

22-5-2025

1) Εργασίες

install. packages ("keras")

2) R package (keras)

Εξέταση Παραγόμενων σε Νευρωνικά δίκτυα

$$\hat{y} = \underbrace{f(x)}$$

$$\min_{\theta} \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$

n: μέγεθος
training set

$X = (X_1, \dots, X_p)$, Y εξαρ. μεταβ.

θ : διάνυσμα παραγόμενων.

Unconstrained Optimization

- $f(x; \theta)$
- 1) όχι κυρία ως προς θ
 - 2) $\theta \in \mathbb{R}^M$, M πολύ μεγάλο
 - 3) f πιθανώς χωρίς ελάχιστο.

Ψάχνουμε τοπικά ελάχιστα (ενα;)

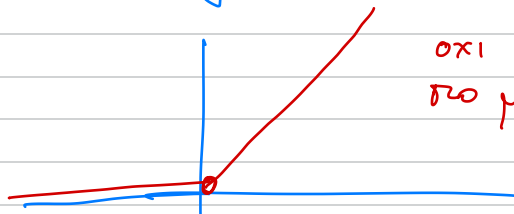
Εύρεση τοπικών ελαχίστων κ' όχι προσεγγιστικά.

Αν υποθ. f είναι παραγωγίσιμη

g : sigmoid



g : ReLU



οχι παραγωγίσιμη
στο μηδέν

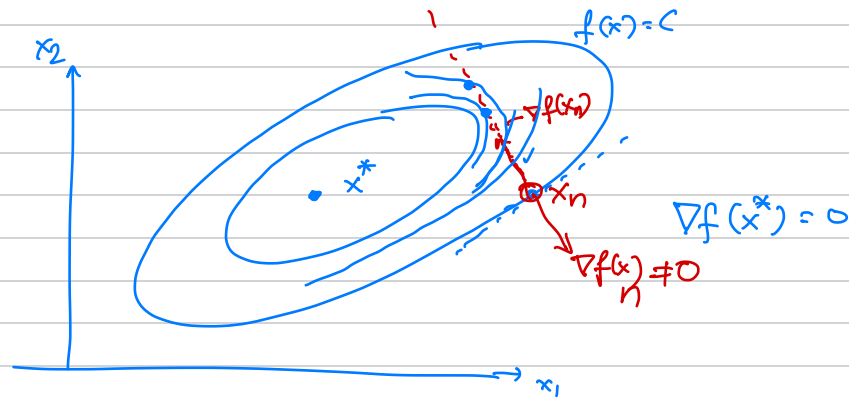
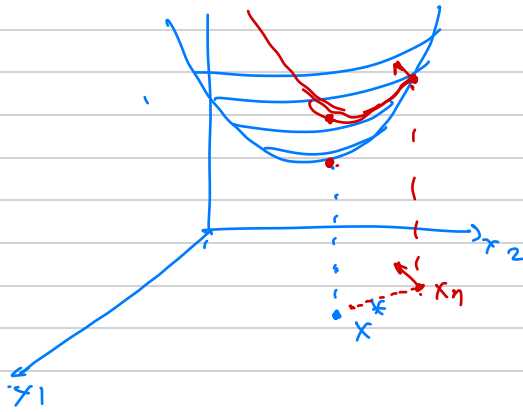
First order Conditions : $\frac{\partial}{\partial \theta_i} \sum (y_i - f(x_i; \theta))^2 = 0 \quad \forall \theta_i$ } (??)

Αλγόριθμος Βαθμίδας (Gradient-descent algorithm)

Εστω f : κυρτή

$f: \mathbb{R}^2 \rightarrow \mathbb{R}$, λείων
παράγυγιση

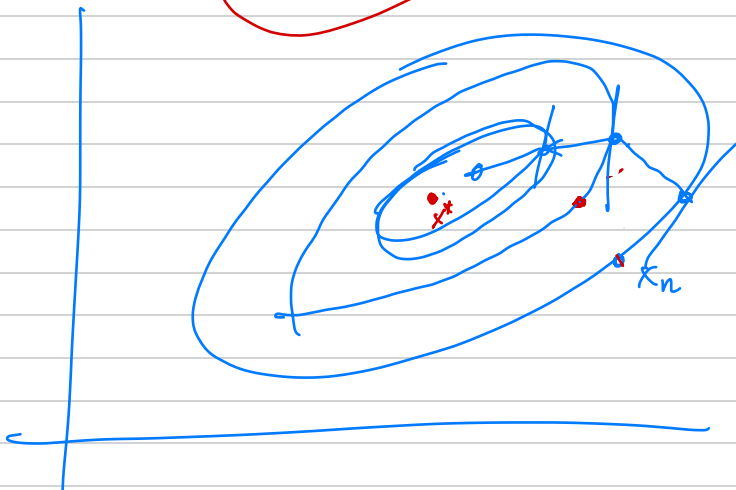
Εστω x^* το ελάχιστο ελαττωστό



Εστω ότι σε κάθε x μπορούμε να υπολογίσουμε
τα $f(x)$ & $\nabla f(x) = \begin{pmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{pmatrix}$

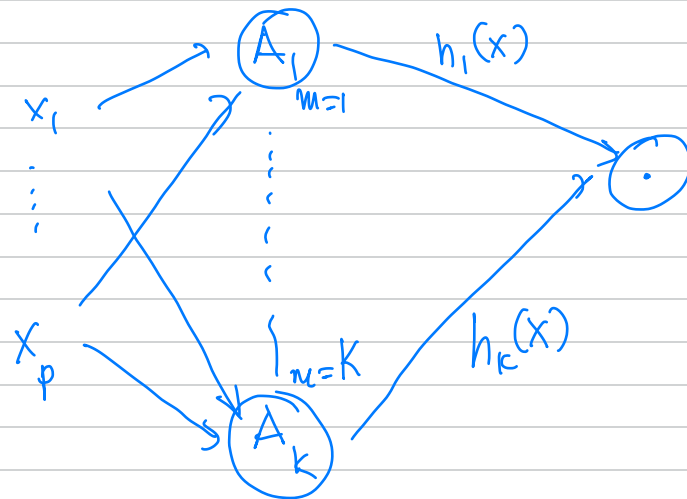
$$x_{n+1} = x_n - \delta_n \nabla f(x_n) \quad \text{εναρμονίζουμε στο } x_{n+1}$$

$\delta_n \rightarrow 0 !!$



This is analogous to gradient, say. $\frac{\partial}{\partial \theta_i} \sum (y_i - f(x_i; \theta))^2$

A.X.



$$\hat{y} = \hat{f}(x) = b_0 + \sum_{m=1}^K b_m h_m(x)$$

$$h_m(x) = g\left(w_{m0} + \sum_{j=1}^p w_{mj} x_j\right)$$

g : activation function

$$\hat{f}(x) = b_0 + \sum_{m=1}^K b_m g\left(w_{m0} + \sum_{j=1}^p w_{mj} x_j\right)$$

$$R(\theta) = \frac{1}{2} \sum_{i=1}^n \underbrace{\left(y_i - \hat{f}(x_i; \theta)\right)^2}_{R_i(\theta)}$$

$$\theta = (b_0, b_1, \dots, b_K, w_{10}, \dots, w_{1p}, w_{20}, \dots, w_{2p}, \dots, w_{K0}, \dots, w_{Kp})$$

$$\nabla R(\theta) = ?$$

$$\frac{\partial R(\theta)}{\partial \theta_i}$$

$$R(\theta) = \sum_{i=1}^n R_i(\theta)$$

$$R_i(\theta) = \frac{1}{2} \left(y_i - \underbrace{f(x_i; \theta)}_{b_0 + \sum b_m h_m} \right)^2$$

$$\frac{\partial R_i(\theta)}{\partial b_m} = - (y_i - f(x_i; \theta)) \cdot \frac{\partial f(x_i; \theta)}{\partial b_m}$$

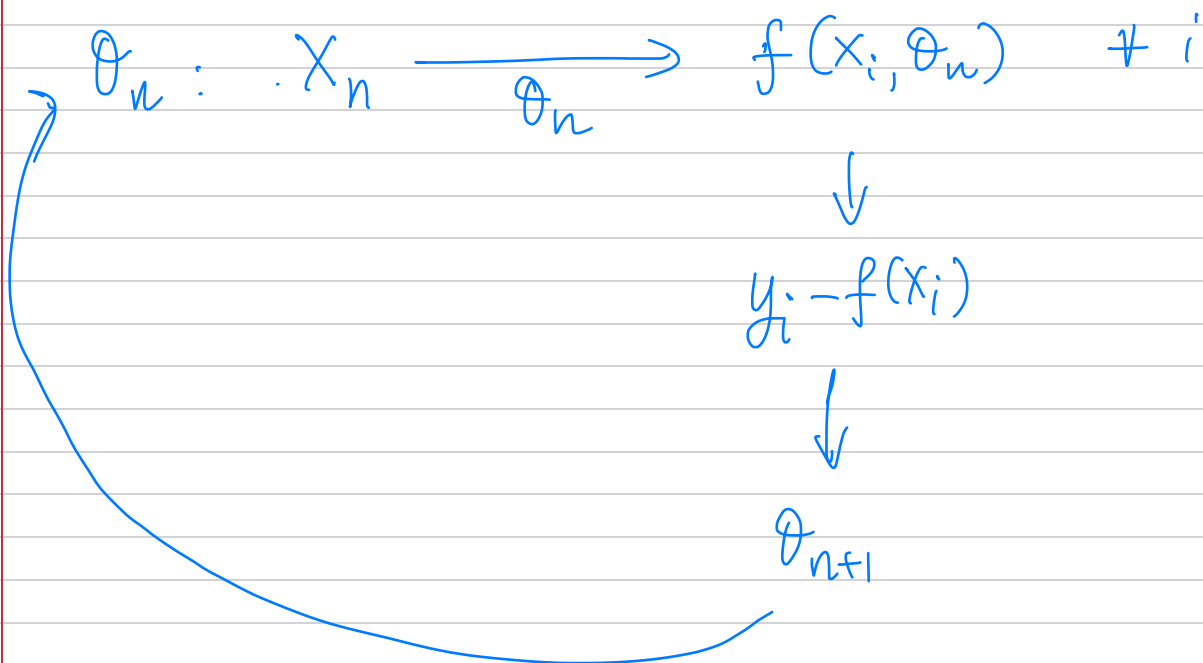
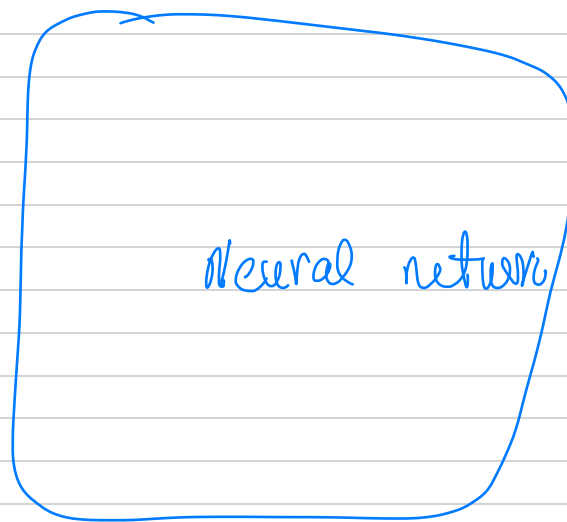
$$= - \boxed{(y_i - \hat{f}(x_i))} \cdot g(z_{im})$$

$$z_{im} = w_{m0} + \sum w_{mj} x_{ij}$$

$$\frac{\partial R_i(\theta)}{\partial w_{mj}} = \dots = - \boxed{(y_i - \hat{f}(x_i))} \cdot b_m \cdot g'(z_{im}) x_{ij}$$

σφάλμα των παρατηρήσεων i.

back propagation scheme



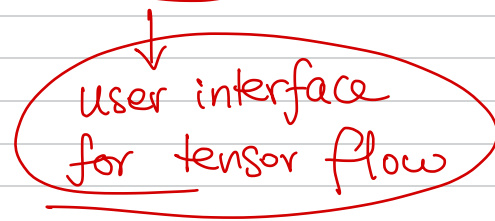
Stochastic gradient

Σε κάθε επανάληψη κάνει fit χρησιμοποιώντας υποσύνολο του training set

Dropout (ιδέες από ridge/random forests)

Σε κάθε επανάληψη παραλείπει ένα μέρος από τους κόμβους του δικτύου.

Tensorflow or R (keras)



$$f(x) = g_1(g_2(g_3(x)))$$

pipe syntax $x \rightarrow g_3 \rightarrow g_2 \rightarrow g_1$

$$x \rightarrow g_3 \Leftrightarrow g_3(x)$$

R syntax

$\%>\%$

$$x \% > \% f() \Leftrightarrow f(x)$$