

9/1/17

נושאים מתקדמים בראייה חישובית - הרצאה #11:

נושאי השיעור:

- Perceptron
- Non Linear Trans.
- XOR Problem
- Multi layer perceptron
- Back propagation.

Perceptron

נתון $\{x_n, y_n\}_{n=1}^N$ כאשר $x_n \in \mathbb{R}^d$, $y_n \in \{-1, +1\}$.

מחרתנו: למצוא מסווג מהצורה: $h(x) = w^T x > \text{thresh}$

$$\begin{cases} h(x) = \text{sign}(w^T x) \\ x = [1, x] \end{cases}$$

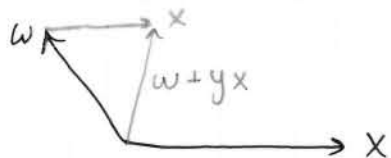
כך "נבט" מההיפוס thresh .

אלגוריתם אימון - Perceptron:

- 1) Pick misclassified point (x_n, y_n)
- 2) update $w \leftarrow w + y_n x_n$
- 3) repeat untill convergence

משפט: אם הקבוצות הן separable, אזי, האלגוריתם מתכנס במספר סופי איטרציות לבין.

הסבר: נניח $y_n = \pm 1$; $\text{sign}(w^T x_n) < 0$ (misclassified x_n)



$$w \leftarrow w - \frac{1}{2} \underbrace{(h(x) - y_n)}_{\text{error}} x_n$$

צדק אחת ארשום את כל העדכון:

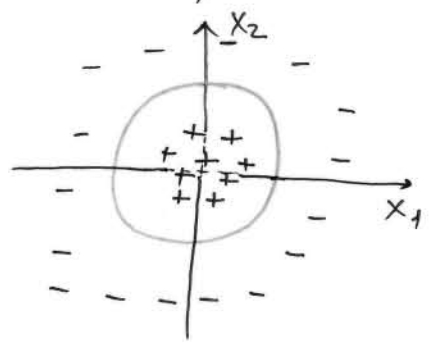
$$\text{error} = \pm 2 \quad h(x) \neq y_n \text{ plc}$$

$$\text{error} = 0 \quad h(x) = y_n \text{ plc}$$

9/1/17

Non Linear Trans.

מה נעשה כאשר הקבוצות הן בנות הפרדה אך המפריד אינו מפריד ליניארי?

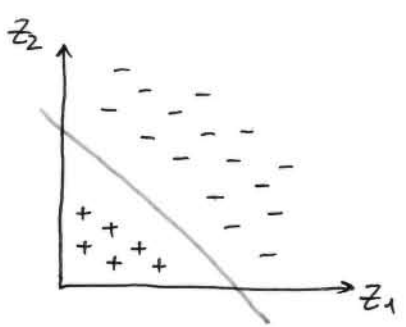


דרך אחת להתמודד עם בעיה היא ע"י תמורה לא ליניארית.

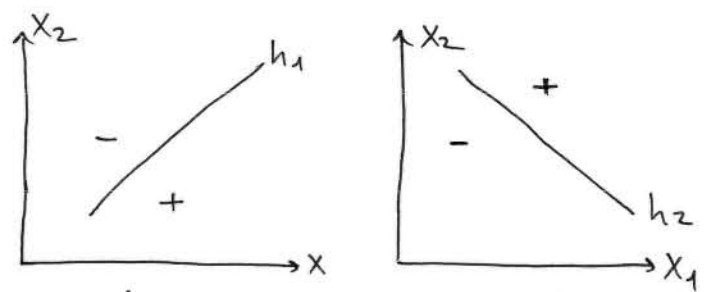
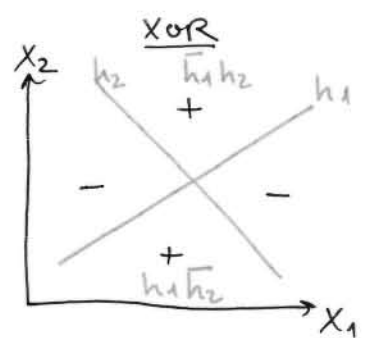
$$(x_1, x_2) \xrightarrow{\Phi} (x_1^2, x_2^2) = (z_1, z_2)$$

$$\text{sign}(w^T x) \xrightarrow[\text{תכל'ה}]{\text{במקום זה}} \text{sign}(w^T z)$$

למחרת ההערה, המילוי יראה כך:



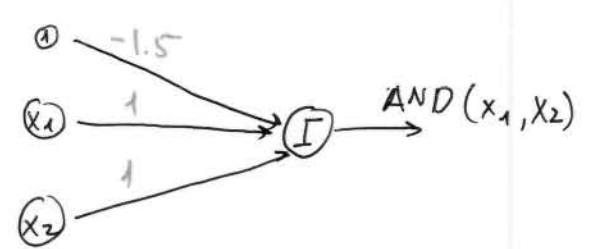
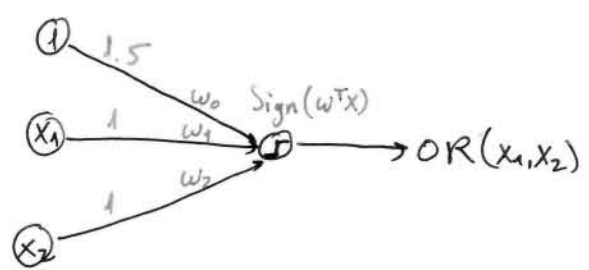
XOR Problem



Subsets

לכל נק' אופריד מצאם
Perceptron ע"י זה

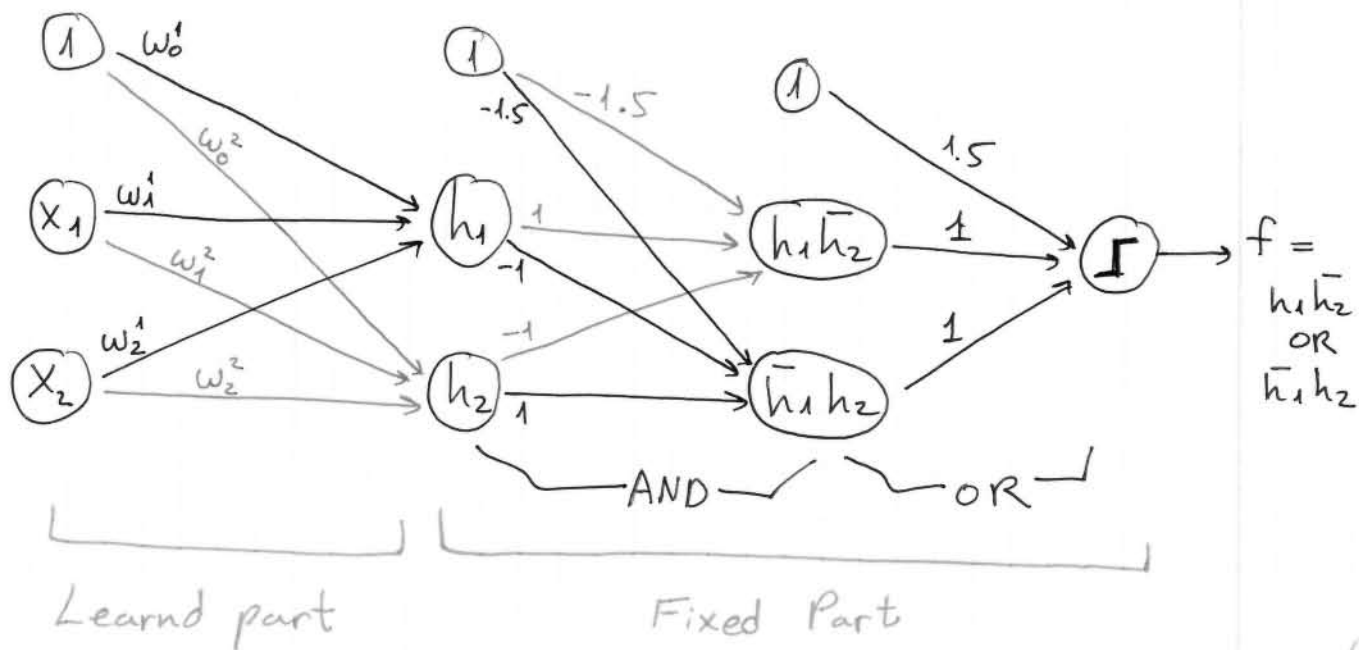
נבנה שני אוג"ים:



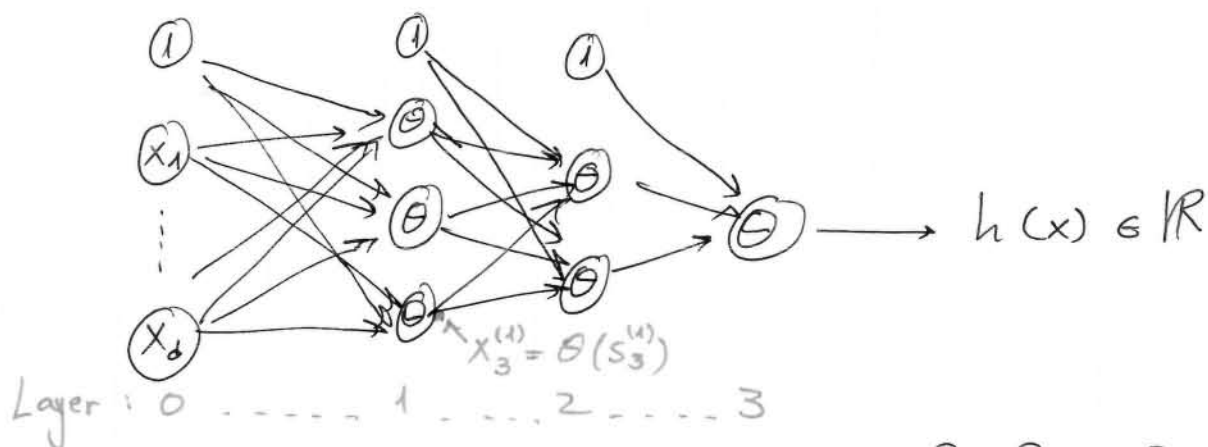
$x_1 = x_2 = -1 \Rightarrow 1.5 - 1 - 1 = -0.5$
 \Downarrow
 $OR(-1, -1) = -1$
 עבור x_1, x_2 אחרים נקבל +1

9/1/17

רשת נוירונים לא דיפרנציאלי: XOR

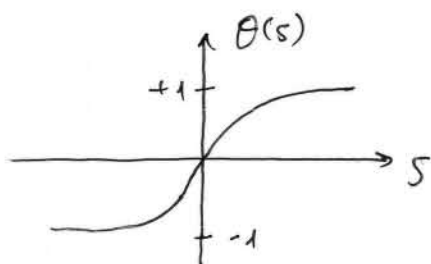


רשת נוירונים



$$s = w^T x, \quad \theta(s) = \tanh(s) \text{ רשת נוירונים}$$

$$\theta(s) = \tanh(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}$$



רשת נוירונים עם $\tanh(s)$ ו- $\text{sign}(s)$ מכיוון e - \tanh
 עוצר ואלו עוצר רשת GD

$$w_{ij}^{(l)} \begin{cases} 1 \leq l \leq L & \text{layers} \\ 0 \leq i \leq d^{(l-1)} & \text{inputs} \\ 1 \leq j \leq d^{(l)} & \text{outputs} \end{cases}$$

הרשת נקבעת ע"י המסקנות.

$$x_j^{(l)} = \theta(s_j^{(l)}) = \theta\left(\underbrace{w_{ij}^{(l)}}_{R^{d^{(l-1)}}} \cdot \underbrace{x_i^{(l-1)}}_{R^{d^{(l-1)}}}\right)$$

dot product

זכור מפורטת את זה

$$\sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} \cdot x_i^{(l-1)}$$

לגדיר פונקציה של טעות: $e(w) = (h(x_n) - y_n)^2$

נחפש: $\forall i, j, l: \nabla e(w) = \frac{\partial e(w)}{\partial w_{ij}^{(l)}}$

$$\frac{\partial e(w)}{\partial w_{ij}^{(l)}} = \frac{\partial e(w)}{\partial s_j^{(l)}} \cdot \underbrace{\frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}}}_{x_i^{(l-1)}}$$

ע"י כלל שרשרת:

$s = w^T x$: מכיוון $x_i^{(l-1)}$

(backprop ע"י שרשרת) $\frac{\partial e(w)}{\partial s_j^{(l)}} = \delta_j^{(l)}$ נשאר לשלוק את זה:

מקבלים את ה forward pass

$$\Delta w = -\eta \nabla e(w) = \frac{\partial e(w)}{\partial w_{ij}^{(l)}} = -\eta x_i^{(l-1)} \delta_j^{(l)}$$

בדיומה אפרססטרין:

העדכון יהיה: $w \leftarrow w + \Delta w$