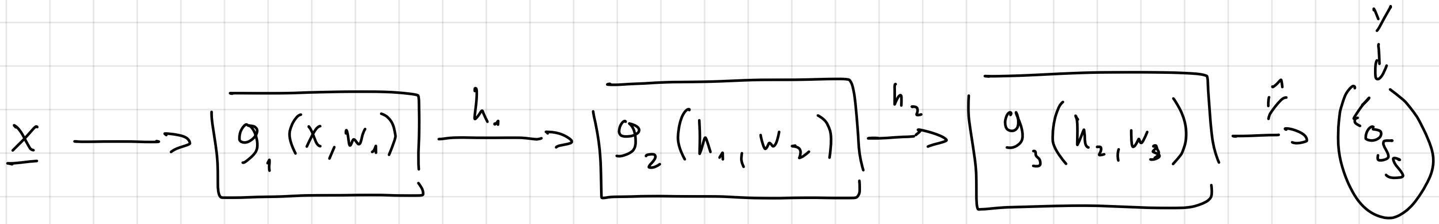
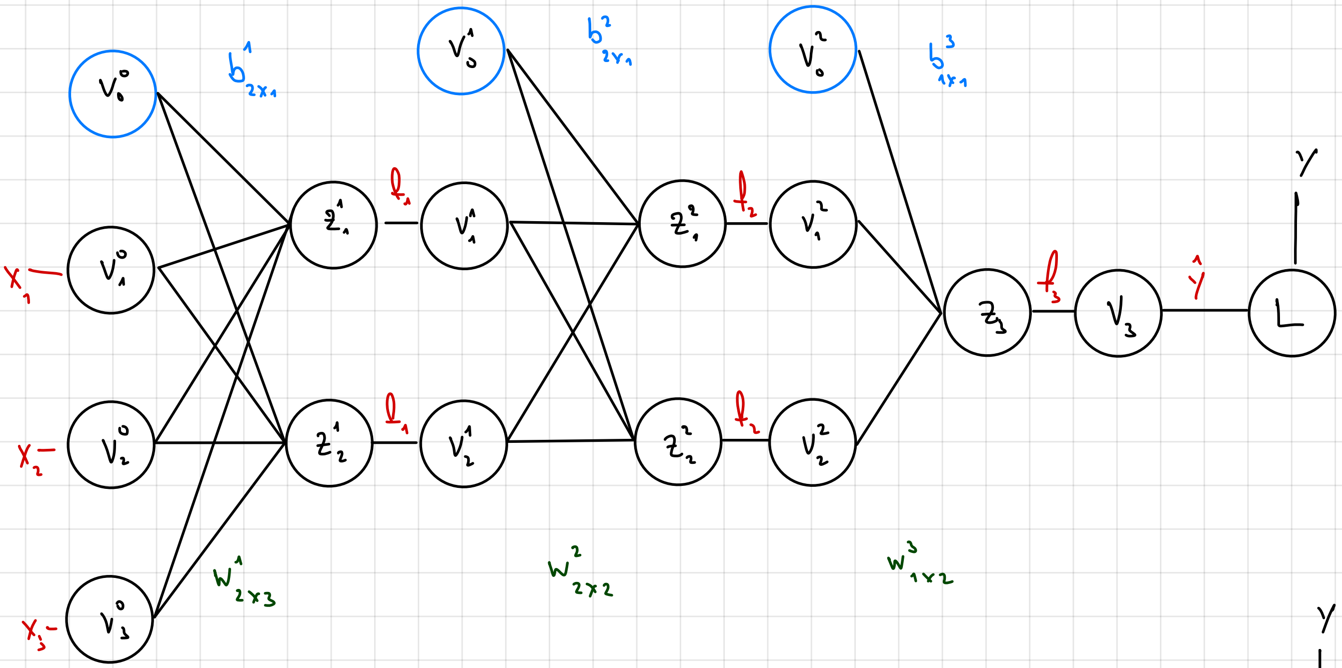
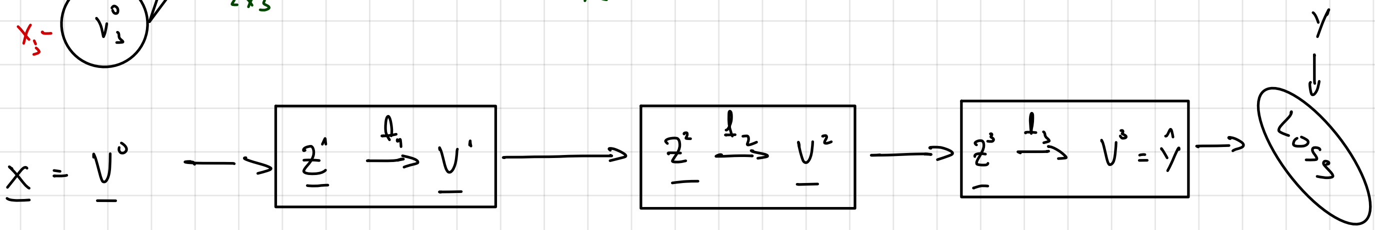


$$z^{\ell} = W^{\ell} z^{\ell-1} + b^{\ell}$$

$$v^{\ell} = f_{\ell}(z^{\ell}) \in \mathbb{R}^{d_\ell \times 1}$$





$$L(\hat{y}, y) = RSS(\hat{y}, y) = (y - \hat{y})^2 = (y - \hat{y})^T (y - \hat{y})$$

$$\hat{y} = v^3 = z_3 = W^3 \cdot v^2 + b^3 = W^3 \cdot f(W^2 v^1 + b^2) + b^3$$

$$= W^3 \cdot f\left(W^2 \cdot f[W^1 \cdot x + b^1] + b^2\right) + b^3 \quad ; \quad f = ReLU$$

$$\frac{\partial L}{\partial W^3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W^3} = 2(\hat{y} - y) \cdot \frac{\partial [W^3 \cdot v^2 + b^3]}{\partial W^3} = 2(\hat{y} - y) \cdot v^{2^T}$$

$$\frac{\partial L}{\partial b^3} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b^3} = 2(\hat{y} - y) \cdot 1.$$

$$\frac{\partial L}{\partial W^2_{2\times 2}} = \frac{\partial L}{\partial V^2} \cdot \frac{\partial V^2}{\partial W^2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V^2} \cdot \frac{\partial V^2}{\partial W^2} = 2(\hat{y}-y) \cdot W^3 \cdot \frac{\partial f(W^2 \cdot V^1 + b^2)}{\partial W^2}$$

$$= \left[f'(W^2 V^1 + b^1) \circ W^{3^T} \cdot 2(\hat{y}-y) \right] \cdot V^{1^T}$$

$$\frac{\partial L}{\partial b^2} = \frac{\partial L}{\partial V_2} \cdot \frac{\partial V_2}{\partial b_2} = f'(W^2 V^1 + b^1) \circ W^{3^T} \cdot 2(\hat{y}-y) \cdot 1$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial V_1} \cdot \frac{\partial V_1}{\partial W_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V^1} \cdot \frac{\partial V^1}{\partial W_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V_2} \cdot \frac{\partial V_2}{\partial V_1} \cdot \frac{\partial V_1}{\partial W_1}$$

$$= W^{2^T} \left[f'(W^2 V^1 + b^1) \circ W^{3^T} \cdot 2(\hat{y}-y) \right] \cdot \frac{\partial f(W^1 \cdot X + b^1)}{\partial W^1}$$

$$= \left[f'(W^1 \cdot X + b^1) \circ W^{2^T} \left[f'(W^2 V^1 + b^1) \circ W^{3^T} \cdot 2(\hat{y}-y) \right] \right] \cdot X^T$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial V_1} \cdot \frac{\partial V_1}{\partial b_1} = \left[f'(W^1 \cdot X + b^1) \circ W^{2^T} \left[f'(W^2 V^1 + b^1) \circ W^{3^T} \cdot 2(\hat{y}-y) \right] \right] \cdot 1$$

$$X = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \quad b^i = \underline{1}^T, \quad W^i = \underline{1}^{n \times m}, \quad y = 0$$

$$z^1 = W^1 \cdot V^0 + b^1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$V^1 = f(z^1) = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$z^2 = W^2 \cdot V^1 + b^2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \end{pmatrix}$$

$$V^2 = f(z^2) = \begin{pmatrix} 7 \\ 7 \end{pmatrix}$$

$$z^3 = W^3 \cdot V^2 + b^3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{pmatrix} 7 \\ 7 \end{pmatrix} + 1 = 15$$

$$V^3 = z^3 \Rightarrow \underline{\underline{\hat{y} = 15}}$$

$$\frac{\partial L}{\partial W^3} = 2(\hat{y}-y)\cdot V^{2^T} = \Big|_{(x,y,w,b)} 2\cdot(15-0)\cdot\begin{pmatrix}7\\7\end{pmatrix} = 30\cdot\begin{pmatrix}7\\7\end{pmatrix}$$

$$\frac{\partial L}{\partial b^3} = 2(\hat{y}-y)\cdot 1 = 30$$

$$\underset{2\times 2}{\frac{\partial L}{\partial W^2}} = \left[f'(W^2V^1+b^1)\circ W^{3^T}\cdot 2(\hat{y}-y)\right]\cdot V^{1^T} = \Big|_{(x,y,w,b)} \left[f'\left(\begin{pmatrix}7\\7\end{pmatrix}\right)\circ\begin{pmatrix}1\\1\end{pmatrix}\cdot 30\right]\cdot[3,3]$$

$$= \begin{pmatrix}30\\30\end{pmatrix}\cdot[3,3] = 3\cdot\begin{bmatrix}30 & 30\\30 & 30\end{bmatrix}$$

$$\frac{\partial L}{\partial b^2} = f'(W^2V^1+b^1)\circ W^{3^T}\cdot 2(\hat{y}-y)\cdot 1 = \begin{pmatrix}30\\30\end{pmatrix}$$

$$\frac{\partial L}{\partial W_1} = \left[f'(W^1\cdot x+b^1)\circ W^{2^T}\left[f'(W^2V^1+b^1)\circ W^{3^T}\cdot 2(\hat{y}-y)\right]\right]\cdot x^T$$

$$= \Big|_{(x,y,w,b)} \left[f'\left(\begin{pmatrix}3\\3\end{pmatrix}\right)\circ\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\cdot\begin{pmatrix}30\\30\end{pmatrix}\right]\begin{pmatrix}1 & 2 & -1\end{pmatrix}$$

$$= \left[\begin{pmatrix}1\\1\end{pmatrix}\circ\begin{pmatrix}60\\60\end{pmatrix}\right]\cdot\begin{pmatrix}1 & 2 & -1\end{pmatrix} = \begin{pmatrix}60\\60\end{pmatrix}\cdot(1,2,-1) = 60\cdot\begin{bmatrix}1 & 2 & -1\\1 & 2 & -1\end{bmatrix}$$

$$\frac{\partial L}{\partial b_1} = \left[f'(W^1\cdot x+b^1)\circ W^{2^T}\left[f'(W^2V^1+b^1)\circ W^{3^T}\cdot 2(\hat{y}-y)\right]\right]\cdot 1 = \Big|_{(x,y,w,b)}\begin{pmatrix}60\\60\end{pmatrix}$$