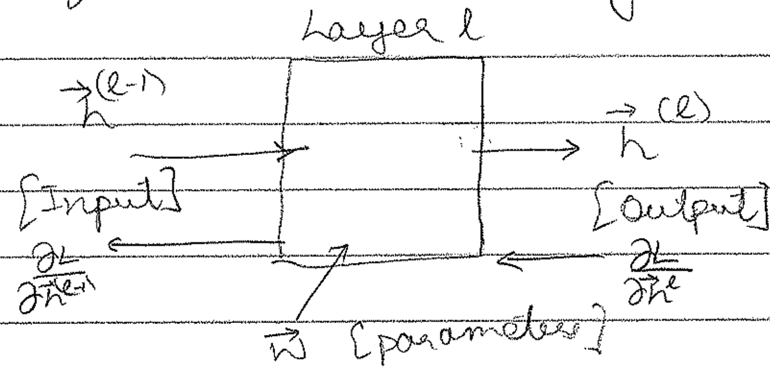
9/03/15



# MLPs + CNNs

(1) Recall from last time, key abstraction:




$$\vec{h}^{(l)} = g(\vec{h}^{l-1}, \vec{w})$$

Loss $L = f(\vec{h}^{l})$

Abstractly: (a) $\dfrac{\partial L}{\partial \vec{in}} = \dfrac{\partial L}{\partial \vec{out}} \cdot \dfrac{\partial \vec{out}}{\partial \vec{in}}$

(b) $\dfrac{\partial L}{\partial \vec{w}} = \dfrac{\partial L}{\partial \vec{out}} \cdot \dfrac{\partial \vec{out}}{\partial \vec{in}}$

[A bit more]
Concretely:

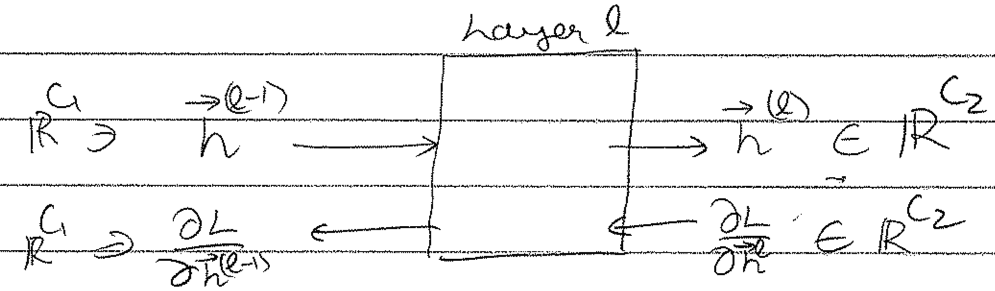
(a) $$\frac{\partial L}{\partial \vec{h}^{(l-1)}} = \frac{\partial L}{\partial \vec{h}^{l}} \times \frac{\partial \vec{h}^{l}}{\partial \vec{h}^{l-1}}$$

$$[\quad]_{1\times|\vec{h}^{l-1}|} \quad [\quad]_{1\times|\vec{h}^{l}|} \quad [\quad]_{|\vec{h}^{l}|\times|\vec{h}^{l-1}|}$$

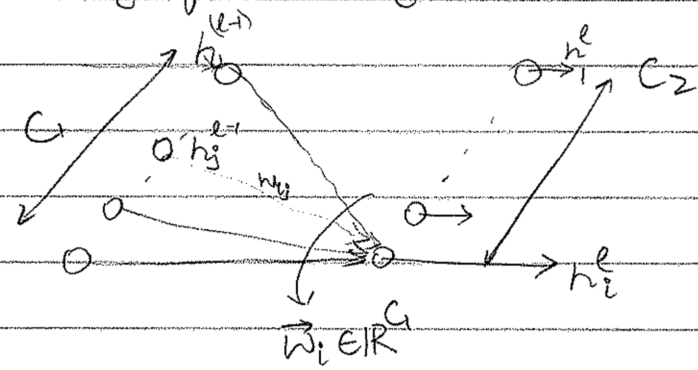
(b) $$\frac{\partial L}{\partial \vec{w}} = \frac{\partial L}{\partial \vec{h}^{(l)}} \times \frac{\partial \vec{h}^{l}}{\partial \vec{w}}$$

$$[\quad]_{1\times|\vec{w}|} \quad [\quad]_{1\times|\vec{h}^{l}|} \quad [\quad]_{|\vec{h}^{l}|\times|\vec{w}|}$$

## ② Multi-Layer Perceptrons / Fully-Connected Layer / Inner-Product Layer




Another way of drawing this




[Why draw in 3D? You'll see. Wait for CNNs!]

Forward-Pass:

$$h_i^{\ell} = \sum_{j=1}^{C_1} w_{ij}\, h_j^{(\ell-1)}$$

$$= \vec{w}_i^{T}\, \vec{h}^{(\ell-1)}$$

Let's vectorize some more

$$\vec{h}^{\ell} = W^{(\ell)}\, \vec{h}^{(\ell-1)}$$

$$\begin{bmatrix} h_1^{\ell} \\ \vdots \\ \vdots \\ h_{C_2}^{\ell} \end{bmatrix} = \begin{bmatrix} \leftarrow w_1^T \rightarrow \\ \vdots \\ \leftarrow w_i^T \rightarrow \\ \vdots \\ \leftarrow w_{C_2} \rightarrow \end{bmatrix} \begin{bmatrix} h_1^{\ell-1} \\ \vdots \\ \vdots \\ h_{C_1}^{\ell-1} \end{bmatrix}$$

So all of F-PASS = 1 matrix-mult

Why did we write this as matrix mult?

1. Because it's cooler!
2. Gradients, my dear Watson, Gradients...

$$\vec{h}^{(\ell)} = W^{(\ell)} \vec{h}^{(\ell-1)}$$

$$\Rightarrow \frac{\partial \vec{h}^{(\ell)}}{\partial \vec{h}^{\ell-1}} = W^{(\ell)}$$

[Now, isn't that beautiful?]

$$\frac{\partial \vec{out}}{\partial \vec{in}}$$

What about $\frac{\partial \vec{out}}{\partial \vec{w}}$?

Simple $\quad h_i^{\ell} = \vec{w}_i^{T} \vec{h}^{\ell-1}$

$$\Rightarrow \frac{\partial h_i^{\ell}}{\partial \vec{w}_i} = \vec{h}^{(\ell-1)\,T}$$

Now put it all together:

(a) $$\frac{\partial L}{\partial \vec{h}^{\ell-1}} = \frac{\partial L}{\partial \vec{h}^{\ell}} \frac{\partial \vec{h}^{\ell}}{\partial \vec{h}^{\ell-1}} = \boxed{\frac{\partial L}{\partial \vec{h}^{\ell}} \times W^{(\ell)}}$$ Very nice

So B-PASS(a) $\equiv$ 1 matrix mult

$\boxed{\text{Left multiply} \equiv \text{gradient} \quad \text{Right multiply} \equiv \text{output}}$

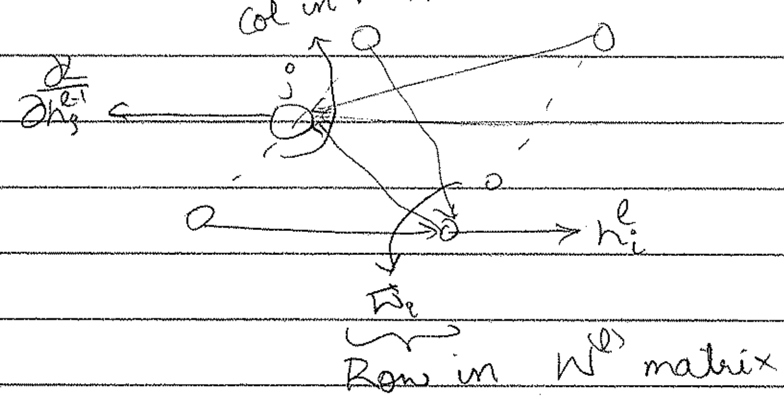
(b) $$\frac{\partial L}{\partial \vec{w}_i} = \frac{\partial L}{\partial h_i^{\ell}} \frac{\partial h_i^{(\ell)}}{\partial \vec{w}_i} = \underbrace{\frac{\partial L}{\partial h_i}}_{\text{scalar}} \times \underbrace{\vec{h}^{(\ell-1)\,T}}_{[\quad]_{1\times c_1}}$$

So $W^{(l)} = W^{(l)} - \eta \dfrac{\partial L}{\partial W^{(l)}}$

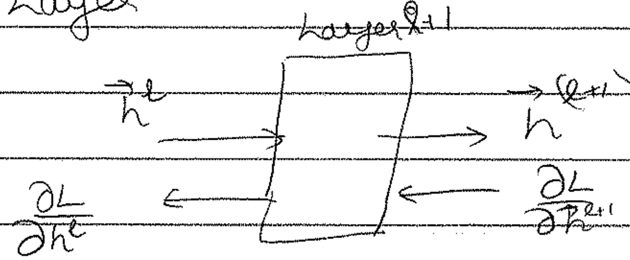
where

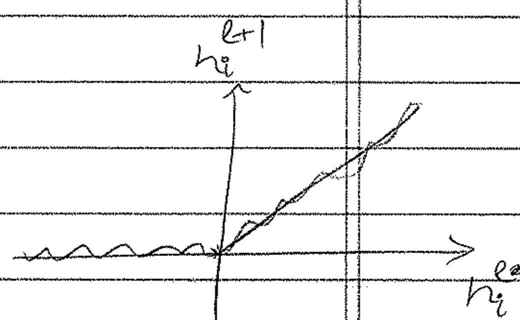
$$\frac{\partial L}{\partial W^{(l)}} = \begin{bmatrix} \longleftarrow \dfrac{\partial L}{\partial h_1^l} \cdot h^{(l-1)T} \longrightarrow \\ \vdots \\ \longleftarrow \dfrac{\partial L}{\partial h_{c_2}^l} \cdot h^{(l-1)T} \longrightarrow \end{bmatrix}$$

One more intuition




③ ReLU Layer




F-PASS $\quad h_i^{(l+1)} = \max\{0, h_i^l\}$

B-PASS [No params] $\quad \dfrac{\partial h_i^{l+1}}{\partial h_i^l} = \begin{cases} +1 & \text{if } h_i^l \geq 0 \\ 0 & \text{else} \end{cases}$

$$\Rightarrow \frac{\partial L}{\partial h_i^l} = \frac{\partial L}{\partial h_i^{l+1}} \times \mathbb{1}\{h_i^l \geq 0\}$$