Exercise 1 / Blackboard 1

1

(a) $\frac{\partial L}{\partial \omega^{(1)}_{1135}} = -(t_0 - \hat{y}_0)\frac{\partial \hat{y}_0}{\partial \omega^{(1)}_{1135}}$ ← $\hat{y}$ from slide

a_{ijk} from slide

$$= -(t_0 - \hat{y}_0)\sum_{ijk} \omega^{(2)}_{ijk}\, \sigma'(a_{ijk})\frac{\partial a_{ijk}}{\partial \omega^{(1)}_{1135}}$$

$$= -(t_0 - \hat{y}_0)\sum_{ijk} \omega^{(2)}_{ijk}\, \sigma'(a_{ijk})\, I_{ij3}$$

(b) $x^{(1)\,(out)}_{ijk} = \max\{x_{ijk}, x_{i+1,jk}, x_{ij+1k}, x_{i+1,j+1}\} = \sigma(a_{i^*(i,j)\,j^*(i,j)})$

$$\frac{\partial L}{\partial \omega^{(1)}_{1135}} = -(t_0 - \hat{y}_0)\sum_{i,jk} \omega^{(2)}_{ijk}\, \sigma'(a_{i^*(i,j)\,j^*(i,j)}) \cdot I_{i^*(i,j),\,j^*(i,j)}$$

$i^*(i,j)$ may be the same for several locations (i,j)

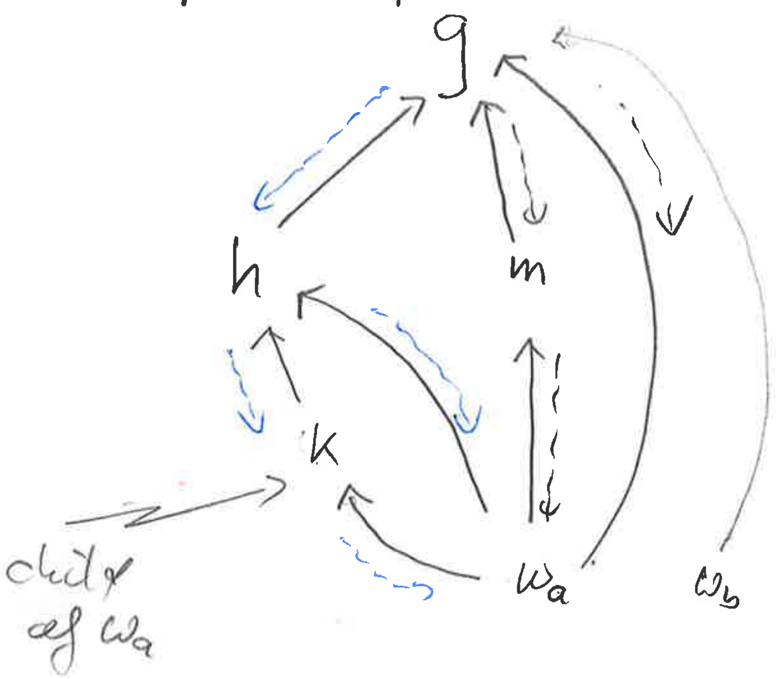
focussing effect of gradient:
it always selects the winner

"filter update focuses on the location that fits best the filter"

Example: Autodiff

$$f(w_a) = g\left[h(k(w_a), w_a), m(w_a), w_a, w_b\right]$$

forward path



w_a has 4 children

backward path (in red)

$$\frac{\partial f}{\partial w_a} = \frac{\partial g}{\partial h}\left[\frac{\partial h}{\partial k}\frac{\partial k}{\partial w_a} + \frac{\partial h}{\partial w_a}\right] + \frac{\partial g}{\partial m}\frac{\partial m}{\partial w_a} + \frac{\partial g}{\partial a}$$

sum over children of w_a

($\Rightarrow$ 4 terms in summation)

Blackboard 3:

Outer - Product representation

$$w_{xyck} = w_{xyk} \cdot \underset{\substack{\uparrow \\ \text{color / depth}}}{w_{ck}}$$

$$a_{ijk} = b_k + \sum_x \sum_y \overset{\text{depth}}{\sum_{c=1}^{d}} I_{i+x-1,j+y-1,c} \cdot w_{xyck}$$

$$= b_k + \sum_x \sum_y \left[\sum_{c=1}^{d} I_{i+x-1,j+y-1,c} \cdot w_{ck} \right] \cdot w_{xyk}$$

some over depth:

$$\left\| \tilde{I}_{i,j} = \sum_{c=1}^{d} I_{i,j,c} \cdot w_{ck} \right\|$$

average over color / depth

Skip connections / Resnet

$$x_i^{(n+2)} = \left[x_i^{(n)} + F_i(\vec{x}^{(n)}) \right]$$

$$F_i(\vec{x}_n) = \sum_j \omega_{ij}^{(n+2)} \, g\left(\sum_k \omega_{jk}^{(n+1)} x_k^{(n)} \right)$$

derivative:

$$\frac{\partial x_i^{(n+2)}}{\partial x_k^{(n)}} = \delta_{ik} + \frac{\partial F_i}{\partial x_k^{(n)}}$$

identity
(layer
(n+1) not there)

$\sim g'$

some paths vanish,
bias problem
linearity problem

initialize with small weights $\omega_{ij}^{(n+2)}$, $\omega_{jk}^{(n+1)}$

$\Rightarrow$ gradient close to unity

avoids vanishing gradient problem