

## Lekce 1

Considering a neural network with  $D$  input neurons, a single hidden layer with  $H$  neurons,  $K$  output neurons, hidden activation  $f$  and output activation  $a$ , list its parameters (including their size) and write down how is the output computed. [5]

Parametry

- Matice  $W \in \mathbb{R}^{D \times H}$  a  $V \in \mathbb{R}^{H \times K}$
- Biasy  $b$  a  $p$

Výpočet

- Výstup vnitřní vrstvy  $h = f(Wx + b)$
- Finální výstup  $o = a(Vh + p)$

List the definitions of frequently used MLP output layer activations (the ones producing parameters of a Bernoulli distribution and a categorical distribution). Then write down three commonly used hidden layer activations (sigmoid, tanh, ReLU). [5]

Výstupní vrstvy

- $\sigma(x) = 1/(1 + e^{-x})$  pro binární klasifikaci
- $\text{softmax}(x)_i = e^{x_i} / \sum_j e^{x_j}$ , rozšíření sigmoidu na více tříd

Vnitřní vrstvy

- $\sigma(x) = 1/(1 + e^{-x})$ , není ideální
- $\tanh(x) = 2\sigma(2x) - 1$ , sigmoid upravený tak, aby byl symetrický (tj. aby jeho opakování nekonvergovalo k 1) a aby jeho derivace v 0 byla 1
- $\text{ReLU}(x) = \max(0, x)$ , jednoduchá nelinearita

Formulate the Universal approximation theorem. [5]

Nechť  $\varphi(x)$  je nekonalatní, omezená, neklesající spojitá funkce (později dokonce jakákoli nepolynomiální). Poté  $\forall \varepsilon > 0$  a  $\forall f$  spojitě na  $[0, 1]^D$  existuje  $N \in \mathbb{N}$ ,  $v \in \mathbb{R}^N$ ,  $b \in \mathbb{R}^N$ ,  $W \in \mathbb{R}^{N \times D}$  takové, že pokud máme  $F(x)$  jako

$$F(x) = v^T \varphi(Wx + b),$$

tak pro všechny  $x \in [0, 1]^D$  platí

$$|F(x) - f(x)| < \varepsilon.$$

Jinými slovy, pokud máme vhodnou aktivační funkci, umíme pomocí ní a pomocí vhodné lineární transformace  $W$  libovolně dobře aproximovat jakoukoli spojitou funkci.