

* פור 3.5 ק"מ 2012 -

תחבול : 2M

100K

- 132

15% של זכר

25%

2012
כ"ב
37
2011

הלכה: רשת ניירונים
Neural Network

logistic regression : 87%

$$a = w_1x_1 + w_2x_2 + \cdots + w_Nx_N + b = \mathbf{w}^T \mathbf{x} + b \quad (13.1)$$

$$z = g(a) \quad (13.2)$$

$$z = g(a)$$

יציאה מ-NN (ההפוך) פונקציה (Sigmoid) במוקדים

(2) עיסת פל ימח

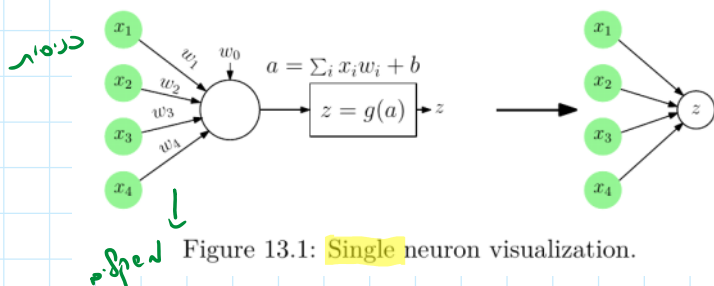
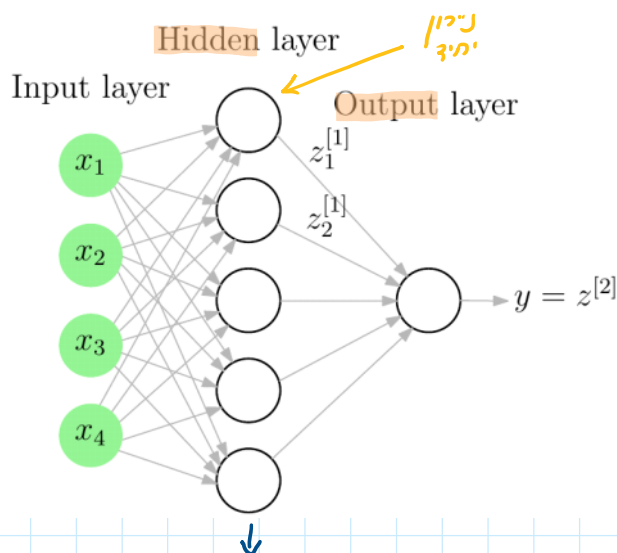


Figure 13.1: Single neuron visualization.

נצח ומה

$$a^{[1]} = \mathbf{w}^T \mathbf{x} + b$$

$$z_1^{[1]} = g(a^{[1]})$$



Multi-layer model:

$$\mathbf{a}^{[k]} = \left(\mathbf{W}^{[k]}\right)^T \mathbf{z}^{[k-1]} + \mathbf{b}^{[k]}$$

$$\mathbf{z}^{[k]} = g_k(\mathbf{a}^{[k]})$$

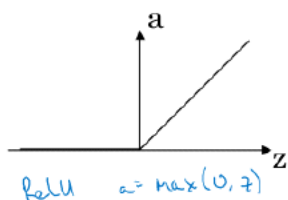
$z^{[0]} = x_i \rightarrow$ כניסה ראשונה
 תבנית ראשונה
 סכום הניבים $y = z^{[k]}$, k סדר

$$\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}}_{\mathbf{a}^{[1]}} = \underbrace{\begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} \\ w_{12} & w_{22} & w_{32} & w_{42} \\ w_{13} & w_{23} & w_{33} & w_{43} \\ w_{14} & w_{24} & w_{34} & w_{44} \\ w_{15} & w_{25} & w_{35} & w_{45} \end{bmatrix}}_{(\mathbf{W}^{[1]})^T} \underbrace{\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{bmatrix}}_{\mathbf{x}_i^T} + \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}}_{\mathbf{b}^{[1]}}$$

ቅጥር ማረጋገጫ

activation function : g

Rectified linear unit



Name	Function	Range
Sigmoid	$f(z) = \frac{1}{1 + e^{-z}}$	$[0, 1]$
Tanh	$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$[-1, 1]$

ReLU $z = \max(0, z)$

Tanh	$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$[-1, 1]$
ReLU	$f(z) = \max(0, z)$	$[0, \infty]$

Back-propagation: הפקת גרדיינטים

MSE loss

כדי לחשב את הפסד, $\hat{y} = f(\underline{x})$, מחשבים את הפסד.

* אלגוריתם gradient descent

$$\mathcal{L}(y, \hat{y}) = \frac{1}{2M} \sum_{i=1}^M (y_i - \hat{y}_i)^2$$

$$\frac{\partial}{\partial \hat{y}_i} \mathcal{L}(y, \hat{y}) = \frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i) (-1)$$

מפני שככה

$$\hat{y}_i = \sigma(a_i^{[l]})$$

$$\frac{\partial}{\partial a_i^{[l]}} \mathcal{L}(y, \hat{y}) = \frac{\partial \mathcal{L}(y, \hat{y})}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial a_i^{[l]}}$$

$$\frac{\partial}{\partial a_i^{[l]}} \hat{y}_i = a_i^{[l]} (1 - a_i^{[l]})$$

$$\frac{\partial \mathcal{L}(y, \hat{y})}{\partial w_{ij}^{[l]}} = \frac{\partial \mathcal{L}(y, \hat{y})}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial a_i^{[l]}} \frac{\partial a_i^{[l]}}{\partial w_{ij}^{[l]}}$$

דפינום: נתן פרמטר של α כדי להתחיל את ההפסד

נסיח את המידה

GD

$$x_{n+1} = x_n - \alpha f'(x_n)$$

וקטור

$$x_{n+1} = x_n - \alpha \nabla_x f'(x_n)$$

GD Variations

Batch-GD

מישור כל אחד של GD
פרמטר של Batch, אחד אחד הפסד

מחלקה: עבודה עם נתונים לא נתון ערכים
ערכיו של הפרמטר בו-באופן

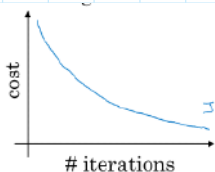
שיטה: חלוקה של נתונים לזוגות קטנים

אקראי

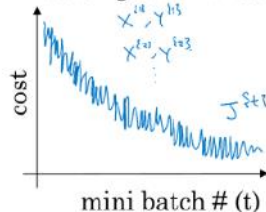
מקובל, שבדרך כלל הוא חלק של 2

Epoch: מספר צעדים של GD

GD רגיל



Mini-batch gradient descent



Batch size

↕

Batch

⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋

סעיף
מני

למנוע עזר

Stop conditions

- Pre-defined number of epoch,
- Early stop conditions on loss or metrics changes during training.

Learning rate control

- Exponential decrease as function of an epoch.
- Loss or metrics related, for example loss does not decrease for a few epochs

- Pre-defined number of epoch,
- Early stop conditions on loss or metrics changes during training.
- Exponential decrease as function of an epoch.
- Loss or metrics related, for example loss does not decrease for a few epochs
- Constant, pre-defined.