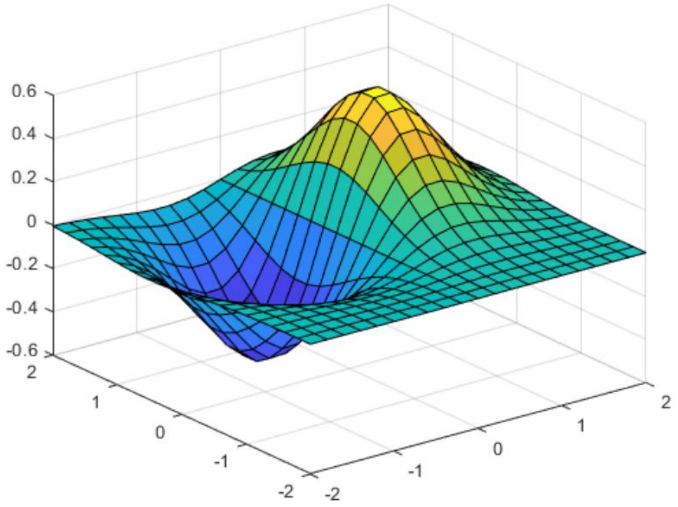
Jak se neuronové sítě učí

Backpropagation



Účelová funkce

- Funkce všech vah a biasů

Účelová funkce

- Funkce všech vah a biasů
- Měří, jak “dobře” se síť chová

Účelová funkce

- Funkce všech vah a biasů
- Měří, jak “dobře” se síť chová
- Rozdíl mezi výstupem ze sítě a labelů

Účelová funkce

- Funkce všech vah a biasů
- Měří, jak “dobře” se síť chová
- Rozdíl mezi výstupem ze sítě a labelů
- např.

$$C(w,b) \equiv \frac{1}{2n} \sum_x \| y(x) - a \|^2$$

Gradient descent

- Váhy a biasy upravujeme v každém kroku gradient descentu na základě tréninkových dat v batchi. Pohybujeme se ve “směru” záporného gradientu účelové funkce

Gradient descent

- Váhy a biasy upravujeme v každém kroku gradient descentu na základě tréninkových dat v batchi. Pohybujeme se ve “směru” záporného gradientu účelové funkce
- Tím docílíme nejvyššího poklesu hodnoty účelové funkce

Gradient descent

- Váhy a biasy upravujeme v každém kroku gradient descentu na základě tréninkových dat v batchi. Pohybujeme se ve “směru” záporného gradientu účelové funkce
- Tím docílíme nejvyššího poklesu hodnoty účelové funkce
- Poznámka: čím větší je hodnota prvku gradientu, tím citlivější je účelová funkce na změnu příslušné proměnné

Jak gradient spočítat?

Backpropagation

- Backward propagation of errors = zpětná propagace chyb

Backpropagation

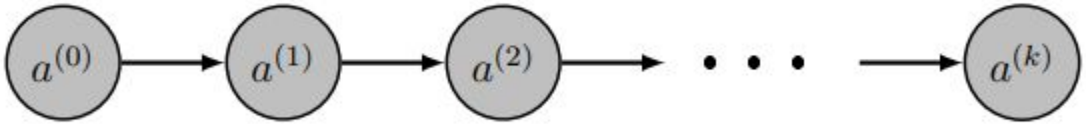
- Backward propagation of errors = zpětná propagace chyb
- Chceme určit, jak upravit váhy a biasy aby byla síť trochu lepší pro jeden konkrétní tréninkový data point, respektive batch

Backpropagation

- Backward propagation of errors = zpětná propagace chyb
- Chceme určit, jak upravit váhy a biasy aby byla síť trochu lepší pro jeden konkrétní tréninkový data point, respektive batch
- video od 3blue1brown

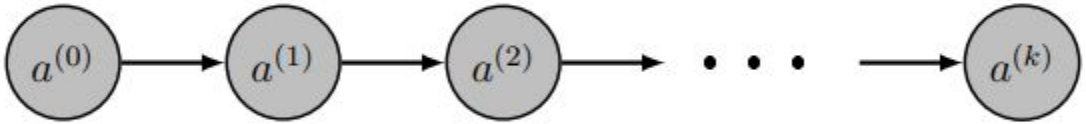
Backpropagation

- Vezmeme si jednodušší síť, odvodíme algoritmus a poté ho zobecníme



Backpropagation

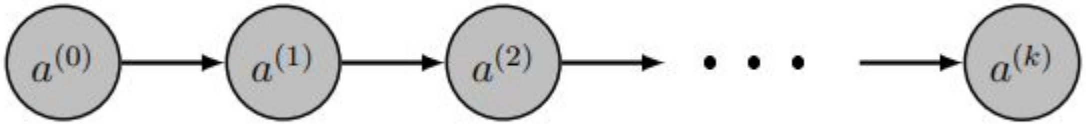
- Vezmeme si jednodušší síť, odvodíme algoritmus a poté ho zobecníme



- *k+1* vrstev, 1 neuron v každé vrstvě

Backpropagation

- Vezmeme si jednodušší síť, odvodíme algoritmus a poté ho zobecníme



- *k+1* vrstev, 1 neuron v každé vrstvě
- prozkoumáme dopad, který bude mít malá změna ve váze $w^{(k)}$, biasu $b^{(k)}$ nebo aktivaci předchozí vrstvy $a^{(k-1)}$ na účelovou funkci.

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W^s} a^{s-1} + b^s).$

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W}^\mathbf{s} a^{s-1} + b^s).$
- pro naší síť je výstup dán jako $a^{(k)} = \sigma(w^{(k)} a^{(k-1)} + b^{(k)})$

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W}^\mathbf{s} a^{s-1} + b^s).$
- pro naší síť je výstup dán jako $a^{(k)} = \sigma(w^{(k)} a^{(k-1)} + b^{(k)})$
- definujme si $z^{(k)} := w^{(k)} a^{(k-1)} + b^{(k)}$

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W}^{\mathbf{s}} a^{s-1} + b^s).$
- pro naší síť je výstup dán jako $a^{(k)} = \sigma(w^{(k)} a^{(k-1)} + b^{(k)})$
- definujme si $z^{(k)} := w^{(k)} a^{(k-1)} + b^{(k)}$
- nyní můžeme psát $a^{(k)} = \sigma(z^{(k)})$

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W}^{\mathbf{s}} a^{s-1} + b^s).$
- pro naší síť je výstup dán jako $a^{(k)} = \sigma(w^{(k)} a^{(k-1)} + b^{(k)})$
- definujme si $z^{(k)} := w^{(k)} a^{(k-1)} + b^{(k)}$
- nyní můžeme psát $a^{(k)} = \sigma(z^{(k)})$
- označme si (cost) jednoho tréninkového příkladu, přičemž ***y*** je správný label, jako

$$C_{i_j} = \|a^{(k)} - y\|^2 = (a^{(k)} - y)^2$$

Backpropagation

- aktivace (vrstvy) neuronů je dána: $a^s = \sigma(\mathbf{W}^{\mathbf{s}} a^{s-1} + b^s).$
- pro naší síť je výstup dán jako $a^{(k)} = \sigma(w^{(k)} a^{(k-1)} + b^{(k)})$
- definujme si $z^{(k)} := w^{(k)} a^{(k-1)} + b^{(k)}$
- nyní můžeme psát $a^{(k)} = \sigma(z^{(k)})$
- označme si (cost) jednoho tréninkového příkladu, přičemž ***y*** je správný label, jako

  $$C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- dohromady $C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

$$C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

$$C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- chceme najít $\frac{\partial C_{i_j}}{\partial w^{(k)}}, \frac{\partial C_{i_j}}{\partial b^{(k)}}$, ale taky $\frac{\partial C_{i_j}}{\partial a^{(k-1)}}$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y) \qquad C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- chceme najít $\frac{\partial C_{i_j}}{\partial w^{(k)}}, \frac{\partial C_{i_j}}{\partial b^{(k)}}$, ale taky $\frac{\partial C_{i_j}}{\partial a^{(k-1)}}$
- jinými slovy jak citlivá je účelová funkce na malou změnu váhy, biasu a předchozí aktivace

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y) \qquad C_{i_j} = \|a^{(k)} - y\|^2 = (a^{(k)} - y)^2$$

- chceme najít $\frac{\partial C_{i_j}}{\partial w^{(k)}}, \frac{\partial C_{i_j}}{\partial b^{(k)}}$, ale taky $\frac{\partial C_{i_j}}{\partial a^{(k-1)}}$
- jinými slovy jak citlivá je účelová funkce na malou změnu váhy, biasu a předchozí aktivace
- použijeme řetězové pravidlo (derivace složené funkce)

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

$$C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- derivujeme složenou funkci $\dfrac{\partial C_{ij}}{\partial w^{(k)}} = \dfrac{\partial C_{ij}}{\partial a^{(k)}} \cdot \dfrac{\partial a^{(k)}}{\partial z^{(k)}} \cdot \dfrac{\partial z^{(k)}}{\partial w^{(k)}}$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

$$C_{i_j} = \left\| a^{(k)} - y \right\|^2 = (a^{(k)} - y)^2$$

- derivujeme složenou funkci $\dfrac{\partial C_{ij}}{\partial w^{(k)}} = \dfrac{\partial C_{ij}}{\partial a^{(k)}} \cdot \dfrac{\partial a^{(k)}}{\partial z^{(k)}} \cdot \dfrac{\partial z^{(k)}}{\partial w^{(k)}}$
- přepíšeme jako $\dfrac{\partial C_{i_j}}{\partial w^{(k)}} = \dfrac{\partial z^{(k)}}{\partial w^{(k)}} \dfrac{\partial a^{(k)}}{\partial z^{(k)}} \dfrac{\partial C_{i_j}}{\partial a^{(k)}}$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y) \qquad C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- derivujeme složenou funkci $\dfrac{\partial C_{ij}}{\partial w^{(k)}} = \dfrac{\partial C_{ij}}{\partial a^{(k)}} \cdot \dfrac{\partial a^{(k)}}{\partial z^{(k)}} \cdot \dfrac{\partial z^{(k)}}{\partial w^{(k)}}$
- přepíšeme jako $\dfrac{\partial C_{i_j}}{\partial w^{(k)}} = \dfrac{\partial z^{(k)}}{\partial w^{(k)}} \dfrac{\partial a^{(k)}}{\partial z^{(k)}} \dfrac{\partial C_{i_j}}{\partial a^{(k)}}$
- nyní spočítáme tyto derivace

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$

$$z^{(k)} := w^{(k)} a^{(k-1)} + b^{(k)}$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$

- $$\frac{\partial a^{(k)}}{\partial z^{(k)}} = \sigma'(z^{(k)}) = \frac{\exp\left(-z^{(k)}\right)}{\left(\exp\left(-z^{(k)}\right) + 1\right)^2}$$

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1+e^{-z}-e^{-z}}{1+e^{-z}} = 1 - \frac{e^{-z}}{1+e^{-z}}$$

$$a^{(k)} = \sigma(z^{(k)})$$

Backpropagation

$$
\begin{aligned}
\frac{d}{dx}\sigma(x) &= \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] \\
&= \frac{(0)(1+e^{-x}) - (-e^{-x})(1)}{(1+e^{-x})^2} \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}}\,\frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\,\frac{e^{-x}+(1-1)}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\,\frac{(1+e^{-x})-1}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\left[\frac{(1+e^{-x})}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right] \\
&= \frac{1}{1+e^{-x}}\left[1 - \frac{1}{1+e^{-x}}\right] \\
&= \sigma(x)(1-\sigma(x))
\end{aligned}
$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$
- $$\frac{\partial a^{(k)}}{\partial z^{(k)}} = \sigma'(z^{(k)}) = \frac{\exp(-z^{(k)})}{(\exp(-z^{(k)}) + 1)^2} = \sigma(z^{(k)})(1 - \sigma(z^{(k)}))$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$
- $$\frac{\partial a^{(k)}}{\partial z^{(k)}} = \sigma'(z^{(k)}) = \frac{\exp(-z^{(k)})}{(\exp(-z^{(k)}) + 1)^2} = \sigma(z^{(k)})(1 - \sigma(z^{(k)}))$$
- $$\frac{\partial C_{i_j}}{\partial a^{(k)}} = 2(a^{(k)} - y)$$

$$C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$
- $$\frac{\partial a^{(k)}}{\partial z^{(k)}} = \sigma'(z^{(k)}) = \frac{\exp\left(-z^{(k)}\right)}{\left(\exp\left(-z^{(k)}\right) + 1\right)^2} = \sigma(z^{(k)})(1 - \sigma(z^{(k)}))$$
- $$\frac{\partial C_{i_j}}{\partial a^{(k)}} = 2(a^{(k)} - y)$$
- dosadíme do $$\frac{\partial C_{i_j}}{\partial w^{(k)}} = \frac{\partial z^{(k)}}{\partial w^{(k)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}}$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y) \qquad C_{i_j} = \left\|a^{(k)} - y\right\|^2 = (a^{(k)} - y)^2$$

- $$\frac{\partial z^{(k)}}{\partial w^{(k)}} = a^{(k-1)}$$
- $$\frac{\partial a^{(k)}}{\partial z^{(k)}} = \sigma'(z^{(k)}) = \frac{\exp(-z^{(k)})}{(\exp(-z^{(k)}) + 1)^2} = \sigma(z^{(k)})(1 - \sigma(z^{(k)}))$$
- $$\frac{\partial C_{i_j}}{\partial a^{(k)}} = 2(a^{(k)} - y)$$

$$\frac{\partial C_{i_j}}{\partial w^{(k)}} = a^{(k-1)}\sigma'(z^{(k)})2(a^{(k)} - y) = 2a^{(k-1)}\sigma(z^{(k)})(1 - \sigma(z^{(k)}))(a^{(k)} - y)$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- Analogicky pro bias a předchozí aktivaci

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- Analogicky pro bias a předchozí aktivaci

$$\frac{\partial C_{i_j}}{\partial b^{(k)}} = \frac{\partial z^{(k)}}{\partial b^{(k)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}} = \sigma'(z^{(k)}) 2(a^{(k)} - y)$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- Analogicky pro bias a předchozí aktivaci

$$\frac{\partial C_{i_j}}{\partial b^{(k)}} = \frac{\partial z^{(k)}}{\partial b^{(k)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}} = \sigma'(z^{(k)}) 2(a^{(k)} - y)$$

$$\frac{\partial C_{i_j}}{\partial a^{(k-1)}} = \frac{\partial z^{(k)}}{\partial a^{(k-1)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}} = w^{(k)} \sigma'(z^{(k)}) 2(a^{(k)} - y)$$

Backpropagation

$$C_{ij} = C_{ij}(a^{(k)}(z^{(k)}(w^{(k)}, a^{(k-1)}, b^{(k)})), y)$$

- Analogicky pro bias a předchozí aktivaci

$$\frac{\partial C_{i_j}}{\partial b^{(k)}} = \frac{\partial z^{(k)}}{\partial b^{(k)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}} = \sigma'(z^{(k)}) 2(a^{(k)} - y)$$

$$\frac{\partial C_{i_j}}{\partial a^{(k-1)}} = \frac{\partial z^{(k)}}{\partial a^{(k-1)}} \frac{\partial a^{(k)}}{\partial z^{(k)}} \frac{\partial C_{i_j}}{\partial a^{(k)}} = w^{(k)} \sigma'(z^{(k)}) 2(a^{(k)} - y)$$

- iterativně můžeme postupovat pro
  $w^{(k-1)}, \ldots, w^{(1)}$ a pro $b^{(k-1)}, \ldots, b^{(1)}$

Backpropagation

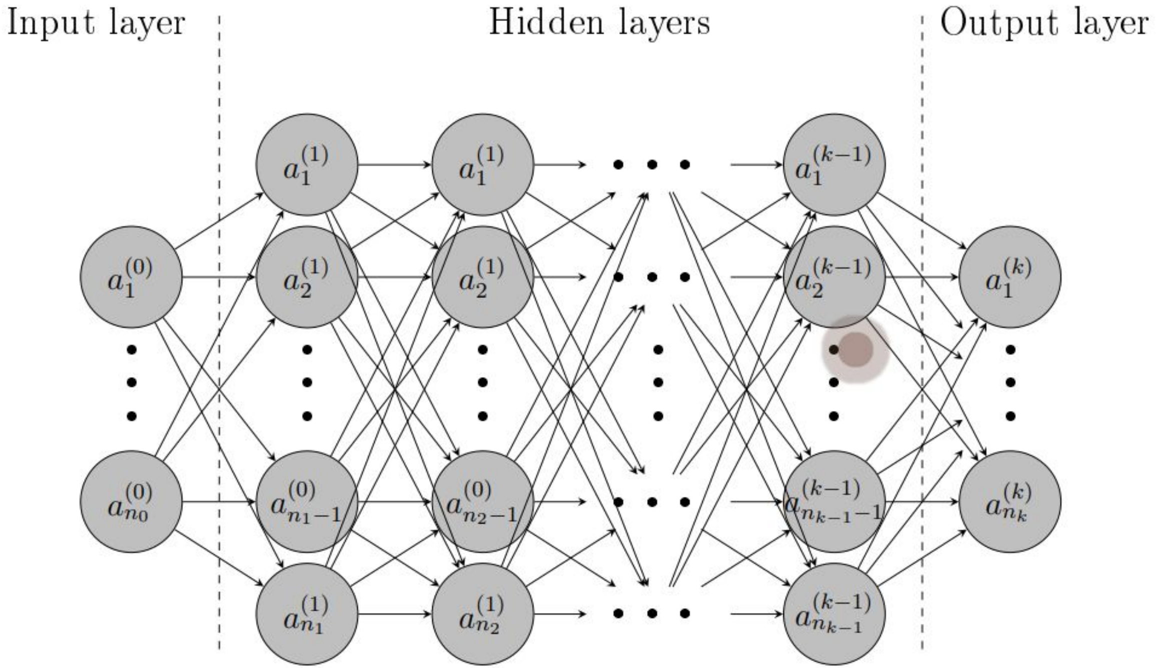
- máme vše co potřebujeme pro gradient descent

Backpropagation

- máme vše co potřebujeme pro gradient descent
- zbývá nám algoritmus zobecnit pro libovolnou síť

Backpropagation

- máme vše co potřebujeme pro gradient descent
- zbývá nám algoritmus zobecnit pro libovolnou síť



Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy

Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy
- $$C_{i_j} = \sum_{l=1}^{n_{k-1}} (a_l^{(k)} - y_l)^2$$

Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy

- $$C_{i_j} = \sum_{l=1}^{n_{k-1}} (a_l^{(k)} - y_l)^2$$

- $$z_g^{(k)} = \sum_{l=1}^{n_{k-1}} w_{gl}^{(k)} a_l^{(k-1)} + b_g^{(k)}$$

Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy
- $C_{i_j} = \sum_{l=1}^{n_{k-1}} (a_l^{(k)} - y_l)^2$
- $z_g^{(k)} = \sum_{l=1}^{n_{k-1}} w_{gl}^{(k)} a_l^{(k-1)} + b_g^{(k)}$
- $a_g^{(k)} = \sigma(z_g^{(k)})$

Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy

- $C_{i_j} = \sum_{l=1}^{n_{k-1}} (a_l^{(k)} - y_l)^2$

- $z_g^{(k)} = \sum_{l=1}^{n_{k-1}} w_{gl}^{(k)} a_l^{(k-1)} + b_g^{(k)}$

- $a_g^{(k)} = \sigma(z_g^{(k)})$

- pro derivace získáme
  velmi podobné výsledky:

Backpropagation

- index **g** značí **g**-tý neuron **k**-té vrstvy, **h** značí **h**-tý neuron **(k-1)-**vrstvy

- $C_{i_j} = \sum_{l=1}^{n_{k-1}} (a_l^{(k)} - y_l)^2$

- $z_g^{(k)} = \sum_{l=1}^{n_{k-1}} w_{gl}^{(k)} a_l^{(k-1)} + b_g^{(k)}$

- $a_g^{(k)} = \sigma(z_g^{(k)})$

- pro derivace získáme velmi podobné výsledky:

$$\frac{\partial z_g^{(k)}}{\partial w_{gh}^{(k)}} = a_h^{(k-1)}$$

$$\frac{\partial a_g^{(k)}}{\partial z_g^{(k)}} = \sigma'(z_g^{(k)}) = \frac{\exp(-z_g^{(k)})}{(\exp(-z_g^{(k)}) + 1)^2} = \sigma(z_g^{(k)})(1 - \sigma(z_g^{(k)}))$$

$$\frac{\partial C_{i_j}}{\partial a_g^{(k)}} = 2(a_g^{(k)} - y_g)$$

$$\frac{\partial z_g^{(k)}}{\partial b_g^{(k)}} = 1$$

$$\frac{\partial z_g^{(k)}}{\partial a_h^{(k-1)}} = w_{gh}^{(k)}.$$

Backpropagation

- znovu použijeme vzorec pro derivaci složené funkce

Backpropagation

- znovu použijeme vzorec pro derivaci složené funkce

$$\frac{\partial C_{i_j}}{\partial w_{gh}^{(k)}} = \frac{\partial z_g^{(k)}}{\partial w_{gh}^{(k)}} \frac{\partial a_g^{(k)}}{\partial z_g^{(k)}} \frac{\partial C_{i_j}}{\partial a_g^{(k)}} = a_h^{(k-1)} \sigma'(z_g^{(k)}) 2(a_g^{(k)} - y_g)$$

$$\frac{\partial C_{i_j}}{\partial b_g^{(k)}} = \frac{\partial z_g^{(k)}}{\partial b_g^{(k)}} \frac{\partial a_g^{(k)}}{\partial z_g^{(k)}} \frac{\partial C_{i_j}}{\partial a_g^{(k)}} = \sigma'(z_g^{(k)}) 2(a_g^{(k)} - y_g)$$

$$\frac{\partial C_{i_j}}{\partial a_h^{(k-1)}} = \sum_{l=1}^{n_{k-1}} \frac{\partial z_l^{(k)}}{\partial a_h^{(k-1)}} \frac{\partial a_l^{(k)}}{\partial z_l^{(k)}} \frac{\partial C_{i_j}}{\partial a_l^{(k)}} = 2 \sum_{l=1}^{n_{k-1}} w_{lh}^{(k)} \sigma'(z_l^{(k)}) (a_l^{(k)} - y_l)$$

Backpropagation

- znovu použijeme vzorec pro derivaci složené funkce

$$\frac{\partial C_{i_j}}{\partial w_{gh}^{(k)}} = \frac{\partial z_g^{(k)}}{\partial w_{gh}^{(k)}} \frac{\partial a_g^{(k)}}{\partial z_g^{(k)}} \frac{\partial C_{i_j}}{\partial a_g^{(k)}} = a_h^{(k-1)} \sigma'(z_g^{(k)}) 2(a_g^{(k)} - y_g)$$

$$\frac{\partial C_{i_j}}{\partial b_g^{(k)}} = \frac{\partial z_g^{(k)}}{\partial b_g^{(k)}} \frac{\partial a_g^{(k)}}{\partial z_g^{(k)}} \frac{\partial C_{i_j}}{\partial a_g^{(k)}} = \sigma'(z_g^{(k)}) 2(a_g^{(k)} - y_g)$$

$$\frac{\partial C_{i_j}}{\partial a_h^{(k-1)}} = \sum_{l=1}^{n_{k-1}} \frac{\partial z_l^{(k)}}{\partial a_h^{(k-1)}} \frac{\partial a_l^{(k)}}{\partial z_l^{(k)}} \frac{\partial C_{i_j}}{\partial a_l^{(k)}} = 2 \sum_{l=1}^{n_{k-1}} w_{lh}^{(k)} \sigma'(z_l^{(k)}) (a_l^{(k)} - y_l)$$

- v reálu se nepočítá prvek po prvku, ale využívá se maticového násobení a Hadamardova produktu (viz kniha) - příště