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

Scalar cost: $L(f, y)$

Remember strategy:

For every intermediate $\theta$

get $\theta = \frac{\partial c}{\partial \theta}$

Also:
MLPR 2017 L15 (2)
Reverse-Mode Differentiation

Notation: Giles 2008

We're computing a scalar cost, $C$

We'll intermediate terms $Z$ some matrix

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

$\overline{Z}_{ij} = \frac{\partial C}{\partial Z_{ij}}$

Start at end

$C = (f-y)^2$ (example)

$\overline{f} = \frac{\partial C}{\partial f} = 2(f-y)$

If we had $f$ & $y$ $\in \mathbb{R}^{N \times 1}$

Backpropagate

$\overline{f} = 2(f-y)$

Compute $\overline{Z}$ for parents of quantities we have derivative signals for.
Elementwise Functions

\[ \cdots \rightarrow a \rightarrow \square \rightarrow h \rightarrow \cdots \quad h_k = g(a_k) \]

\[ \xleftarrow{\bar{a}} \quad \xleftarrow{\bar{h}} \quad \xleftarrow{\cdots \text{upstream computation of cost } c} \]

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

\[ = \bar{h}_i \cdot g'(a_i) \]

\[ \bar{a} = \bar{h} \odot g'(a) \]

↑ Hadamard product
* Matlab
* Python
Example: Matrix Multiplication

\[ \cdots \times \rightarrow \odot \rightarrow \cdots \]

\[ \cdots \rightarrow \bar{z} \rightarrow \cdots \]

\[ z = x \cdot y \]

\[ z_{mn} = \sum_p x_{mp} y_{pn} \]

\[ \frac{\partial c}{\partial x_{ij}} = \sum_{m,n} \frac{\partial c}{\partial z_{mn}} \frac{\partial z_{mn}}{\partial x_{ij}} = \sum_{n} E_{in} Y_{jn} \]

\[ X = \bar{z} Y^T \]
\[ z_{mn} = X_{mi} y_{in} + X_{m2} y_{zn} + \ldots + X_{mj} y_{jn} + \ldots \]