1.1.

$$L = \frac{\sum^{N} (y_i - \hat{y}_i)^2}{N}$$

$$\hat{y} = h_2 W_3 + b_3$$

$$h_2 = g(W_2 h_1 + b_2)$$

$$h_1 = g(W_1 x + b_1)$$

$$\frac{\partial L}{\partial \hat{y}} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y})$$

$$\frac{\partial \hat{y}}{\partial W_3} = h_2 \qquad \frac{\partial \hat{y}}{\partial b_3} = 1 \qquad \frac{\partial \hat{y}}{\partial h_2} = W_3$$

$$\frac{\partial h_2}{\partial W_2} = h_1 \frac{\partial g(u)}{\partial u} \qquad \frac{\partial h_2}{\partial b_2} = \frac{\partial g(u)}{\partial u} \qquad \frac{\partial h_2}{\partial h_1} = W_2 \frac{\partial g(u)}{\partial u}$$

$$\frac{\partial h_1}{\partial W_1} = x \frac{\partial g(u)}{\partial u} \qquad \frac{\partial h_1}{\partial b_1} = \frac{\partial g(u)}{\partial u}$$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial W_3} \qquad \frac{\partial L}{\partial b_3} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b_3}$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial W_2} \qquad \frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial b_2}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial W_1} \qquad \frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial b_1}$$

$\Rightarrow$

$$\frac{\partial L}{\partial W_3} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y}) h_2 \qquad \frac{\partial L}{\partial b_3} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y})$$

$$\frac{\partial L}{W_2} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y}) W_3 h_1 \frac{\partial g(u)}{\partial u} \qquad \frac{\partial L}{\partial b_2} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y}) W_3 \frac{\partial g(u)}{\partial u}$$

$$\frac{\partial L}{W_1} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y}) W_3 W_2 x \left[\frac{\partial g(u)}{\partial u}\right]^2 \qquad \frac{\partial L}{\partial b_1} = -\frac{2}{N} \sum^{N} \hat{y}(y - \hat{y}) W_3 W_2 \left[\frac{\partial g(u)}{\partial u}\right]^2$$

1.2.

$$W_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ w_{i,3} \\ \vdots \\ w_{i,m} \end{bmatrix} \qquad 0 < \|W_i\|_2 < 1$$

$$\Rightarrow 0 < \sqrt{\sum_{j=1}^{m} w_{i,j}} < 1$$

$$\Rightarrow \forall w_{i,j} : |w_{i,j}| < 1$$

For each element j in $\nabla_{h_1} L$:

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_{k-1}} \prod_{i=2}^{k-1} \frac{\partial h_i}{\partial h_{i-1}}$$

$x = w_{k,j}^T h_{k-1,j} + b_k$
(not input vector)

$$= \frac{\partial L}{\partial \hat{y}} w_{k,j} \left(\prod_{i=2}^{k-1} \frac{\partial g(x)}{\partial x} \frac{\partial x}{\partial h_{i-1}} \right)$$

$$= \frac{\partial L}{\partial \hat{y}} \left(\frac{\partial g(x)}{\partial x}\right)^{k-2} \cancel{\left(\prod_{i=2}^{k} w_{i,j} \right)}$$

$$\begin{matrix} k \to \infty \\ |w_{i,j}| < 1 \end{matrix} \Rightarrow \prod_{i=2}^{k} w_{i,j} = 0$$

$$\Rightarrow \boxed{\nabla_{h_1} L_{k \to \infty} = 0}$$