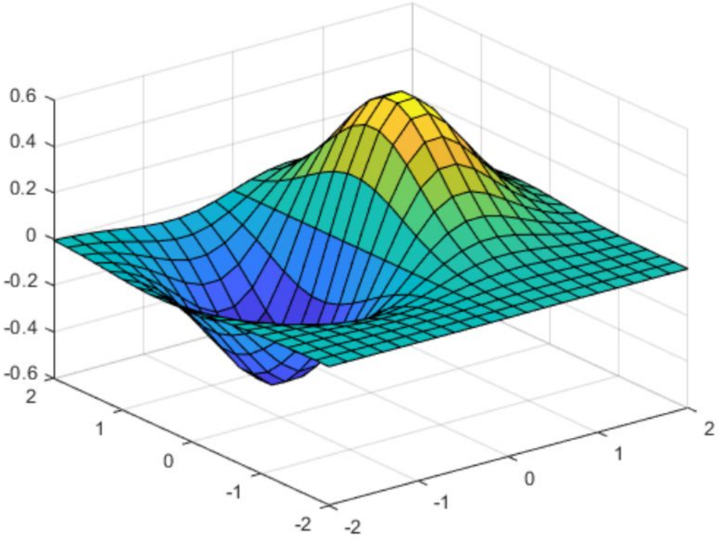
Jak zefektivnit učení NN

Alternativní model neuronu



Alternativní model neuronu

- Síť sigmoid neuronů teoreticky může aproximovat libovolnou funkci, ale v praxi mají často navrch jiné modely neuronu - rychlejší učení i lepší schopnost generalizovat

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neruonem (***l+1***)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neuronem (***l+1***)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}
  - gradient získáme z rovnic backpropagation: $a_k^l \delta_j^{l+1}$

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neuronem (***l+1***)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}
  - gradient získáme z rovnic backpropagation: $a_k^l \delta_j^{l+1}$
  - aktivace jsou pozitivní, takže gradient bude mít stejné znaménko, jako chyba neuronu δ_j^{l+1}

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neuronem (***l+1***)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}
  - gradient získáme z rovnic backpropagation: $a_k^l \delta_j^{l+1}$
  - aktivace jsou pozitivní, takže gradient bude mít stejné znaménko, jako chyba neuronu δ_j^{l+1}
  - pokud $\delta_j^{l+1} > 0$, všechny váhy w_{jk}^{l+1} se v GD zmenší

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neuronem (***l+1***)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}
  - gradient získáme z rovnic backpropagation: $a_k^l \delta_j^{l+1}$
  - aktivace jsou pozitivní, takže gradient bude mít stejné znaménko, jako chyba neuronu δ_j^{l+1}
  - pokud $\delta_j^{l+1} > 0$, všechny váhy w_{jk}^{l+1} se v GD zmenší
  - pokud $\delta_j^{l+1} < 0$, všechny váhy w_{jk}^{l+1} se v GD zvětší

Potenciální nedostatek sigmoid neuronů

- Potenciální nedostatek sigmoid neuronů:
  - uvažujme váhy mezi ***j***-tým neruonem (***l*+1**)-té vrstvy a předchozí vrstvou w_{jk}^{l+1}
  - gradient získáme z rovnic backpropagation: $a_k^l \delta_j^{l+1}$
  - aktivace jsou pozitivní, takže gradient bude mít stejné znaménko, jako chyba neuronu δ_j^{l+1}
  - pokud $\delta_j^{l+1} > 0$, všechny váhy w_{jk}^{l+1} se v GD zmenší
  - pokud $\delta_j^{l+1} < 0$, všechny váhy w_{jk}^{l+1} se v GD zvětší
  - jinými slovy se **všechny tyto váhy** zmenšují nebo zvětšují dohromady

Potenciální nedostatek sigmoid neuronů

- To může být problém - co když by bylo lepší některé váhy zmenšovat a jiné zvětšovat?

Potenciální nedostatek sigmoid neuronů

- To může být problém - co když by bylo lepší některé váhy zmenšovat a jiné zvětšovat?
- To může nastat pouze tehdy, když budou mít aktivace různá znaménka => na to je třeba jiná aktivační funkce

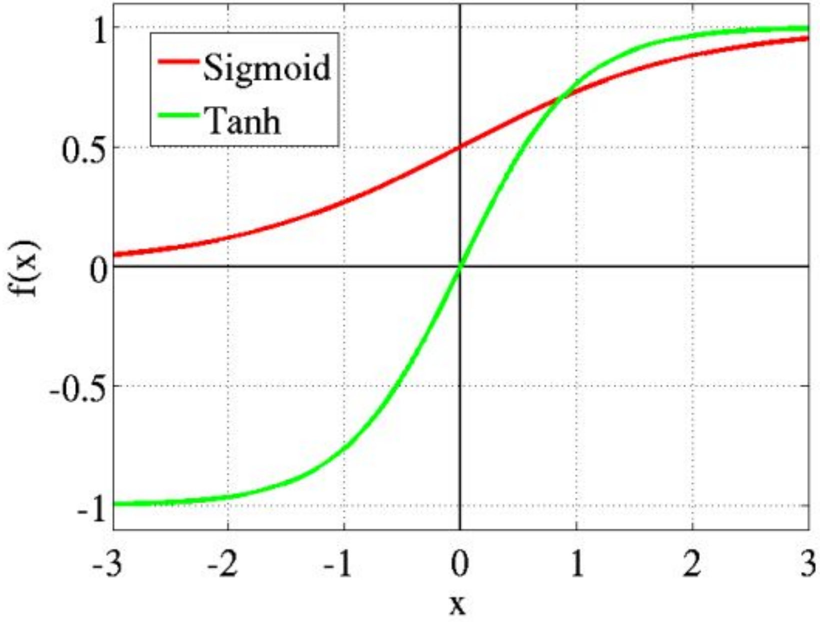
Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

$$\tanh(w \cdot x + b) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

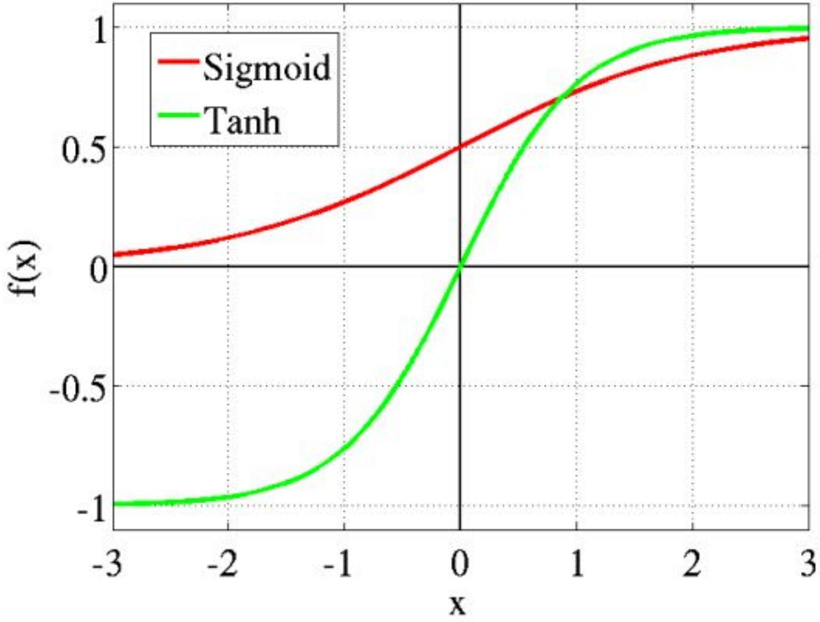


Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

$$\tanh(w \cdot x + b) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

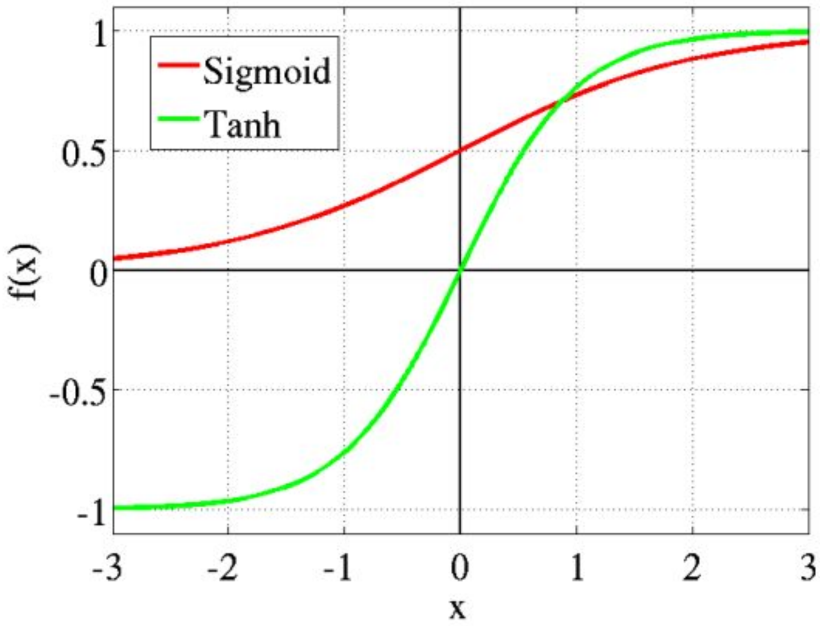
- Hf=(-1,1)



Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

$$\tanh(w \cdot x + b) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

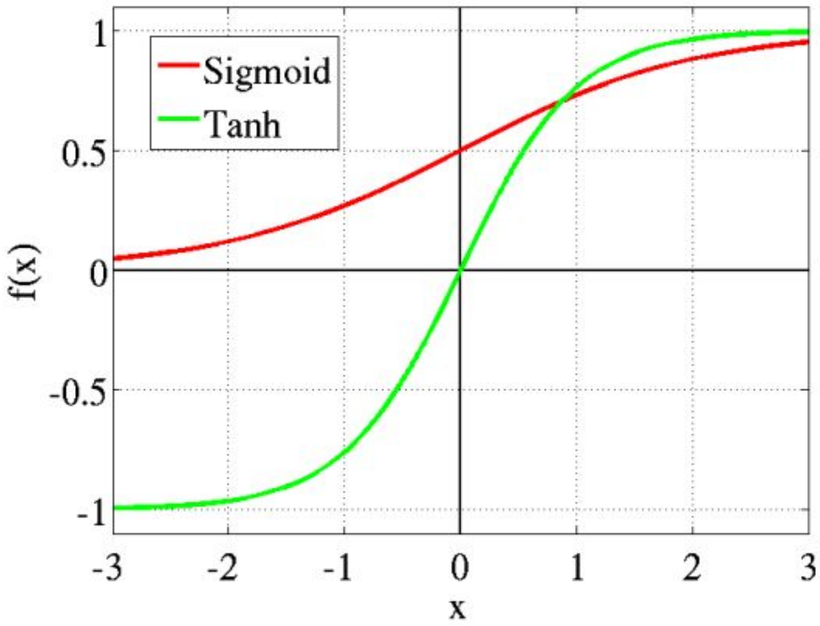


- Hf=(-1,1)
- “Přeškálovaný” sigmoid

Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

$$\tanh(w \cdot x + b) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

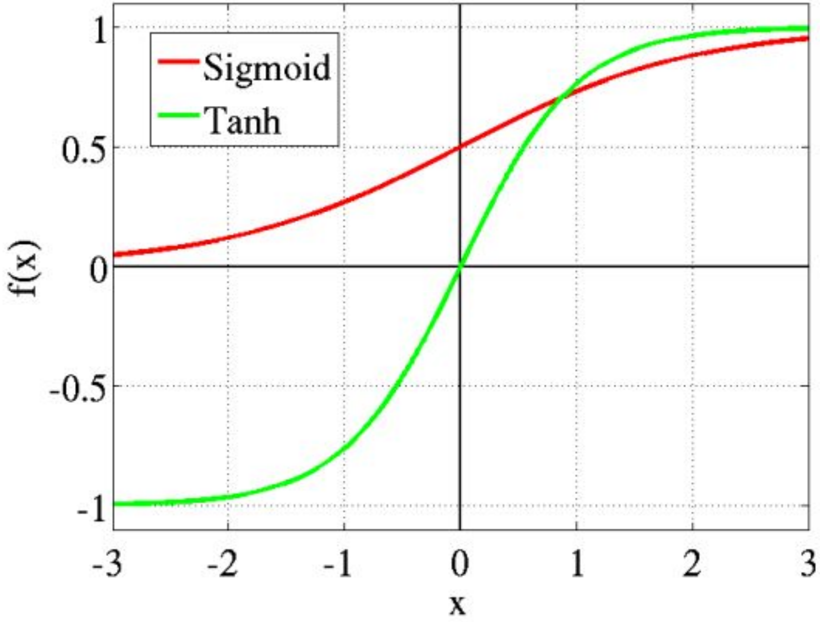


- Hf=(-1,1)
- “Přeškálovaný” sigmoid
- Symetrický okolo 0 => +-polovina aktivací by mohla být záporná, respektive kladná

Aktivační funkce

- Příkladem takové funkce může být hyperbolický tangens

$$\tanh(w \cdot x + b) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



- Hf=(-1,1)
- “Přeškálovaný” sigmoid
- Symetrický okolo 0 => +-polovina aktivací by mohla být záporná, respektive kladná
- Obvykle je třeba normalizovat output sítě (a často i input)

Aktivační funkce: cvičení 1

- Dokažte vztah mezi tanh a sigmoidem: $\sigma(z) = \dfrac{1 + \tanh\left(\frac{z}{2}\right)}{2}$

Aktivační funkce: cvičení 1

- Dokažte vztah mezi tanh a sigmoidem $\sigma(z) = \frac{1 + \tanh\left(\frac{z}{2}\right)}{2}$
  - $$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^z - \frac{1}{e^z}}{e^z + \frac{1}{e^z}} = \frac{\frac{e^{2z}-1}{e^z}}{\frac{e^{2z}+1}{e^z}} = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Aktivační funkce: cvičení 1

- Dokažte vztah mezi tanh a sigmoidem $\sigma(z) = \dfrac{1 + \tanh\left(\frac{z}{2}\right)}{2}$
  - $$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^z - \frac{1}{e^z}}{e^z + \frac{1}{e^z}} = \frac{\frac{e^{2z}-1}{e^z}}{\frac{e^{2z}+1}{e^z}} = \frac{e^{2z} - 1}{e^{2z} + 1}$$
  - $$\sigma(z) = \frac{1 + \tanh\left(\frac{z}{2}\right)}{2} = \frac{1 + \frac{e^z-1}{e^z+1}}{2} = \frac{1}{2}\,\frac{e^z + 1 + e^z - 1}{e^z + 1} = \frac{1}{2}\,\frac{2e^z}{e^z + 1} = \frac{e^z}{e^z + 1} =$$
    $$= \frac{1}{\frac{e^z+1}{e^z}} = \frac{1}{1 + \frac{1}{e^z}} = \frac{1}{1 + e^{-z}} = \sigma(z)$$

Aktivační funkce

- Podobně jako sigmoid může mít tanh problém, pokud je neuron saturován

Aktivační funkce

- Podobně jako sigmoid může mít tanh problém, pokud je neuron saturován
- Můžeme použít nekonečně mnoho aktivačních funkcí

Aktivační funkce

- Podobně jako sigmoid může mít tanh problém, pokud je neuron saturován
- Můžeme použít nekonečně mnoho aktivačních funkcí
- Nevíme, které aktivační funkce/kombinace aktivačních funkcí bude pro daný problém nejlepší

Aktivační funkce

- ReLU = rectified linear unit: relu(z)=max(0,y)
- Leaky ReLU
- ELU
- SELU
- SWISH
- GELU
- MISCH
- softmax
- softplus