

## Vzorečky

- aktivační funkce sigmoid:  $\sigma(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(\sum_i^n w_i x_i + b)}}$ ,  $\sigma(\vec{z}) = \begin{pmatrix} \sigma(z_1) \\ \vdots \\ \sigma(z_n) \end{pmatrix}$
- aktivace k-té vrstvy:  $a^{(k)} = \sigma\left(w^{(k)} \cdot a^{(k-1)} + b^{(k)}\right)$
- pomocná proměnná z:  $z^{(k)} = w^{(k)} \cdot a^{(k-1)} + b^{(k)} \implies a^{(k)} = \sigma\left(z^{(k)}\right)$
- derivace aktivační funkce:  $\sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$
- účelová funkce MSE:  $C(w, b) \equiv \frac{1}{2n} \sum_x \|y(x) - a\|^2$
- derivace účelové funkce:  $\frac{\partial C_{i_j}}{\partial a^{(k)}} = 2(a^{(k)} - y)$

## Vzorečky

- **SGD:**  $w_k \rightarrow \tilde{w}_k = w_k - \frac{\xi}{m} \nabla C_{i_j}^{(k)} = w_k - \frac{\xi}{m} \sum_{j=1}^m \frac{\partial C_{i_j}}{\partial w_k}$

$$b_l \rightarrow \tilde{b}_l = w_k - \frac{\xi}{m} \nabla C_{i_j}^{(l)} = b_l - \frac{\xi}{m} \sum_{j=1}^m \frac{\partial C_{i_j}}{\partial b_l},$$

- **Backpropagation algoritmus:**

1. Input x: nastavíme aktivaci input vrstvy  $a^1$
2. Feedforward: pro  $l = 2, 3, \dots, L$  spočítáme  $z^l = w^l a^{l-1} + b^l$  a  $a^l = \sigma(z^l)$
3. Output error: spočítáme vektor  $\delta^L = \nabla_a C \odot \sigma'(z^L)$
4. Backpropagation: pro  $l = L - 1, L - 2, \dots, 2$  spočítáme  $\delta^l = \left( (w^{l+1})^T \delta^{l+1} \right) \odot \sigma'(z^l)$
5. Output: gradient účelové funkce je dán vztahy  $\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta^l$  a  $\frac{\partial C}{\partial b_j^l} = \delta_j^l$

# Kuchařka

## ● Backpropagation s SGD:

0. Input mini batch. Postupně беру všechny body  $x$ :

1. Input  $x$ : nastavíme aktivaci input vrstvy  $a^{x,1}$

2. Feedforward: pro  $l = 2, 3, \dots, L$  spočítáme  $z^{x,l} = w^l a^{x,l-1} + b^l$  a  $a^l = \sigma(z^{x,l})$

3. Output error: spočítáme vektor  $\delta^{x,L} = \nabla_a C \odot \sigma'(z^{x,L})$

4. Backpropagation: pro  $l = L - 1, L - 2, \dots, 2$  spočítáme  $\delta^{x,l} = \left( (w^{l+1})^T \delta^{x,l+1} \right) \odot \sigma'(z^{x,l})$

5. SGD: pro každé  $l=L, L-1, \dots, 2$  aktualizujeme váhy pomocí pravidla

$$w^l \rightarrow w^l - \frac{\xi}{m} \sum_x \delta^{x,l} \cdot (a^{x,l-1})^T \text{ a biasy podle pravidla}$$

$$b^l \rightarrow b^l - \frac{\xi}{m} \sum_x \delta^{x,l}$$