Aa2) 1.)

Beh: $\max \left| \frac{\partial h_n}{\partial w_i} \right| = |h_{n-1}| \cdot \frac{1}{4}$

Bew: Sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$

Kettenregel

$$\sigma'(x) \overset{\swarrow}{=} \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} = \sigma(x) \cdot \frac{1+e^{-x}-1}{1+e^{-x}}$$

$$= \sigma(x) \cdot (1-\sigma(x))$$

Substitution: $\sigma(x) = t$, $t \in (0,1)$ da $0 < \frac{1}{1+e^{-x}} < 1$

$\Rightarrow$ $g(t) = t(1-t)$ $\Rightarrow$ $g'(t) = 1-2t \overset{!}{=} 0$

$\Rightarrow$ $1-2t = 0$ $\Rightarrow$ $t = \frac{1}{2}$

$g''(t) = -2 < 0$ $\Rightarrow$ $t = \frac{1}{2}$ ist ein Maximum von $g(t)$

$g(\frac{1}{2}) = \frac{1}{2} \cdot (1-\frac{1}{2}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\sigma'(x) = \sigma(x)(1-\sigma(x)) = g(\sigma(x))$

$\Rightarrow$ $\max\limits_{x} \sigma'(x) = \max\limits_{t \in (0,1)} g(t) = g(\frac{1}{2}) = \frac{1}{4}$

$\Rightarrow$ $\forall j$: $|\sigma'(w_j h_{j-1})| \leq \frac{1}{4}$ und nach Annahme $|w_j| < 1$

Laut Aa 1) gilt $\frac{\partial h_n}{\partial w_i} = \prod\limits_{j=i+1}^{n} [\sigma'(w_j h_{j-1}) w_j] \cdot \sigma'(w_i h_{i-1}) \cdot h_{i-1}$

$$\Rightarrow \left| \frac{\partial h_n}{\partial w_i} \right| \leq \left(\prod_{j=i+1}^{n} \frac{1}{4} \cdot \overbrace{|w_j|}^{<1} \right) \cdot \frac{1}{4} |h_{i-1}|$$

$$\leq \left(\frac{1}{4}\right)^{n-(i+1)} \cdot 1 \cdot \frac{1}{4} \cdot |h_{i-1}|$$

$$= \frac{1}{4}^{n-i+1} |h_{i-1}|$$

Maximalwert über i findet man für $i=n$: $|h_{n-1}| \left(\frac{1}{4}\right)$. (Für große Tiefe n geht der Term $\left(\frac{1}{4}\right)^{n-i+1}$ stark gegen 0.

□

Aa2) 2.)

ReLU $\sigma(x) = \max(0, x)$

$$\sigma'(x) = \begin{cases} 1 & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

$$\Rightarrow \frac{\partial h_n}{\partial w_i} \underset{\uparrow}{=} \left(\prod_{j=i+1}^{n} w_j \right) h_{i-1}$$

Formel aus Aa 1)

Da $|w_j| < 1$ schrumpft $\prod\limits_{j=i+1}^{n} w_j$, aber es gibt keinen zusätzlichen Faktor, der schrumpft.

Der Maximalwert über i liegt bei $i=n$: $\frac{\partial h_n}{\partial w_n} = h_{n-1}$