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.

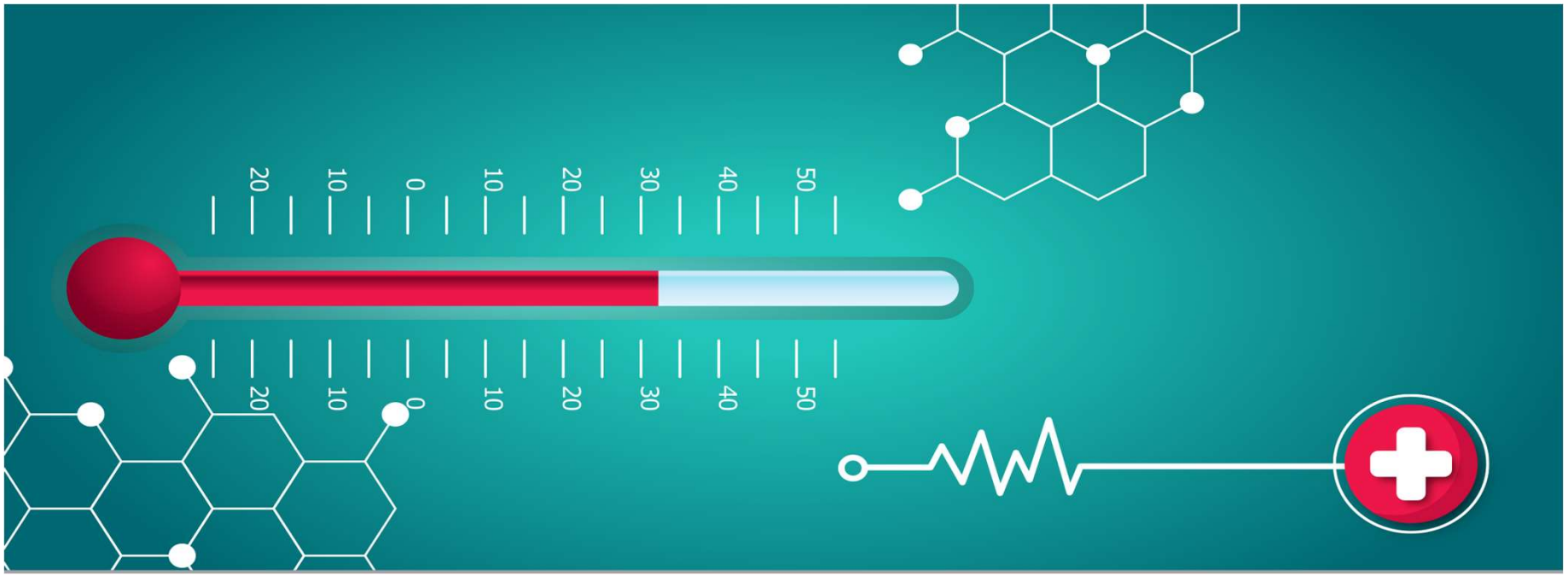
Kis házi feladat

- GitHubon hirdetve
- Beadni: http://smartlab.tmit.bme.hu/oktatas-deep-learning

Nagy házi feladat

- Csapatok toborzása Microsoft Teams-en
- 4. hét végéig csapat minden tagjának regisztrációja a http://smartlab.tmit.bme.hu/oktatas-deep-learning oldalon.
- Témalabor, BSc, MSc diploma
- PhD téma
- 3 fős csapatok
- 1, 2, 4 fős csapatok
- Téma javaslatok

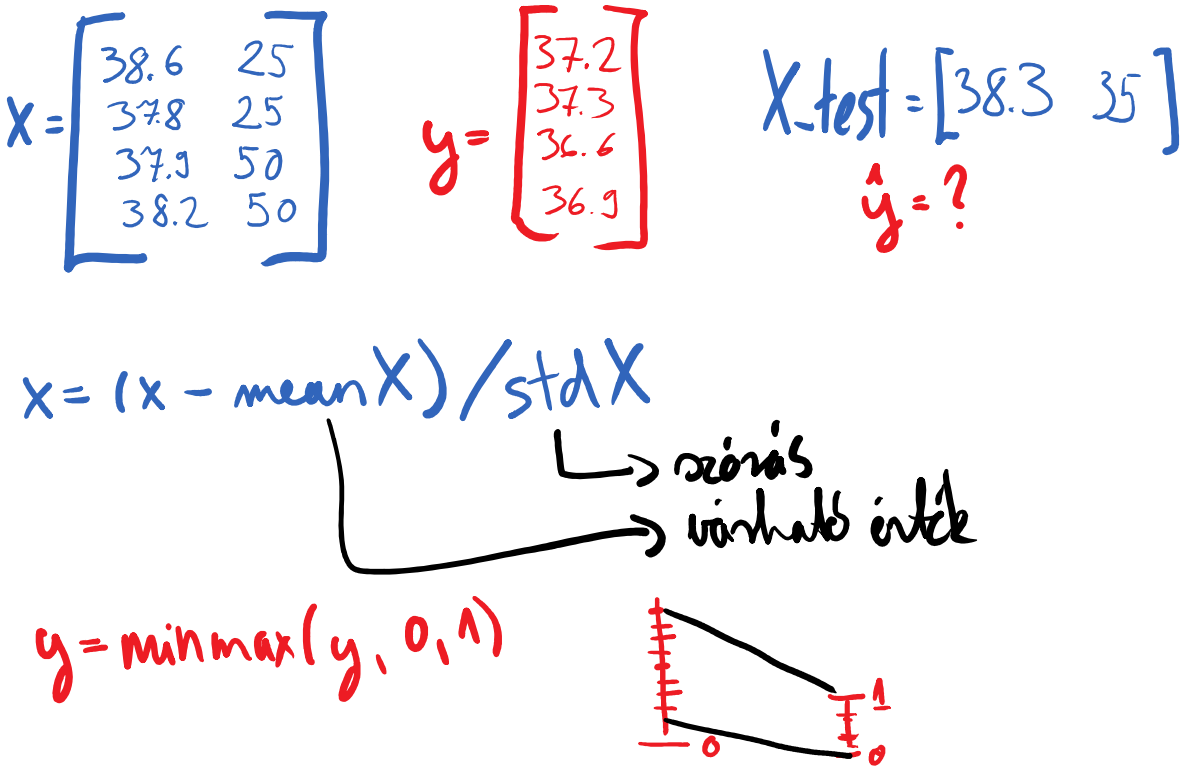
Tanítási feladat, adatok



- 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



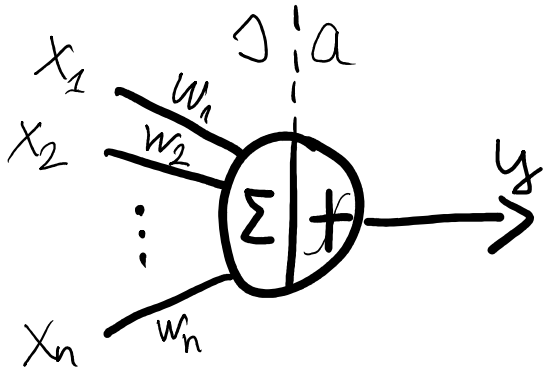
Min max scaler

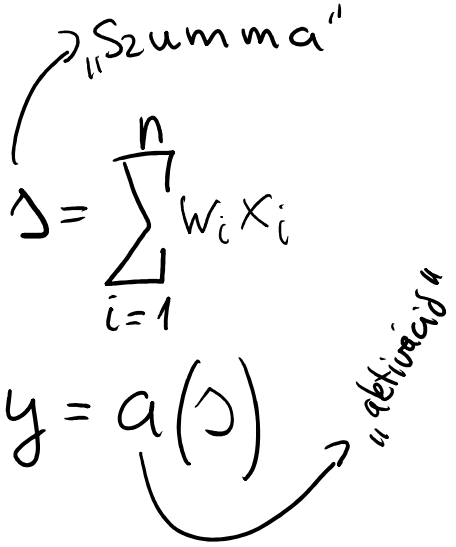
Kimenetek átskálázása

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

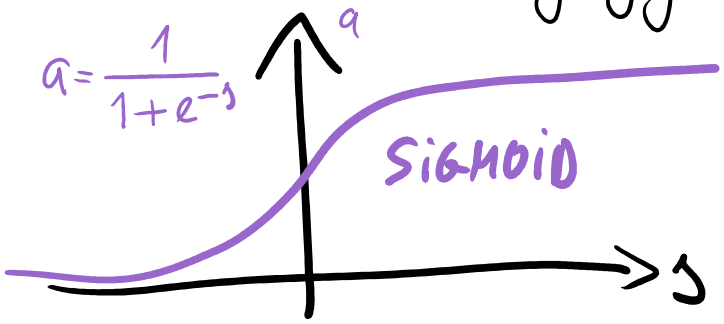
$$y\_scaled = y\_std \cdot (max - min) + min \Big|_{\substack{max=1 \\ min=0}}$$

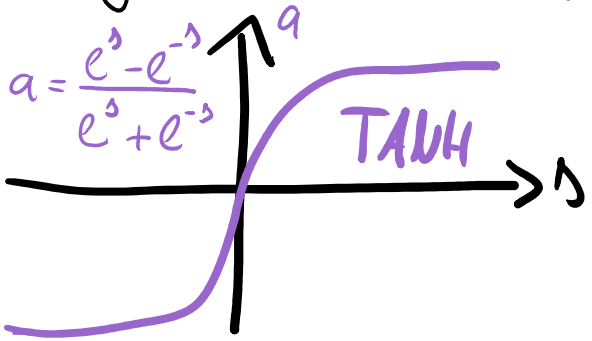
Elemi neuron



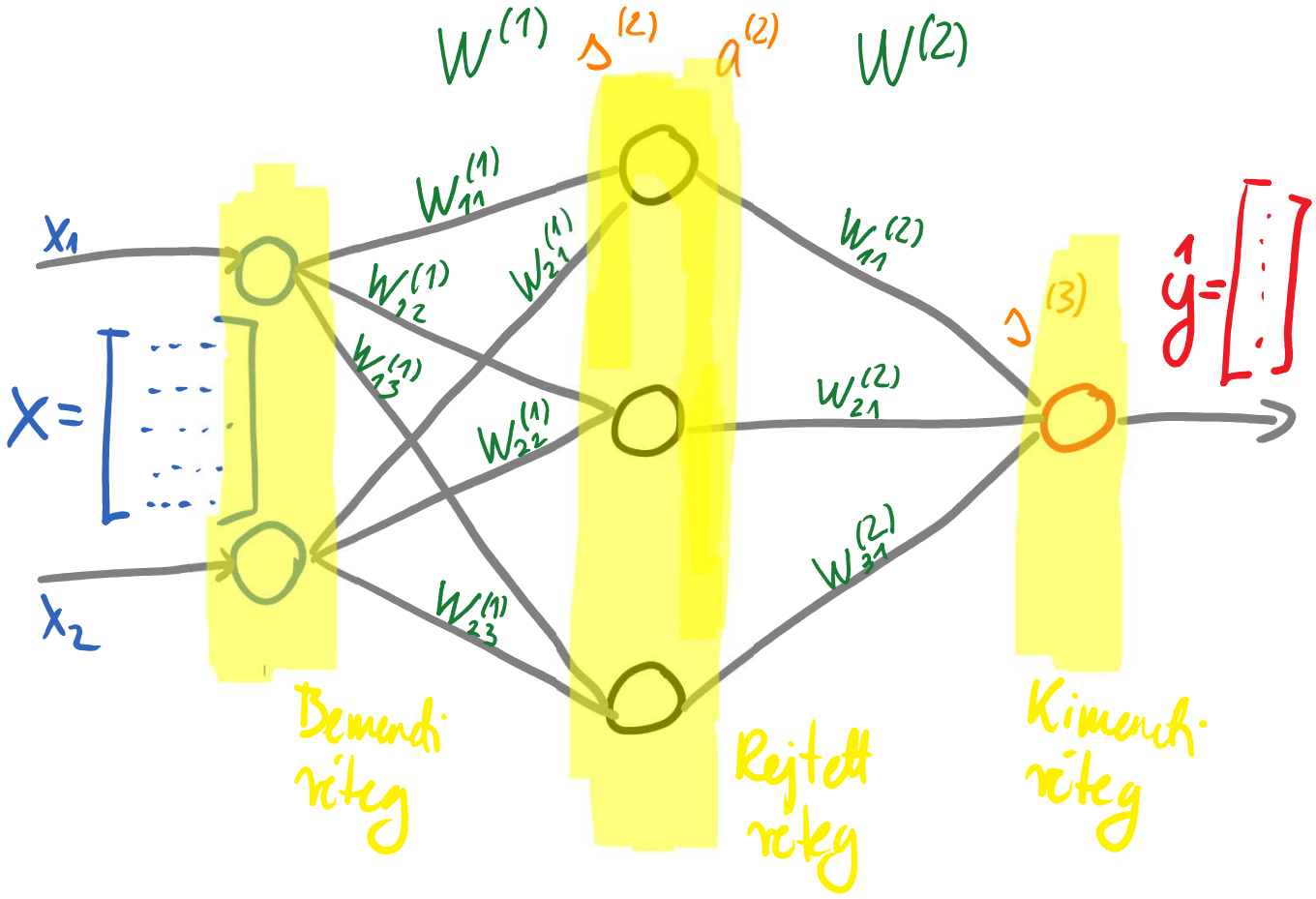


a: aktivációs függvény (sgm, tanh, ReLU...)





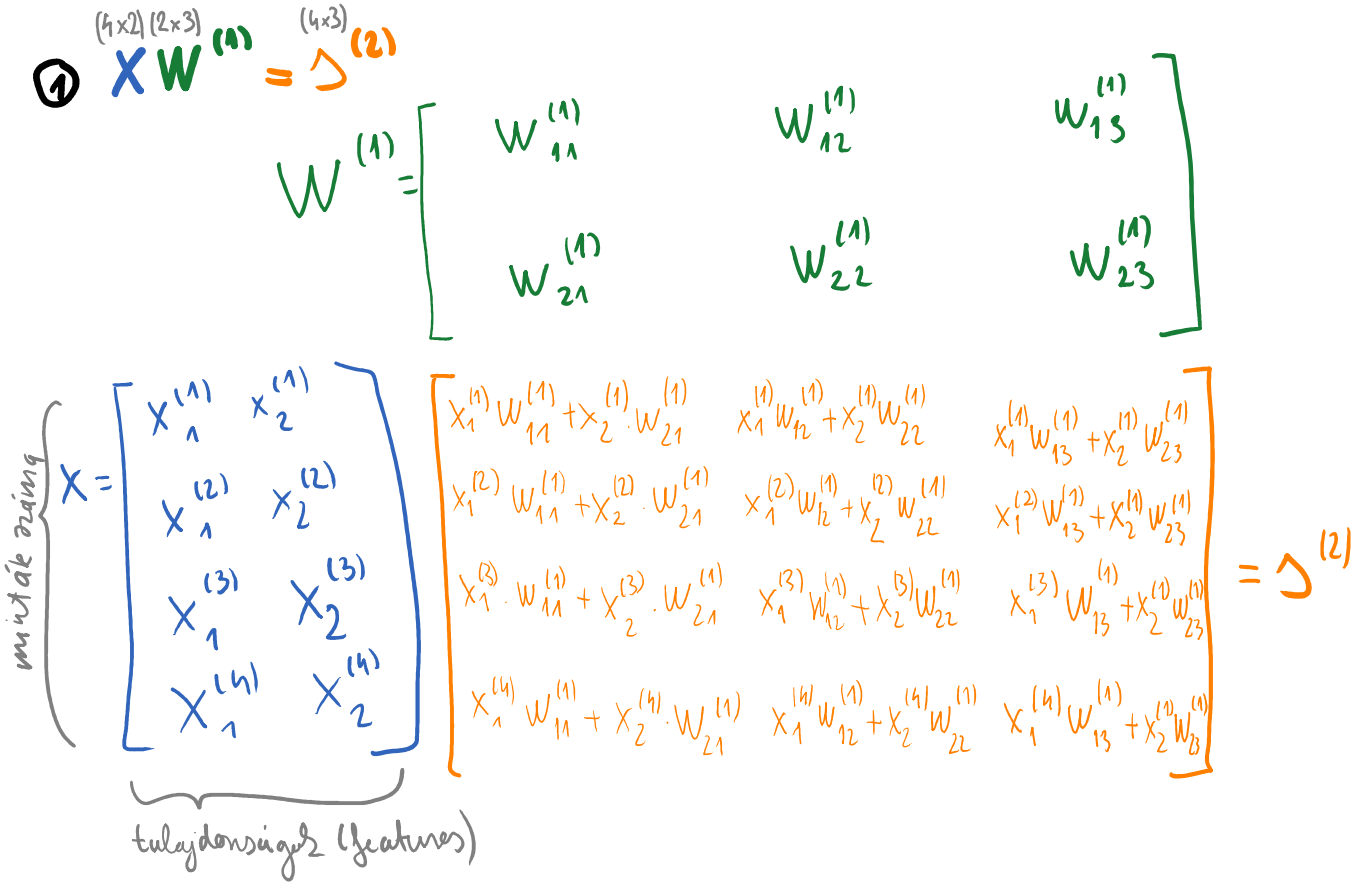
Egyszerű neurális hálózat



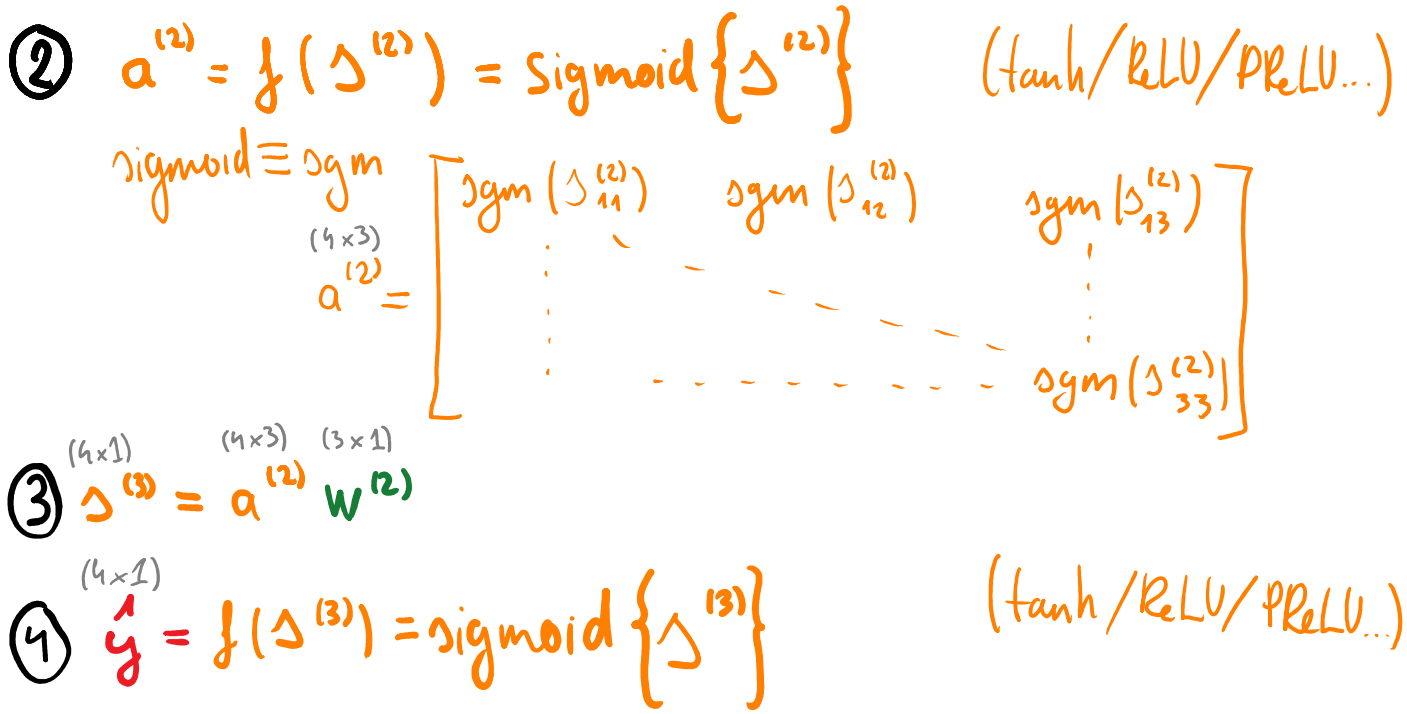
Mátrixalgebra

- Mátrix szorzás
- Transzponálás
- Parciális derivált

Forward lépés I.



Forward lépés II.

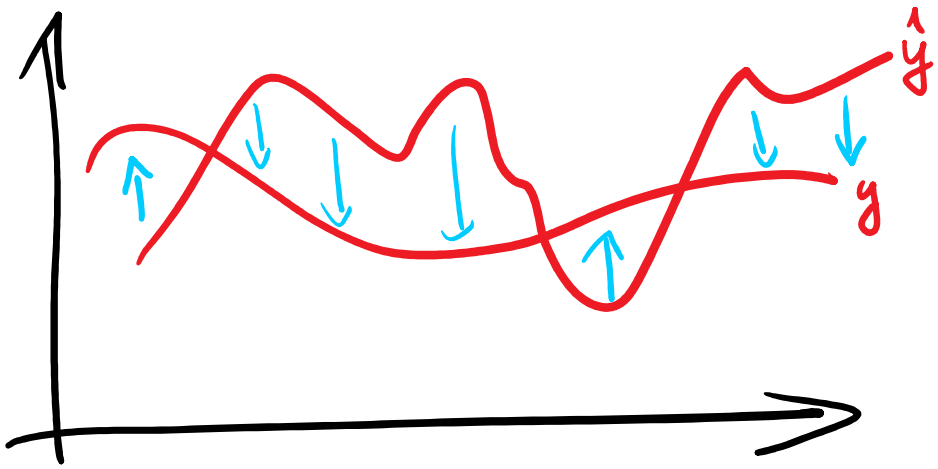


Forward lépés III.: Cost function

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

(5)

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

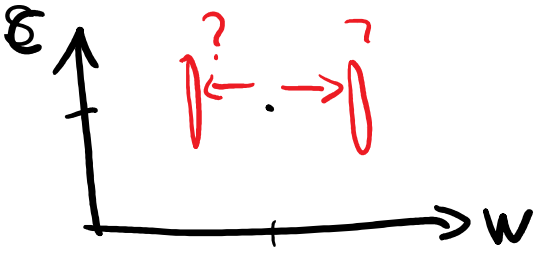


Súlyok finomhangolása

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

  …
  9 súly → $\mathbf{2000^9 = 512 * 10^{27}}$

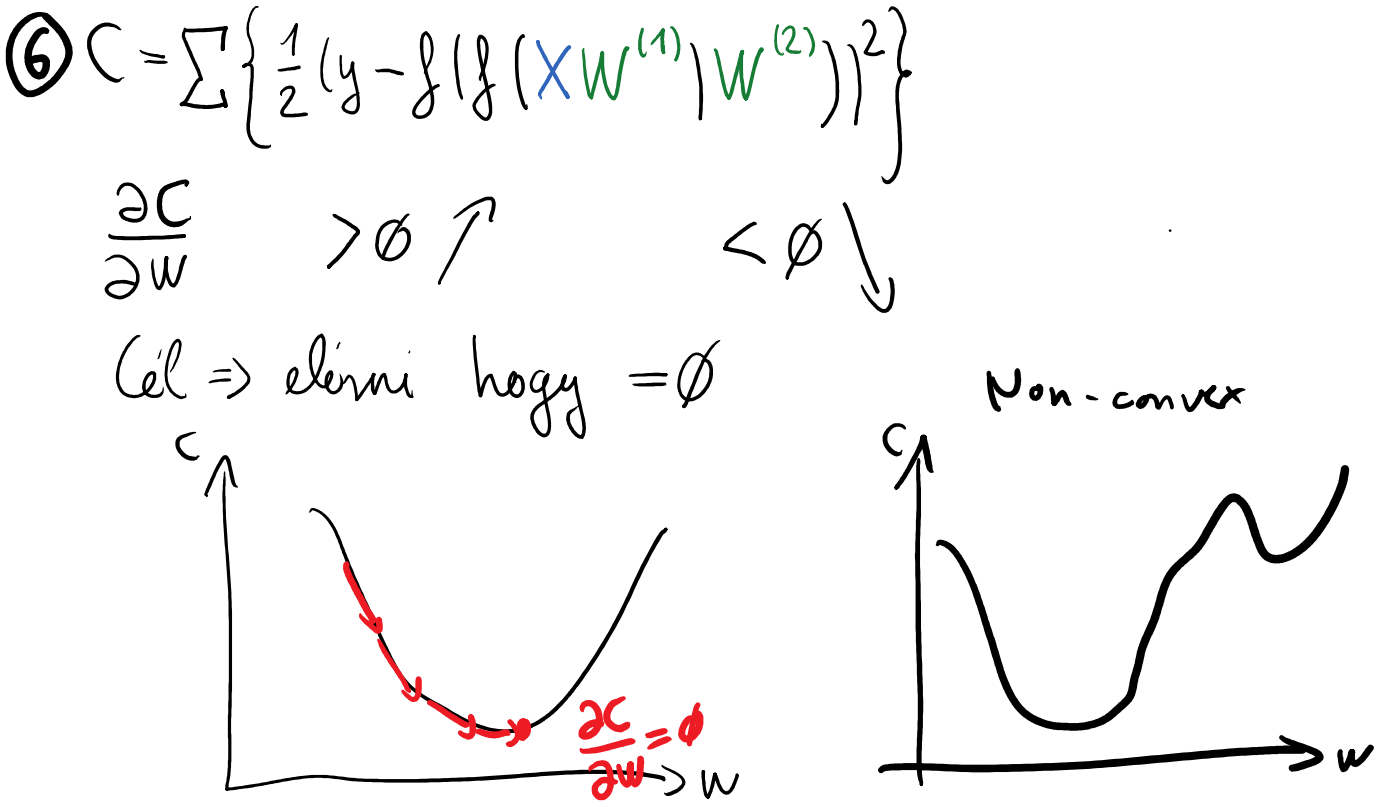
  közepes méretű háló, 6 rejtett réteg, 1000 neuron rétegenként: $\mathbf{6 * 1000^2}$ súly
- Numerikus gradiens keresés



Névnek a NEPTUN kódodat add meg!

https://kahoot.it/

Gradient descent



Backprop: kimeneti – rejtett réteg

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

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

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

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

$\delta^{(3)}$ (1×1) $\quad$ $a^{(2)}$ (1×3)

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

b) lánc szabály:

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

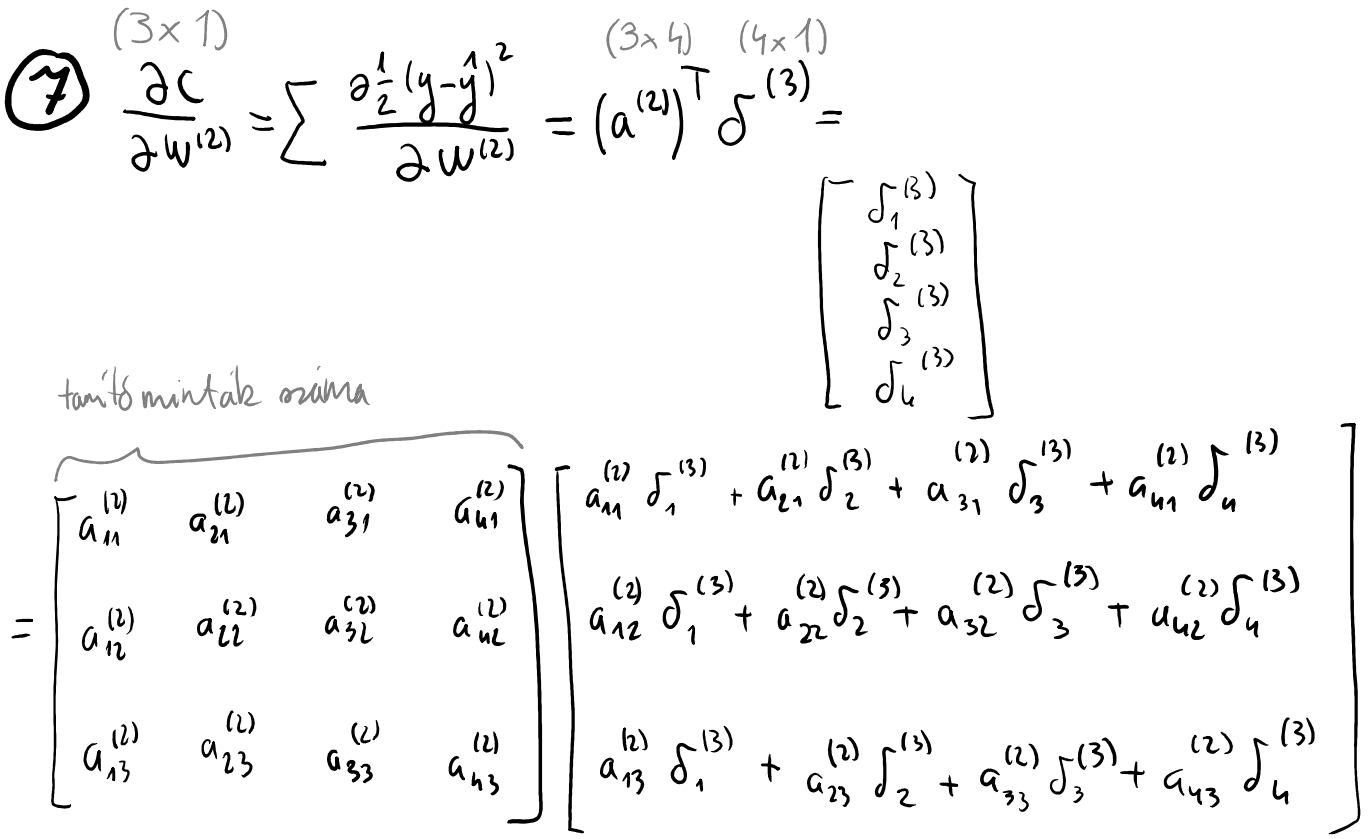
$\frac{ds}{dx} = \frac{ds}{dy} \cdot \frac{dy}{dx}$

d) $f(s) = \frac{1}{1+e^{-s}} = v$

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

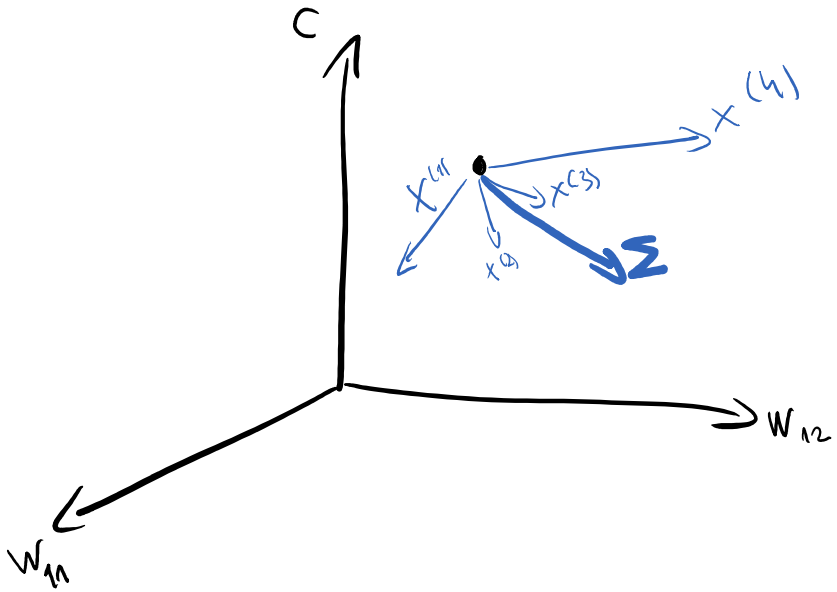
$= \frac{e^{-s}}{(1+e^{-s})^2}$

Backprop: Batch gradient descent



Backprop: Batch gradient descent

Az összes tanítómintára kiszámoljuk a gradienst és ezeket összegezzük.



Backprop: rejtett – bemeneti réteg

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

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

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

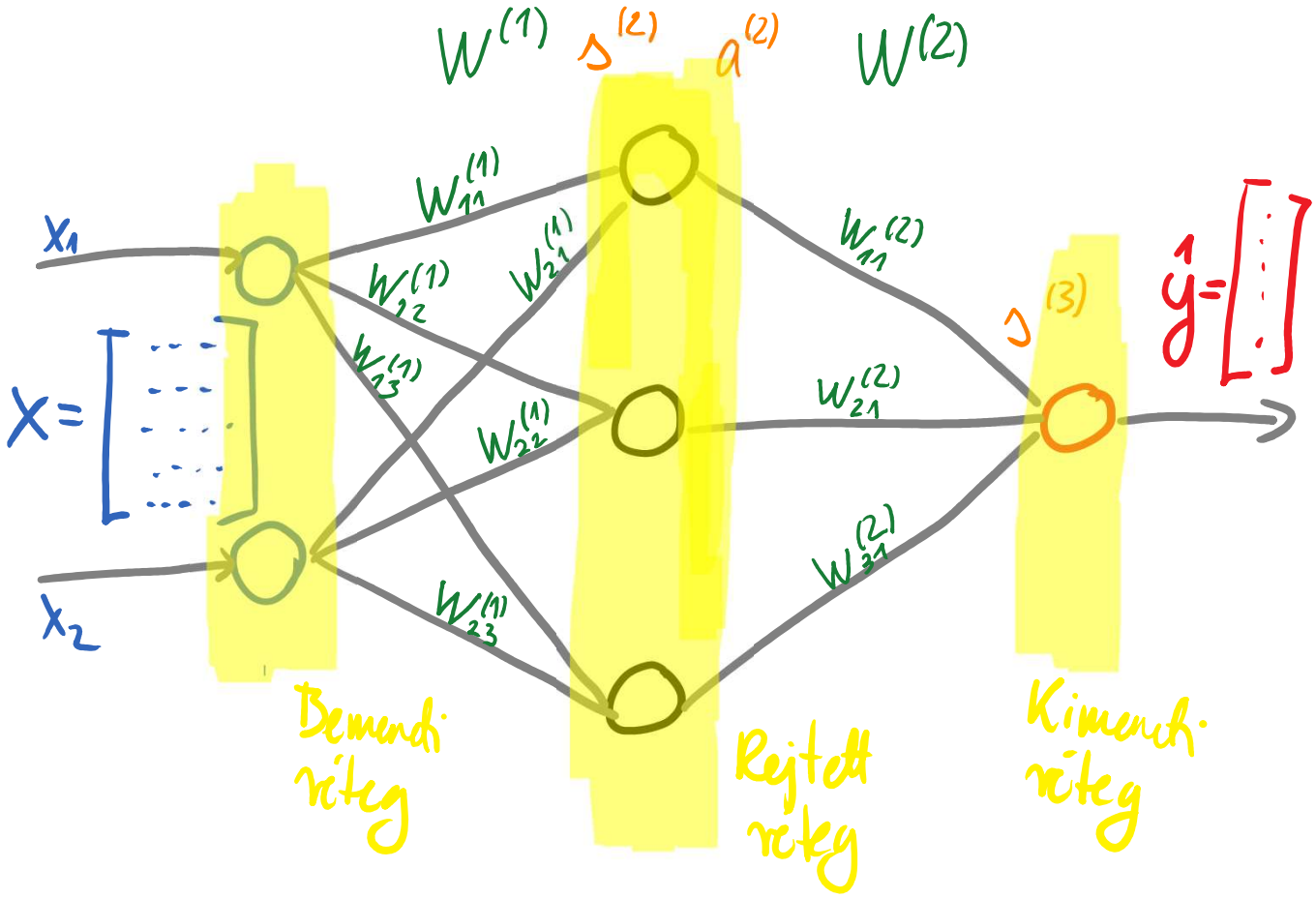
TANULÁS

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

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

μ: tanulási ráta (learning rate)

Egyszerű neurális hálózat



Gradient descent

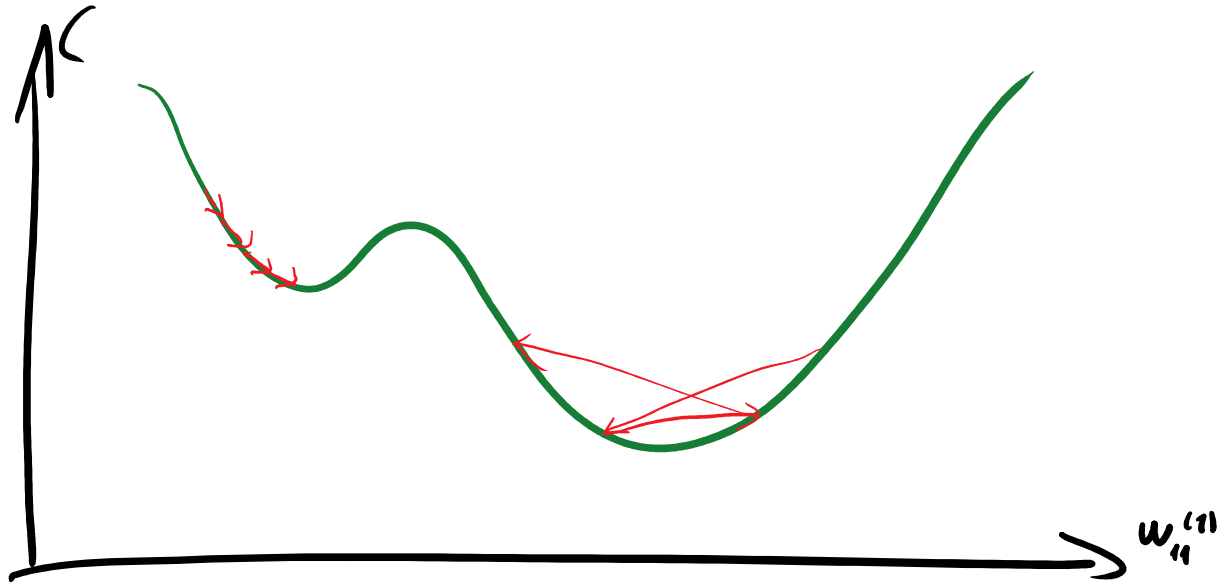
- Stochastic Gradient Descent (SGD)
- Batch GD

- Mini-Batch GD

Gradient descent

Lokális vs. Globális minimum

Batch learning, Mini-batch learning



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

SmartLab
Intelligent Interactions

http://smartlab.tmit.bme.hu

NVIDIA
GPU
EDUCATION
CENTER