MŰEGYETEM 1782

BME TMIT

Gyires-Tóth Bálint

Deep Learning a gyakorlatban Python és LUA alapon
Backpropagation

SmartLab Intelligent Interactions

http://smartlab.tmit.bme.hu

NVIDIA GPU EDUCATION CENTER

Jogi nyilatkozat

Jelen előadás diái a „Deep Learning a gyakorlatban Python és LUA alapon” című tantárgyhoz készültek és letölthetők a http://smartlab.tmit.bme.hu honlapról.

A diák nem helyettesítik az előadáson való részvételt, csupán emlékeztetőül szolgálnak.

Deep Learning Híradó

Hírek az elmúlt 168 órából

SmartLab
Intelligent Interactions

http://smartlab.tmit.bme.hu

NVIDIA GPU EDUCATION CENTER

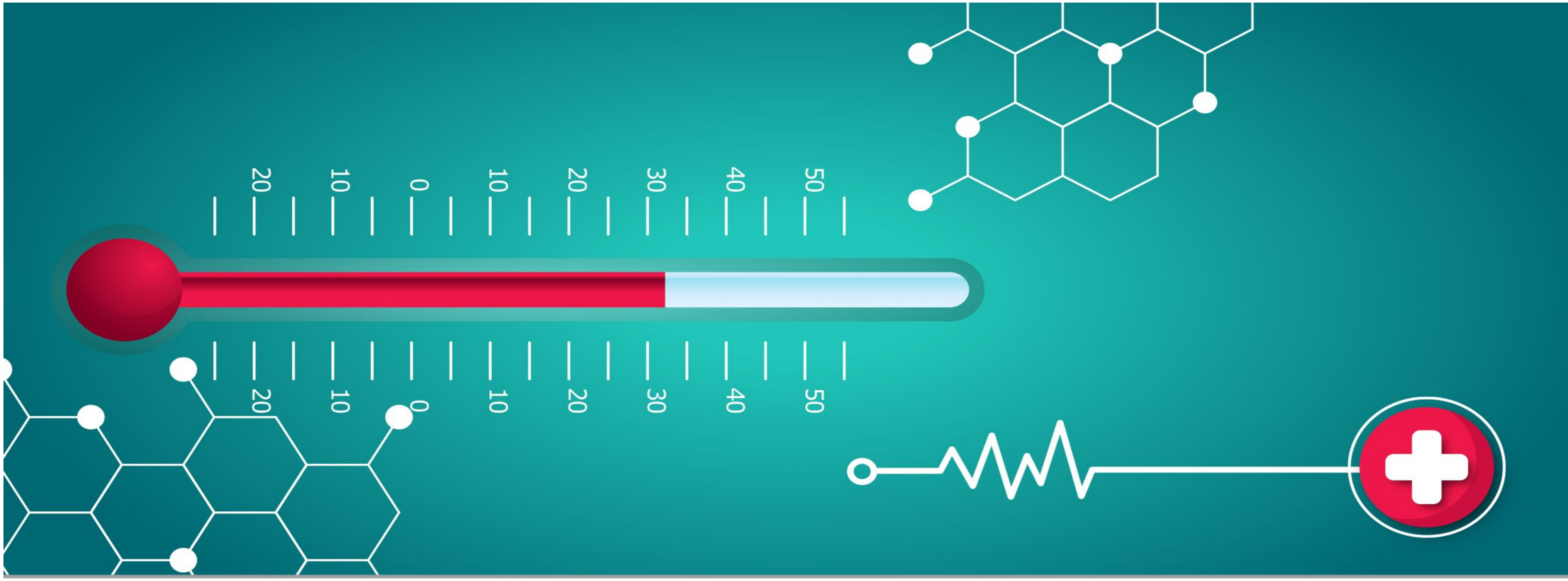

INTERSPEECH
2019

https://nvidiadli-20191001.eventbrite.com

NVIDIA | DEEP LEARNING INSTITUTE

TEACHING YOU TO SOLVE PROBLEMS WITH DEEP LEARNING

Tanítási feladat, adatok



- *Bemenet:* `[láz (°C), gyógyszer (mg)]` $= X$
- *Kimenet*
  - *Regresszió:* `[láz 2 óra múlva]` $= y$
  - *Osztályozás:* `[láz/hőemelkedés/normális 2 óra múlva]`

Tanító és teszt adatok

- *Példa*

$$X = \begin{bmatrix} 38.6 & 25 \\ 37.8 & 25 \\ 37.9 & 50 \\ 38.2 & 50 \end{bmatrix} \qquad y = \begin{bmatrix} 37.2 \\ 37.3 \\ 36.6 \\ 36.9 \end{bmatrix} \qquad X\_test = \begin{bmatrix} 38.3 & 35 \end{bmatrix}$$

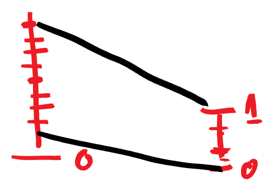
$$\hat{y} = ?$$

$$x = (x - \text{mean}X) / \text{std}X$$

→ szórás

→ várható érték

$$y = \text{minmax}(y, 0, 1)$$



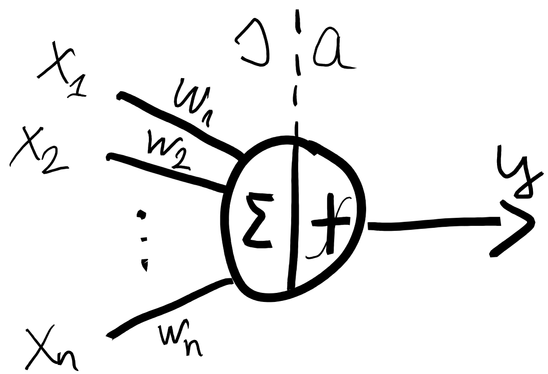
Min max scaler

Kimenetek átskálázása

$$y\_std = (y - \min y) / (\max y - \min y)$$

$$y\_scaled = y\_std \cdot (max - min) + min \Big|_{\substack{max = 1 \\ min = 0}}$$

Elemi neuron



„Szumma"

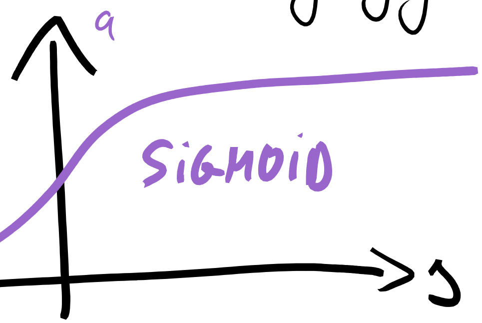
$$s = \sum_{i=1}^{n} w_i x_i$$

$$y = a(s)$$

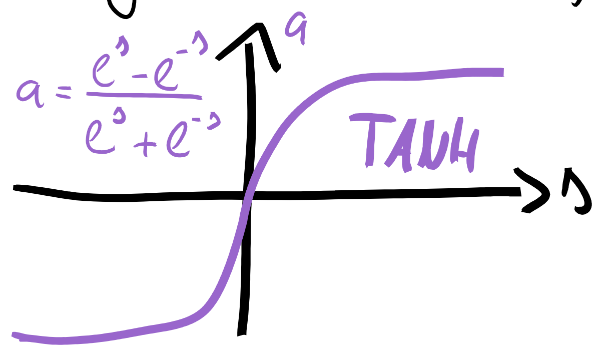
„aktivációs"

a: aktivációs függvény (sgm, tanh, ReLU...)

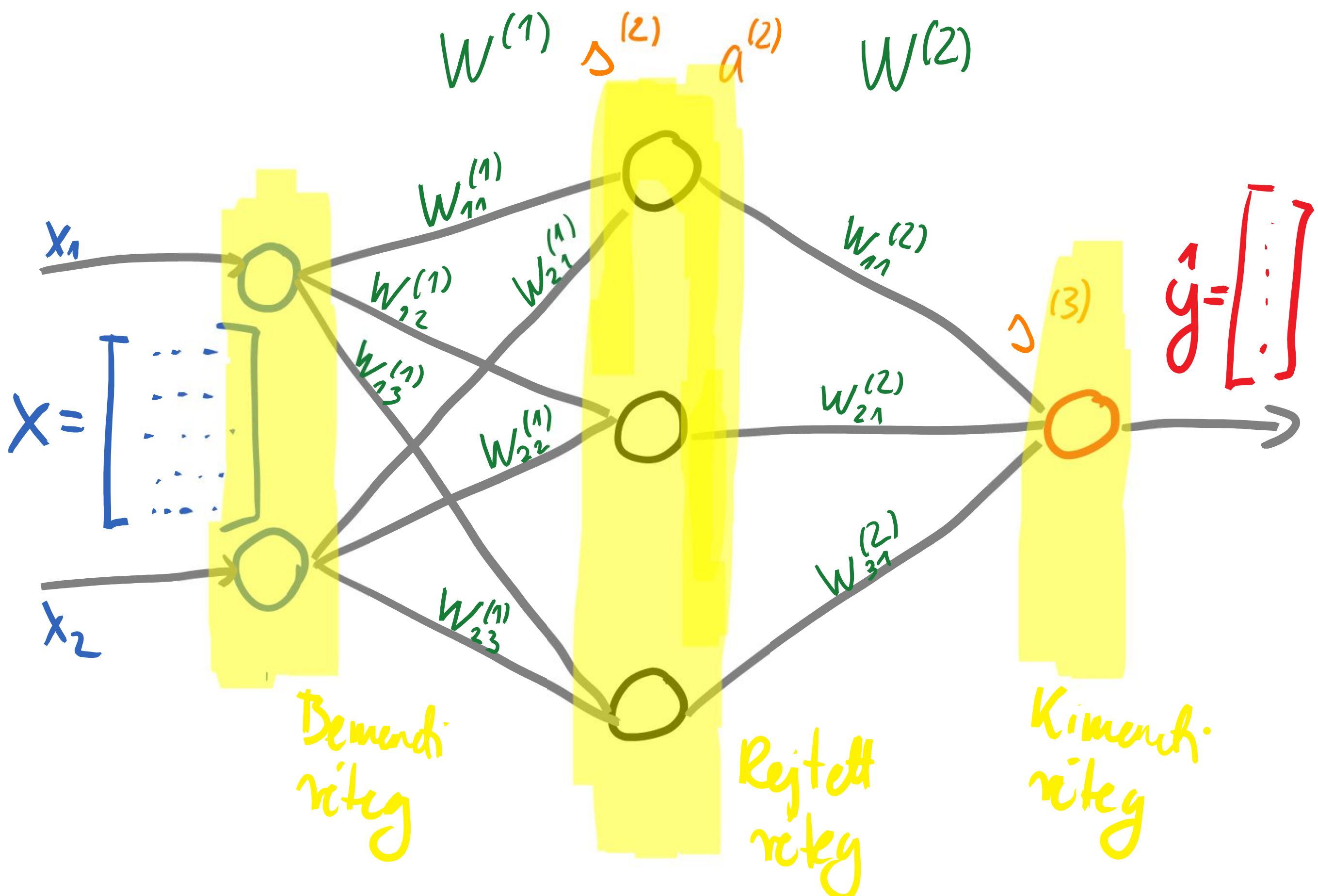
$$a = \frac{1}{1+e^{-s}}$$



$$a = \frac{e^{s}-e^{-s}}{e^{s}+e^{-s}}$$



Előrecsatolt mély neurális hálózat



Mátrixalgebra

- *Mátrix szorzás*
- *Transzponálás*
- *Parciális derivált*

NumPy torch theano TensorFlow

MATLAB

Forward lépés I.

$$\text{①}\quad \overset{(4\times2)}{X}\,\overset{(2\times3)}{W^{(1)}} = \overset{(4\times3)}{z^{(2)}}$$

$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix}$$

$$\text{minták száma}\left\{ X = \underbrace{\begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ x_1^{(3)} & x_2^{(3)} \\ x_1^{(4)} & x_2^{(4)} \end{bmatrix}}_{\text{tulajdonságok (features)}} \right.$$

$$\begin{bmatrix} x_1^{(1)} w_{11}^{(1)} + x_2^{(1)} \cdot w_{21}^{(1)} & x_1^{(1)} w_{12}^{(1)} + x_2^{(1)} w_{22}^{(1)} & x_1^{(1)} w_{13}^{(1)} + x_2^{(1)} w_{23}^{(1)} \\ x_1^{(2)} w_{11}^{(1)} + x_2^{(2)} \cdot w_{21}^{(1)} & x_1^{(2)} w_{12}^{(1)} + x_2^{(2)} w_{22}^{(1)} & x_1^{(2)} w_{13}^{(1)} + x_2^{(2)} w_{23}^{(1)} \\ x_1^{(3)} \cdot w_{11}^{(1)} + x_2^{(3)} \cdot w_{21}^{(1)} & x_1^{(3)} w_{12}^{(1)} + x_2^{(3)} w_{22}^{(1)} & x_1^{(3)} w_{13}^{(1)} + x_2^{(1)} w_{23}^{(1)} \\ x_1^{(4)} w_{11}^{(1)} + x_2^{(4)} w_{21}^{(1)} & x_1^{(4)} w_{12}^{(1)} + x_2^{(4)} w_{22}^{(1)} & x_1^{(4)} w_{13}^{(1)} + x_2^{(1)} w_{23}^{(1)} \end{bmatrix} = z^{(2)}$$

Forward lépés II.

② $a^{(2)} = f(s^{(2)}) = \text{Sigmoid}\{s^{(2)}\}$ (tanh/ReLU/PReLU...)

$\text{sigmoid} \equiv \text{sgm}$

$$\overset{(4\times3)}{a^{(2)}} = \begin{bmatrix} \text{sgm}(s^{(2)}_{11}) & \text{sgm}(s^{(2)}_{12}) & \text{sgm}(s^{(2)}_{13}) \\ \vdots & \ddots & \vdots \\ \dots & \dots & \text{sgm}(s^{(2)}_{33}) \end{bmatrix}$$

③ $\overset{(4\times1)}{s^{(3)}} = \overset{(4\times3)}{a^{(2)}} \, \overset{(3\times1)}{W^{(2)}}$

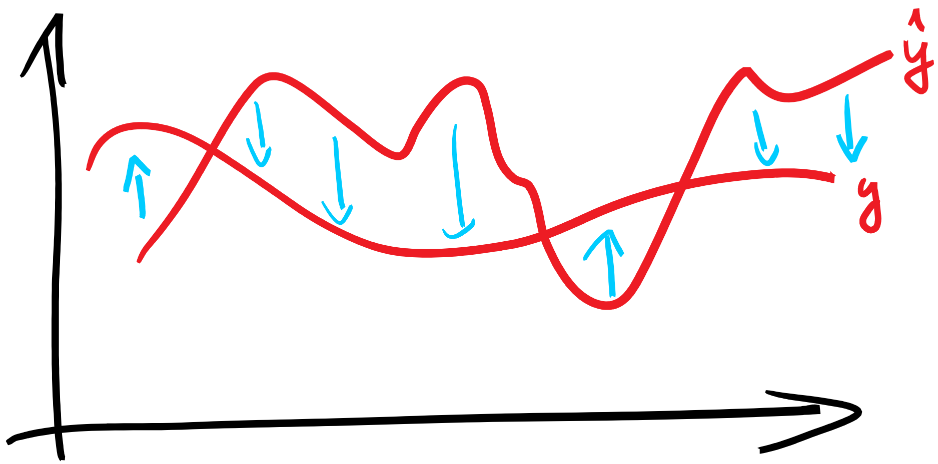
④ $\overset{(4\times1)}{\hat{y}} = f(s^{(3)}) = \text{sigmoid}\{s^{(3)}\}$ (tanh/ReLU/PReLU...)

Forward lépés III.: Cost function

- *Magyarul: költségfüggvény*
- *Négyzetes hiba (regresszió) vagy keresztentrópia (osztályozás)*
- *Kezdetben súlyok értéke véletlenszám*

5

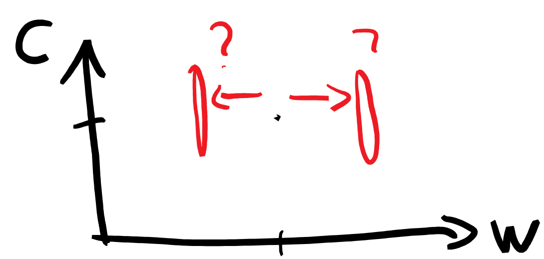
$$C = \sum \frac{1}{2} (y - \hat{y})^2$$



Súlyok finomhangolása

- Brute force
  1 súly → [-100,100] 0.1-es felbontással 2000 eset
  2 súly → $2000^2 = 4 * 10^6$
  …
  9 súly → $2000^9 = 4 * 10^6 = 512 * 10^{27}$

  közepes méretű háló, 6 rejtett réteg, 1000 neuron rétegenként: $6 * 1000^2$ súly
- Numerikus gradiens keresés



Kahoot!

Névnek a NEPTUN kódodat add meg!

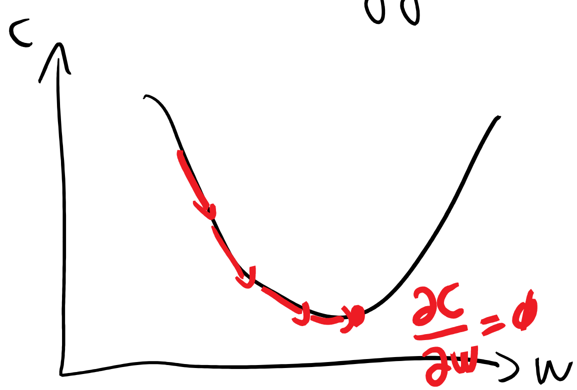
https://kahoot.it/

Gradient descent

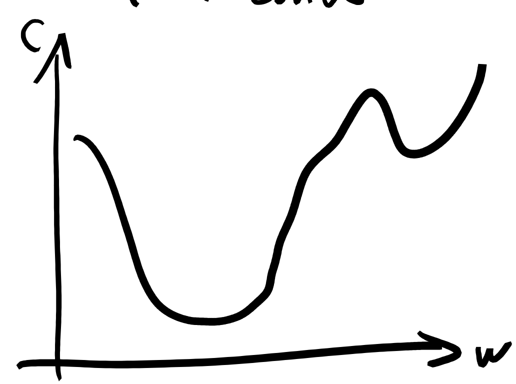
⑥ $C = \sum \left\{ \frac{1}{2} \left(y - f\left(f\left(X W^{(1)} \right) W^{(2)} \right) \right)^2 \right\}$

$\frac{\partial C}{\partial w} \quad > 0 \nearrow \qquad < 0 \searrow$

Cél => elérni hogy $= 0$



Non-convex



Backprop: kimeneti – rejtett réteg

$$\frac{\partial C}{\partial W^{(2)}} = \frac{\partial \sum \frac{1}{2}(y-\hat{y})^2}{\partial W^{(2)}} \overset{ⓐ}{=} \sum \frac{\partial \frac{1}{2}(y-\hat{y})^2}{\partial W^{(2)}}$$

ⓑ

$$\Rightarrow \cancel{\tfrac{1}{2}}(y-\hat{y})^{\cancel{2}} \cdot - \frac{\partial \hat{y}}{\partial W^{(2)}} \overset{ⓒ}{=}$$

$$= -(y-\hat{y}) \cdot \frac{\partial \hat{y}}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial W^{(2)}} =$$

$$= \underbrace{-(y-\hat{y}) \cdot f'(z^{(3)})}_{\substack{\delta^{(3)} \\ (1\times 1)}} \cdot \underbrace{\frac{\partial a^{(2)} W^{(2)}}{\partial W^{(2)}}}_{\substack{a^{(2)} \\ (1\times 3)}} \quad *$$

ⓐ $\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}$

ⓑ lánc szabály:

$(f \cdot g)' = (f' \cdot g) \cdot g$

$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$

ⓓ $f(z) = \frac{1}{1+e^{-z}} \; = v$

$f'(\;) = \frac{u'v - uv'}{v^2} = \frac{0 \cdot \ldots - 1 \cdot -e^{-z}}{(1+e^{-z})^2} =$

$= \frac{e^{-z}}{(1+e^{-z})^2}$

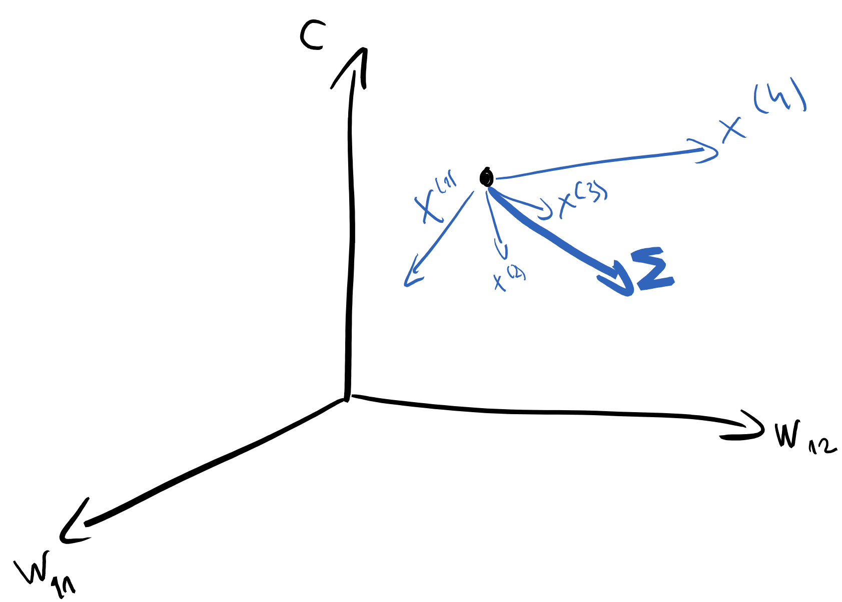
Backprop: Batch gradient descent

$$7\quad \overset{(3\times 1)}{\frac{\partial C}{\partial W^{(2)}}} = \sum \frac{\partial \frac{1}{2}(y-\hat{y})^2}{\partial W^{(2)}} = \overset{(3\times 4)}{\left(a^{(2)}\right)^T} \overset{(4\times 1)}{\delta^{(3)}} = \begin{bmatrix} \delta_1^{(3)} \\ \delta_2^{(3)} \\ \delta_3^{(3)} \\ \delta_4^{(3)} \end{bmatrix}$$

$$= \begin{bmatrix} \overbrace{a_{11}^{(2)} \quad a_{21}^{(2)} \quad a_{31}^{(2)} \quad a_{41}^{(2)}}^{\text{tanító minták száma}} \\ a_{12}^{(2)} \quad a_{22}^{(2)} \quad a_{32}^{(2)} \quad a_{42}^{(2)} \\ a_{13}^{(2)} \quad a_{23}^{(2)} \quad a_{33}^{(2)} \quad a_{43}^{(2)} \end{bmatrix} \begin{bmatrix} a_{11}^{(2)}\delta_1^{(3)} + a_{21}^{(2)}\delta_2^{(3)} + a_{31}^{(2)}\delta_3^{(3)} + a_{41}^{(2)}\delta_4^{(3)} \\ a_{12}^{(2)}\delta_1^{(3)} + a_{22}^{(2)}\delta_2^{(3)} + a_{32}^{(2)}\delta_3^{(3)} + a_{42}^{(2)}\delta_4^{(3)} \\ a_{13}^{(2)}\delta_1^{(3)} + a_{23}^{(2)}\delta_2^{(3)} + a_{33}^{(2)}\delta_3^{(3)} + a_{43}^{(2)}\delta_4^{(3)} \end{bmatrix}$$

Backprop: Batch gradient descent

Az összes tanítómintára kiszámoljuk a gradienst és ezeket összegezzük.



Backprop: rejtett – bemeneti réteg

$$\frac{\partial C}{\partial w^{(1)}} = \frac{\partial \frac{1}{2}(y-\hat{y})^2}{\partial w^{(1)}} = -(y-\hat{y})\cdot\frac{\partial \hat{y}}{\partial w^{(1)}} = -(y-\hat{y})\cdot\frac{\partial \hat{y}}{\partial z^{(3)}}\cdot\frac{d z^{(3)}}{d w^{(1)}} = -(y-\hat{y})\cdot f'(z^{(3)})\frac{d z^{(3)}}{d w^{(1)}} =$$

$$= \delta^{(3)}\cdot\frac{\partial z^{(3)}}{\partial w^{(1)}} = \delta^{(3)}\cdot\frac{\partial z^{(3)}}{\partial a^{(2)}}\cdot\frac{\partial a^{(2)}}{\partial w^{(1)}} = \delta^{(3)}\cdot\left(W^{(2)}\right)^T\cdot\frac{\partial a^{(2)}}{\partial w^{(1)}} =$$

$$= \delta^{(3)}\left(W^{(2)}\right)^T\frac{\partial a^{(2)}}{\partial z^{(2)}}\cdot\frac{d z^{(2)}}{\partial w^{(1)}} = \underbrace{\delta^{(3)}\cdot\left(W^{(2)}\right)^T\cdot f'(z^{(2)})}_{\delta^{(2)}}\underbrace{\frac{\partial z^{(2)}}{\partial w^{(1)}}}_{X} = X^T\cdot\delta^{(2)}$$

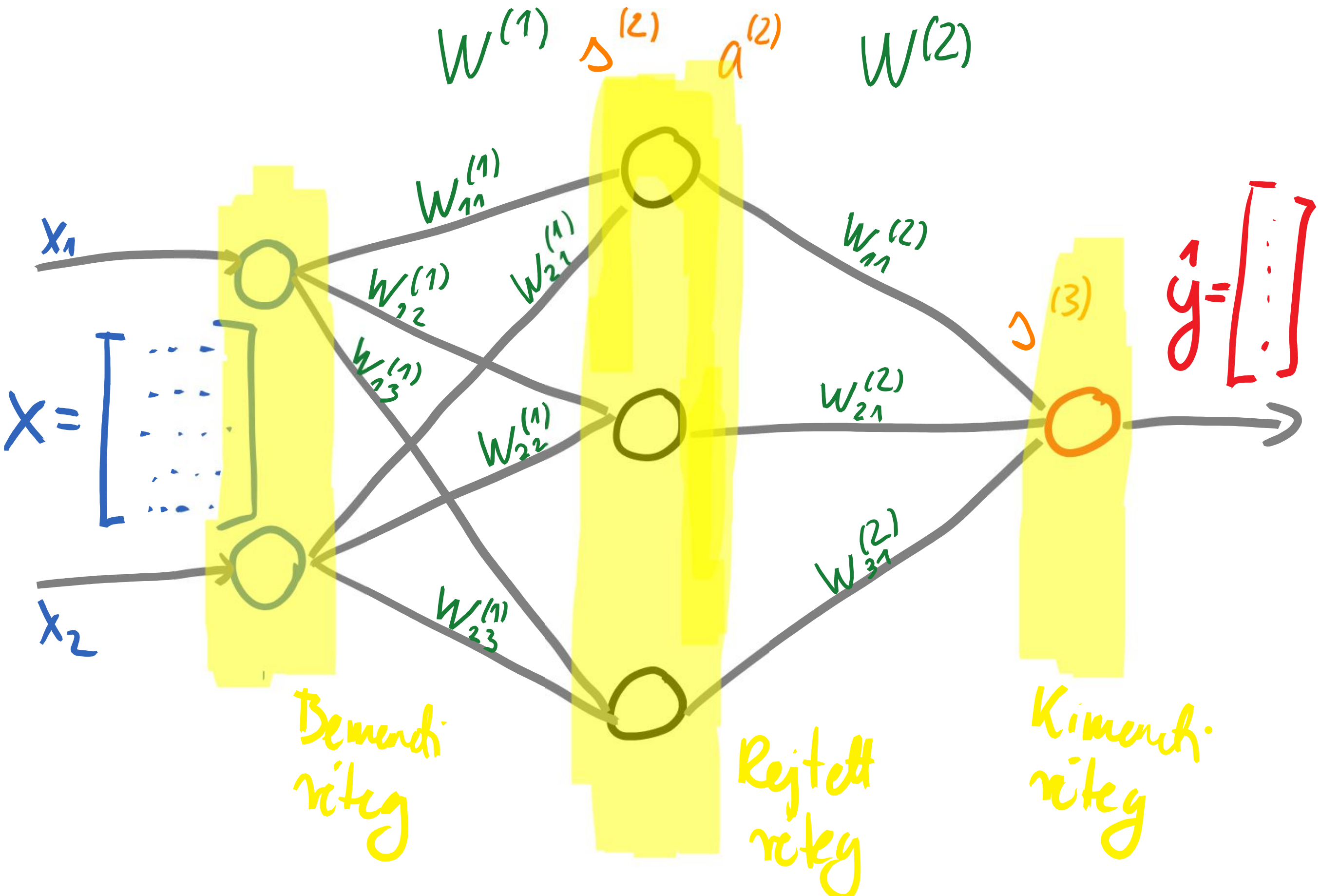
TANULÁS

$$W^{(1)} = W^{(1)} - \mu\frac{\partial C}{\partial w^{(1)}}$$

$$W^{(2)} = W^{(2)} - \mu\frac{\partial C}{\partial w^{(2)}}$$

μ: tanulási ráta (learning rate)

Előrecsatolt mély neurális hálózat



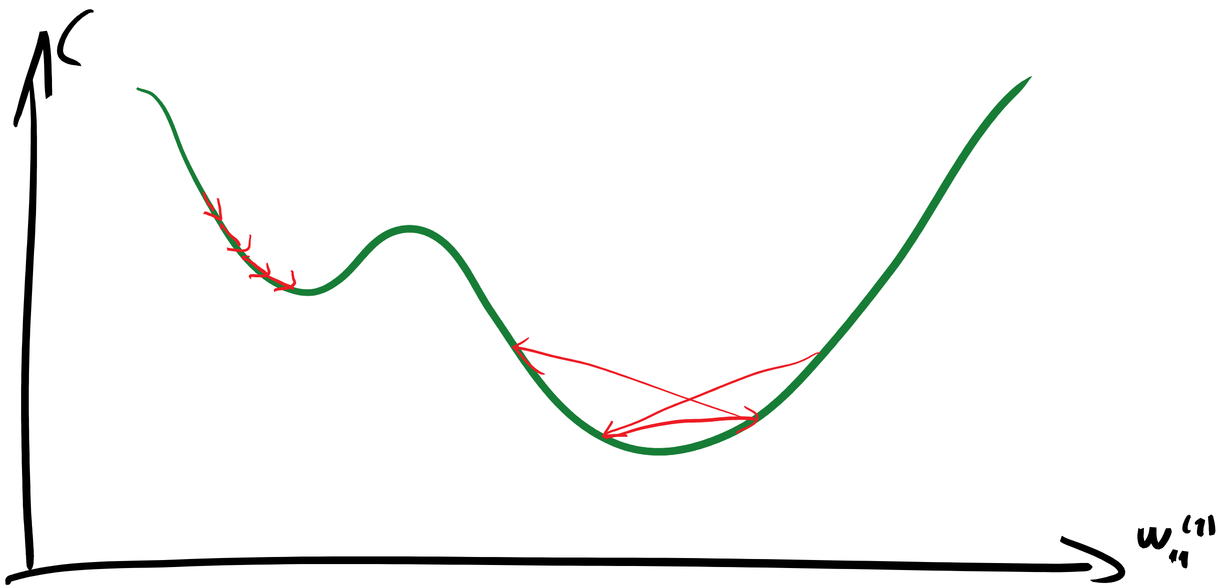
Gradient descent

- *Stochastic Gradient Descent (SGD)*
- *Batch GD*
- *Mini-Batch GD*

Gradient descent

Lokális vs. Globális minimum

Vanilla SGD, Batch learning, Mini-batch



Intuitív magyarázat

- Deep learning: több rétegen keresztül ugyanez
- Yann LeCunn: Efficient Backprop (1998)
  http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf
- További magyarázat és online demo:
  https://www.deeplearning.ai/ai-notes/optimization/

Acknowledgement

The picture of Slide 6 was designed by Freepik.com.

Köszönöm a figyelmet!

toth.b@tmit.bme.hu