Reflection

— John Quinn's talk

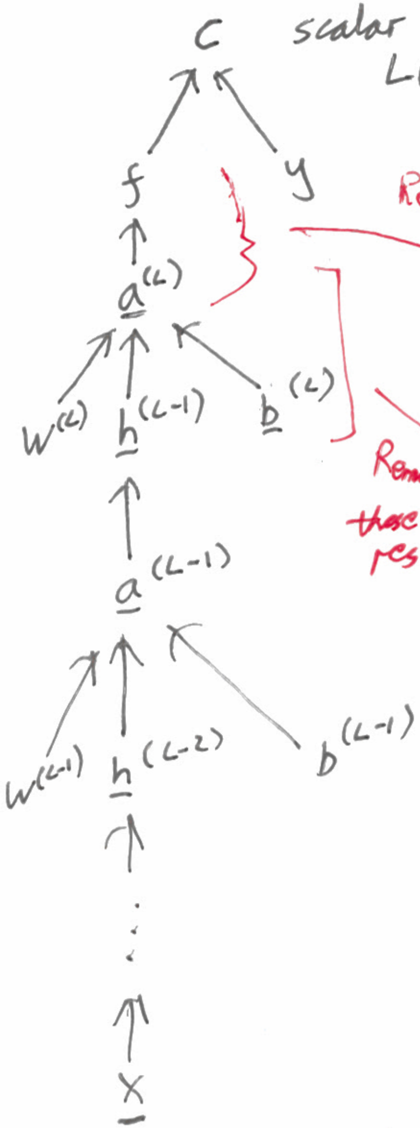
Should we jump straight to a ConvNet?

Why? Why not?

What general messages were there?

Getting gradients – Reverse-mode differentiation

Backpropagation



Remember

Strategy:

$$\bar{f} = \frac{\partial c}{\partial f}$$

$$\overline{a_i^{(L)}} = \frac{\partial c}{\partial a_i^{(L)}}$$

Remember these results

$\overline{W}^{(L)}$ $\quad$ $\overline{\underline{h}}^{(L-1)}$ $\quad$ $\overline{\underline{b}}^{(L)}$

$$\overline{W_{ij}^{(L)}} = \frac{\partial c}{\partial W_{ij}^{(L)}}$$

For every intermediate θ

get $\bar{\theta} = \dfrac{\partial c}{\partial \theta}$

Also:

MLPR 2017 L15 (2) $\qquad$ MLPR 2017 L13 (4)

Reverse-Mode Differentiation

Notation: Giles 2008

We're computing a scalar cost, c

We'll intermediate terms Z some matrix

We want $\bar{Z}$ same size as Z

$$\bar{Z}_{ij} = \frac{\partial c}{\partial Z_{ij}}$$

Start at end

$$c = (f-y)^2 \quad \text{(example)}$$

$$\bar{f} = \frac{\partial c}{\partial f} = 2(f-y)$$

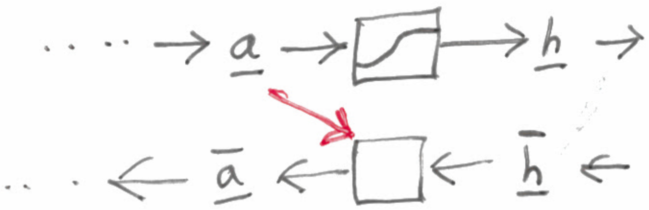
If we had $\underset{N\times 1}{\underline{f}}$ & $\underset{N\times 1}{\underline{y}}$

$$\bar{\underline{f}} = 2(\underline{f}-\underline{y})$$

Backpropagate

Compute $\bar{Z}$ for parents of quantities we have derivative signals for.

Elementwise Functions



$h_k = g(a_k)$

upstream computation of cost c

$$\bar{a}_i = \frac{\partial c}{\partial a_i} = \frac{\partial c}{\partial h_i}\frac{\partial h_i}{\partial a_i}$$

$$= \bar{h}_i \, g'(a_i)$$

$$g'(z) = \left.\frac{\partial g(a)}{\partial a}\right|_{a=z}$$

$$\bar{\underline{a}} = \bar{\underline{h}} \odot g'(\underline{a})$$

Hadamard product

.* Matlab

* Python

Example: Matrix Multiplication

n^2 numbers in an $n \times n$ matrix

$O(n^3)$ cost to multiply.

$$
\begin{tikzcd}
\cdots \arrow[r] & X \arrow[dr] & & & \\
\cdots \arrow[r] & Y \arrow[r] & \boxed{@} \arrow[r] & Z \arrow[r] & \cdots
\end{tikzcd}
$$

$$Z = XY$$

$$
\begin{tikzcd}
\cdots & \bar{X} \arrow[l] & \bar{Z} \arrow[l] \arrow[dl] & \cdots \arrow[l] \\
\cdots & \bar{Y} \arrow[l] & &
\end{tikzcd}
$$

$$Z_{mn} = \sum_p X_{mp} Y_{pn}$$

$$\frac{\partial c}{\partial X_{ij}} = \sum_{m,n} \underbrace{\frac{\partial c}{\partial Z_{mn}}}_{\bar{Z}_{mn}} \underbrace{\frac{\partial Z_{mn}}{\partial X_{ij}}}_{\delta_{mi} Y_{jn}} = \sum_n \bar{Z}_{in} Y_{jn}$$

$O(n^4)$

$$\boxed{\bar{X} = \bar{Z} Y^\top}$$

$$z_{mn} = x_{m1}y_{1n} + x_{m2}y_{2n} + \dots$$

$$\dots \underline{x_{mj}y_{jn}} + \dots$$