow w^L \rightarrow (n^L, n^{L-1})$
$db^L \rightarrow b^L \rightarrow (n^L, m)$
$dz^L \rightarrow z^L \rightarrow (n^L, m)$
$da^L \rightarrow a^L \rightarrow (n^L, m)$
$a^{L-1} \rightarrow (n^{L-1}, m)$

$a^L \rightarrow (n^L, m)$ $\quad y \rightarrow (n^L, m)$

$J \rightarrow (1, m) \Rightarrow J \rightarrow \frac{\sum J}{m}$

for L2 loss,

$$J = \frac{1}{2}(y - a^L)^2$$

$$J = \frac{1}{2}\frac{1}{m}\sum (y - a^L)^2$$

[redacted]