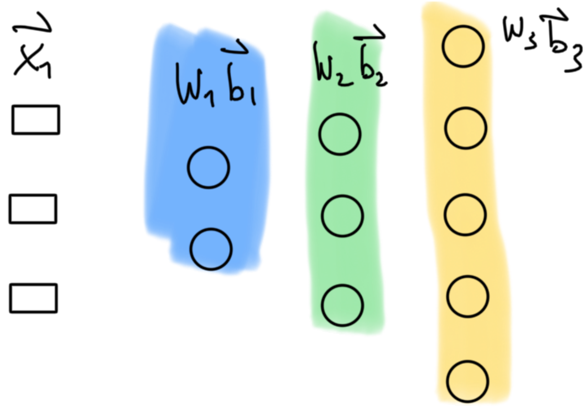
Neural network back propagation 2

New approach



The neural net f_{NN} returns a vector of length 5

$$f_{NN}(\vec{x}) = \vec{y} = g_3(\vec{z}_3)$$

Feed forward

$$f_3(f_2(f_1(\vec{x}))) = f_{NN}(\vec{x})$$

activation

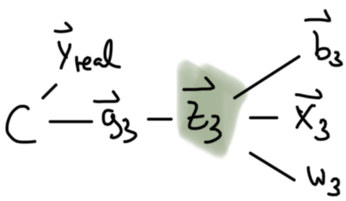
$$f_\ell(\vec{x}_i) = g_\ell(\underbrace{W_\ell \vec{x}_i + b_\ell}_{\vec{z}_i}) = \vec{x}_{i+1} \qquad g_1 = g_2 = g_3 = \sigma$$

Backpropagation

Layer 3

$$C = \sum_{i=0} (g_{3_i} - y_{i\,real})^2$$

$$W_3 = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \\ W_{41} & W_{42} & W_{43} \\ W_{51} & W_{52} & W_{53} \end{bmatrix} \quad \vec{b}_3 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} \quad \vec{X}_3 = [\;]$$



$$\frac{dC}{d\vec{b}_m} = \frac{dC}{d\vec{g}_3}\frac{d\vec{g}_3}{d\vec{z}_3}\frac{d\vec{z}_3}{d\vec{b}_3} = 2(\vec{g}_{3_m} - y_{m\,real}) \cdot \sigma'(\vec{z}_{3_m}) \cdot 1$$

$$\frac{dC}{dW_{mn}} = 2(\vec{g}_{3_m} - y_{m\,real}) \cdot \sigma'(\vec{z}_{3_m}) \cdot \vec{X}_{3_n}$$

5x1

[P1]

$$\frac{\partial C}{\partial \vec{x}_3} = \frac{\partial C}{\partial \vec{g}_3}\,\frac{\partial \vec{g}_3}{\partial \vec{z}_3}\,\frac{\partial \vec{z}_3}{\partial \vec{x}_3} = \overbrace{W_3^{t}}^{3\times 5} \cdot \overbrace{2(\vec{g}_3 - \vec{y}_{real}) \odot \sigma'(\vec{z}_3)}^{\begin{bmatrix} p_1 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} \;\; 3\times 1}$$

3×1

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\vec{z}_3 = W_3 \cdot \vec{X}_3 + \vec{b}_3 = \begin{bmatrix} w_{11} \cdot x_1 + w_{12} \cdot x_2 + w_{13} \cdot x_3 + b_3 \\ w_{21} \cdot x_1 + w_{22} \cdot x_2 + w_{23} \cdot x_3 + b_3 \\ w_{31} \cdot x_1 + w_{32} \cdot x_2 + w_{33} \cdot x_3 + b_3 \\ w_{41} \cdot x_1 + w_{42} \cdot x_2 + w_{43} \cdot x_3 + b_3 \\ w_{51} \cdot x_1 + w_{52} \cdot x_2 + w_{53} \cdot x_3 + b_3 \end{bmatrix}$$

$$\frac{\partial \vec{z}_3}{\partial \vec{x}_3} = \begin{bmatrix} \frac{\partial z_3}{\partial x_1} \\ \frac{\partial z_3}{\partial x_2} \\ \frac{\partial z_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & w_{51} \\ w_{12} & w_{22} & w_{32} & w_{42} & w_{52} \\ w_{13} & w_{23} & w_{33} & w_{43} & w_{53} \end{bmatrix} = W^{t}$$

Layer 2

$$\begin{tikzcd} & & & \vec{b}_2 \\ C \arrow[r, dash] & \vec{g}_2 \arrow[r, dash] & \vec{z}_2 \arrow[ur, dash] \arrow[r, dash] \arrow[dr, dash] & \vec{x}_2 \\ & & & w_2 \end{tikzcd}$$

$$\underbrace{\frac{\partial C}{\partial \vec{x}_2}}_{2\times 1} = \frac{\partial C}{\partial \vec{g}_2}\,\frac{\partial \vec{g}_2}{\partial \vec{z}_2}\,\frac{\partial \vec{z}_2}{\partial \vec{x}_2} = \overbrace{W_2^{t}}^{2\times 3}\;\overbrace{\frac{\partial C}{\partial \vec{x}_3}}^{3\times 1} \odot \sigma'(\vec{z}_2)$$

Layer 1

3×1 $\quad$ 3×2 $\quad$ 2×1

$$\frac{\partial C}{\partial \vec{x}_1} = \frac{\partial C}{\partial \vec{s}_1}\frac{\partial \vec{s}_1}{\partial \vec{z}_1}\frac{\partial \vec{z}_1}{\partial \vec{x}_1} = W_1^t \quad \frac{\partial C}{\partial \vec{x}_2} \odot \sigma'(\vec{z}_1)$$

Layer L ← Last Layer