Lecția 5: Regresia Liniară

living area	nr. of bedrooms	Price (1000) $
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮	⋮	⋮

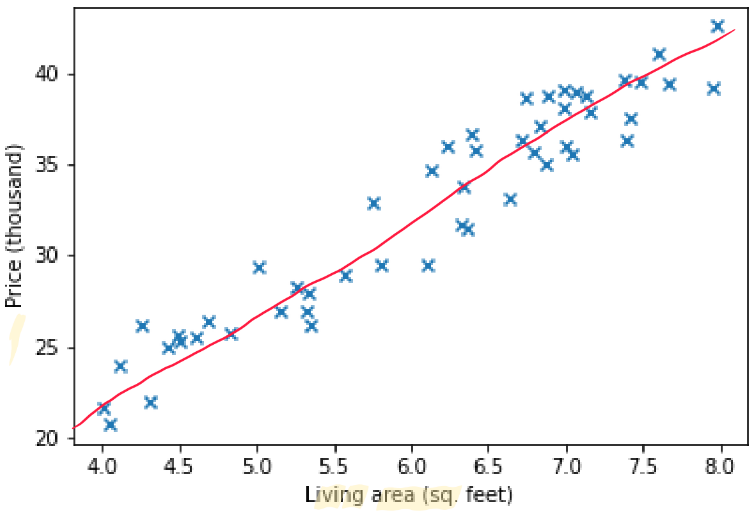
x $\qquad\qquad$ y ← target

$x \in \mathbb{R}^2$

Unde $x_1^{(i)}$ - living area

$x_2^{(i)}$ - nr. of bedrooms

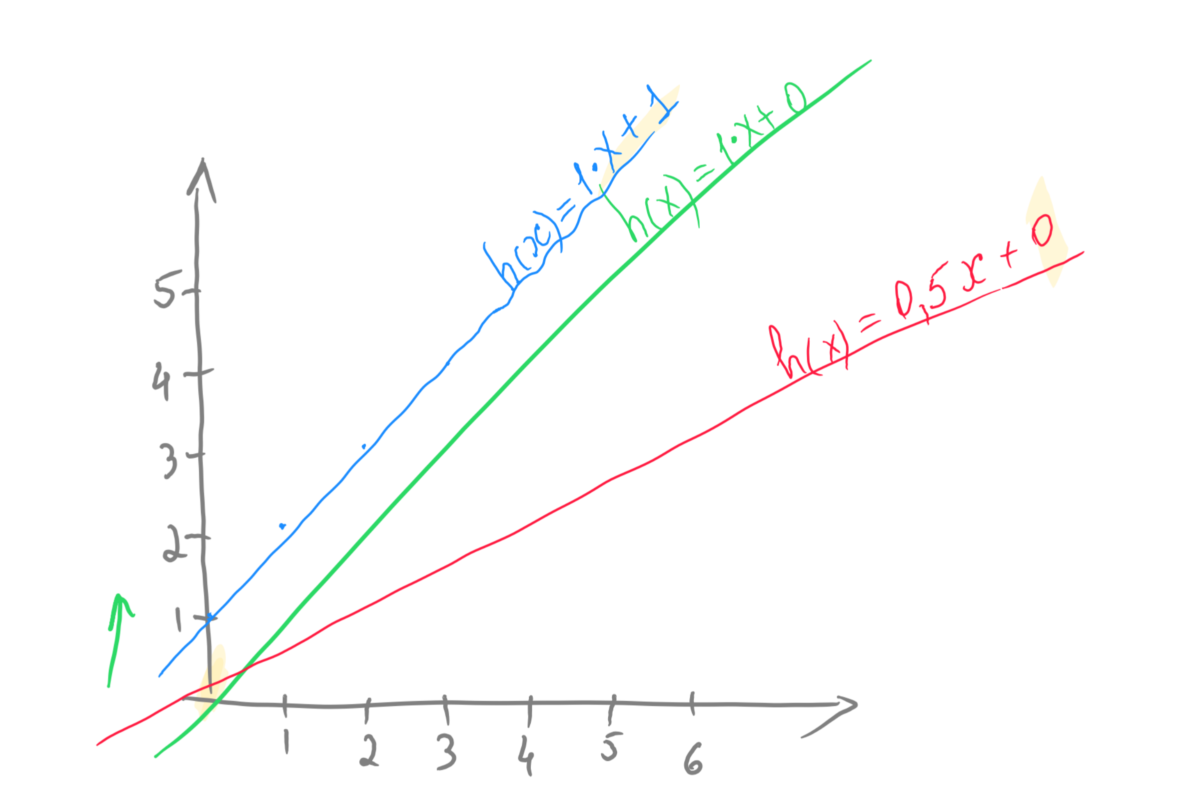
Graficul area vs price at putea arata asa:



Ipoteza: dependenta dintre living area si pret este una lineara. Dar care linie anume?

Ne amintim graficul functiei liniare.

$$h(x) = a \cdot x + b$$



Putem descrie ecuatia anterioara ca

$$h(x) = ax + b \Rightarrow a = \theta_1, b = \theta_0$$

$$h_\theta(x) = \theta_1 x + \theta_0 \cdot 1 = \sum_{i=0}^{m} \theta_i x_i,\ x_0 = 1$$

Unde formula sumei este data de:

$$a_0 + a_1 + a_2 + \ldots + a_n = \sum_{i=0}^{n} a_i$$

Putem rescrie formula ipotezei in format vectorial ca:

$$\sum_{i=0}^{N} \theta_i x_i = \theta^T x \Rightarrow h_\theta(x) = \theta^T x$$

-(-(-?))

Avand:

Exemplu:

$$\theta = \begin{array}{cc} \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \begin{array}{l} \to \theta_0 \\ \to \theta_1 \end{array} \end{array}, \quad x = \begin{array}{cc} \begin{bmatrix} 1 \\ x_1 \end{bmatrix} & \begin{array}{l} \to x_0 \\ \to x_1 \end{array} \end{array}$$

$$\Rightarrow \theta^T x = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$= 1 \cdot 1 + 2 \cdot x$$

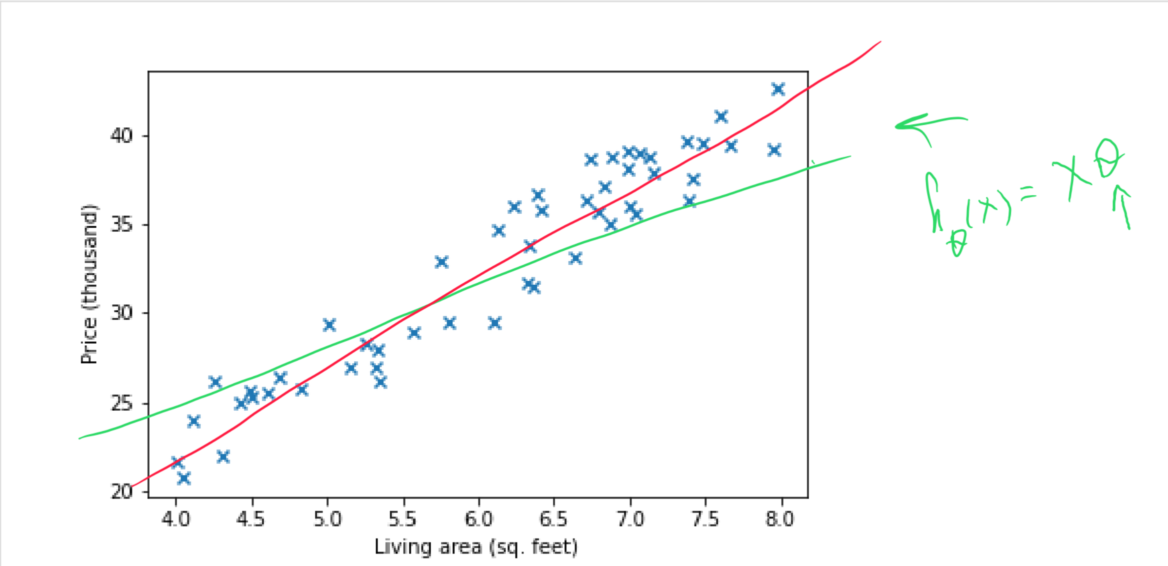
$$= 1 + 2x$$

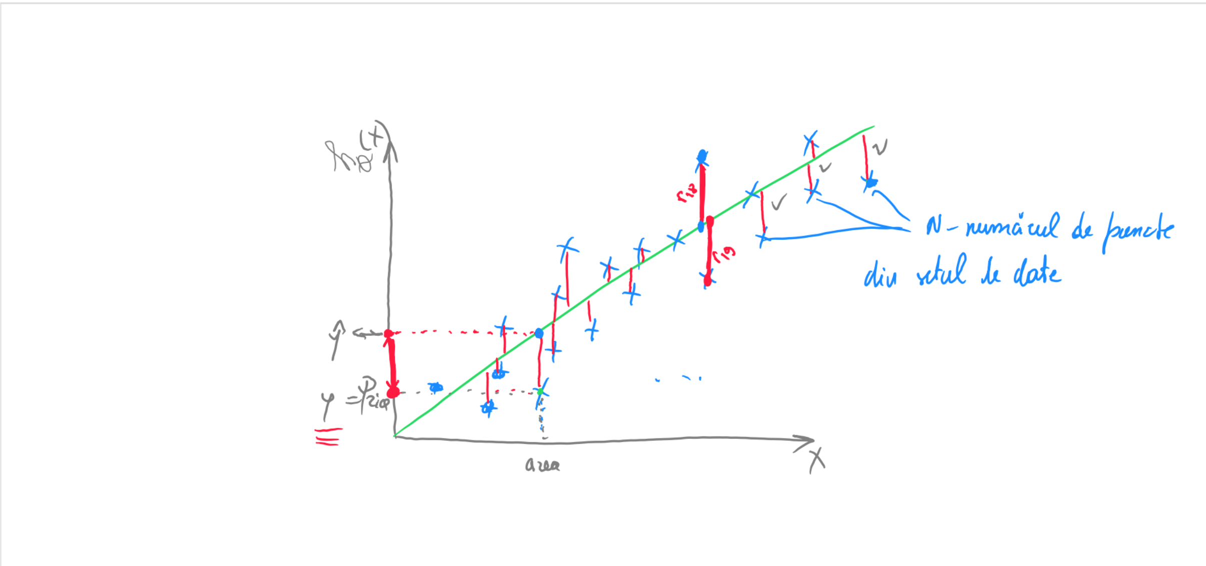
$$= 2x + 1$$

Considerand o matrice de n observatii X:

$$h_\theta(X) = X\theta$$

Cum selectam linia cea mai optimala?





Linia cea mai optimala este cea pentru care in medie eroarea este minima.

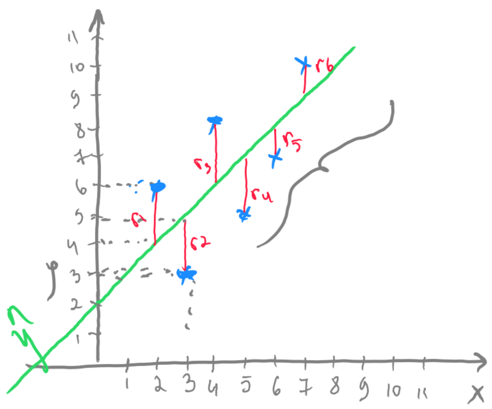
$$R = \frac{r_1 + r_2 + \ldots + r_n}{n} = \frac{1}{N}\sum_{i=1}^{N} r_i$$

$$r = y - h_\theta(x) \Rightarrow ?$$

$$R = \frac{1}{N}\sum \left(y - h_\theta(x)\right);\ h_\theta(x) = \hat{y} \Rightarrow ?$$

$$R = \frac{1}{N}\sum_{i=1}^{N} (y - \hat{y})$$

Care ar fi problema daca adunam erorile asa? Ar putea cumva a rezultatul sa fie zero chiar daca exista erori? Raspunsul este da. Sa ne uitam in graficul de mai jos:



$r = y - \hat{y}$

$r_1 = 6 - 4 = 2$

$r_2 = 3 - 5 = -2$

$r_3 = 8 - 6 = 2$

$r_4 = 5 - 7 = -2$

$r_5 = 7 - 8 = -1$

$r_6 = 10 - 9 = 1$

$$R = \frac{1}{6}\sum_{i=1}^{N=6} r_i = \frac{1}{6}\left(2 + (-2) + 2 + (-2) + (-1) + 1\right) = 0.$$

Însă, noi vedem clar pe grafic că prezicerile noastre nu coincid de fapt cu valorile reale ale lui y. => Înloc să adunăm valorile erorilor -> adunăm valorile erorilor ridicate la pătrat la pătrat =>

Prin urmare, formula pentru suma erorilor devine:

$$R = \frac{1}{N}\sum_{i=1}^{N}\left(y - \hat{y}\right)^2 = \frac{1}{N}\left\|y - \hat{y}\right\|_2^2$$

Ne amintim ca:

$$\hat{y} = X\theta$$

Prin urmare:

$$R = \frac{1}{N}\left\|y - \hat{y}\right\|_2^2 = \frac{1}{N}\left\|y - X\theta\right\|_2^2$$

Care poate fi rescris ca:

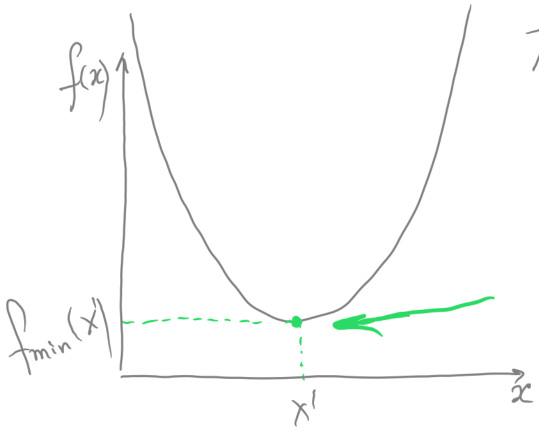
$$R = \frac{1}{N}\left\|y - X\theta\right\|_2^2 = \frac{1}{N}\cdot\left(y - X\theta\right)^T\cdot\left(y - X\theta\right)$$

Functia R este adesea notata cu L, din engleza Loss Function.

Ne amintim ca scopul nostru e de a prezice targetul cu cea mai mica eroare. Acum, avand o functie cum aflam valoarea minima a functiei? Sa atragem atentia unui exemplu:

Care este valoarea minimă pt funcția

$$f = x^2 + 2x + 1$$



Analitic, minimul unei funcții poate fi calculat cu ajutorul derivatei:

$$f'(x) \overset{!}{=} 0 \Rightarrow$$

$$(x^2 + 2x + 1)' = 0$$

$$2 \cdot x + 2 + 0 = 0$$

$$2(x+1) = 0 \quad \Big| \cdot \frac{1}{2}$$

$$x + 1 \overset{!}{=} 0$$

$$x = -1$$

Valoarea funcției pentru $x = -1$ este:

$$f(-1) = (-1)^2 + 2(-1) + 1 = 1 - 2 + 1 = 0$$

Calculam minimumul functiei noastre - afland astfel pentru ce valoare a parametrului theta eroarea este minima.

Ne reamintim urmatoarele proprietati cand lucram cu derivate:

$$\frac{d}{dx} \times \left[f\left(g(x)\right)\right] = f'\left(g(x)\right) \cdot g'(x)$$

$$\left(f(x) \cdot g(x)\right)' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Pentru functia noastra:

$$\frac{\partial L}{\partial \theta} = \frac{\partial}{\partial \theta} \times (y - X\theta)^T \cdot (y - X\theta)$$

$$? = (y - X\theta)^1 \cdot (y - X\theta) + (y - X\theta) \cdot (y - X\theta)^1$$

$$? = -X(y - X\theta) + (y - X\theta) \cdot (0 - X)$$

$$? = -X^T y + X^T X\theta - X^T y + X^T X\theta$$

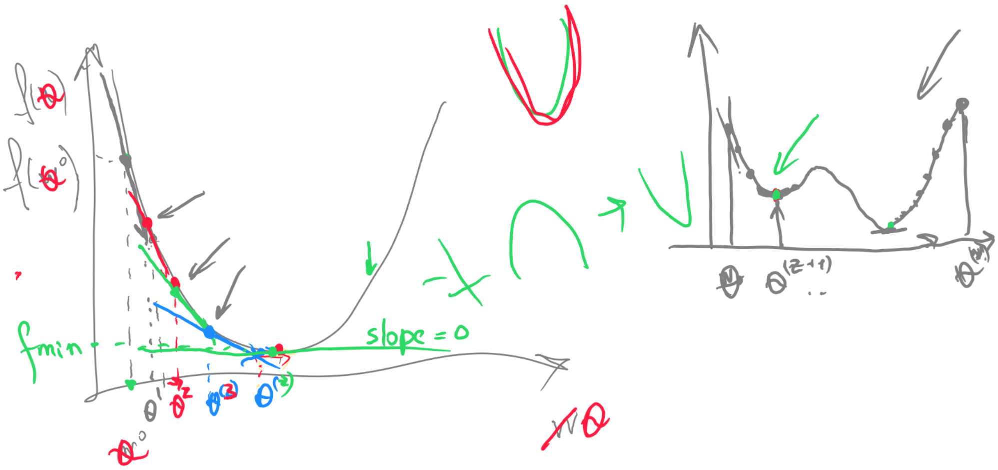
$$0 = 2X^T X\theta - 2X^T y = 0$$

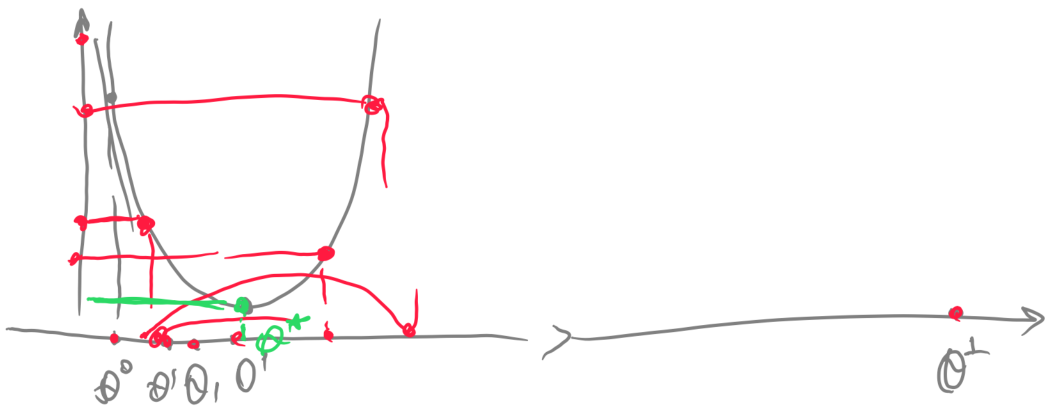
$$2X^T X\theta = 2X^T y$$

$$X^T X\theta = X^T y$$

$$\theta = \left(X^T X\right)^{-1} \cdot X^T y$$

Gradient Descent Intuition





Pentru loss-funcția noastră:

$$\frac{\partial L}{\partial \theta} = \underline{2X^TX\theta - 2X^Ty} = \underline{-2X^T(y - X\theta)}$$

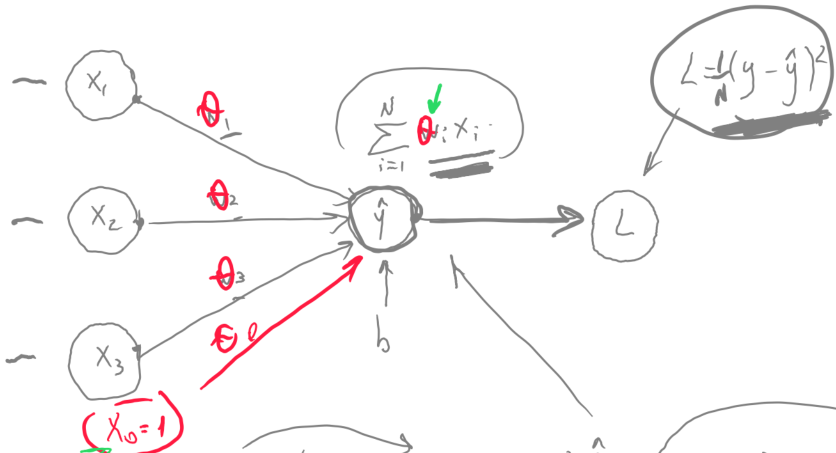
Formula pentru pasul de update devine:

$$\theta^{(t)} = \theta^{(t-1)} - \alpha \cdot \frac{\partial L}{\partial \theta}$$

substituim:

hiperparametru

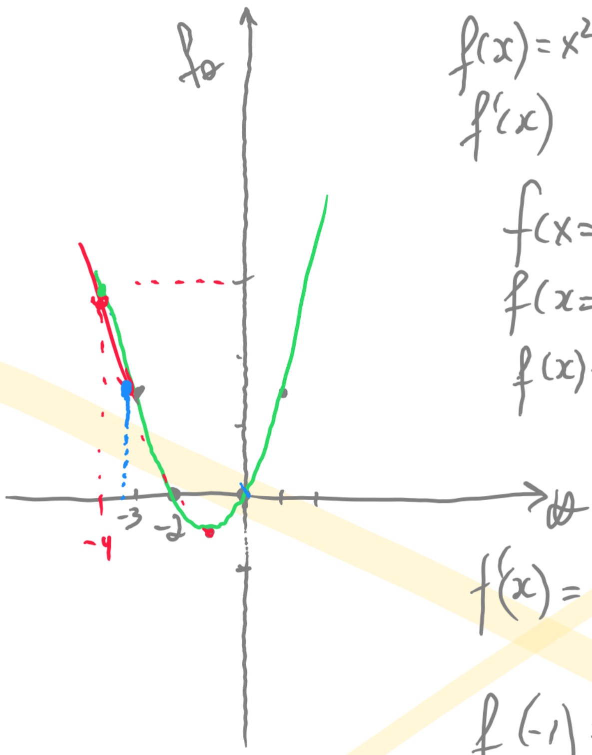
$\hat{y} = h_\theta(X)$



$$\frac{\partial L}{\partial \theta_i} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta_i} = 2(y-\hat{y}) \cdot \frac{\partial \hat{y}}{\partial \theta_i} = \underbrace{2(y-\hat{y})}_{\text{Loss gradien}} \cdot \underbrace{x_i}_{\text{Linear operation gradient}}$$

$$\sum_{i=0}^{N} \theta_i x_i = \overbrace{\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3}^{\alpha} = x_1$$

$\partial \theta_1$



$f(x) = x^2 + 2x$

$f'(x)$

$f(x=1) = 1^2 + 2\cdot 1 = 3$

$f(x=-1) = (-1)^2 - 2 = -1$

$f(x) =$

$f'(x) = 0 \Rightarrow 2x + 2 = 0$

$x + 1 = 0$

$x = -2$

$f(-1) = -1$

$f(-3) = 9 - 6 = +3$

$f(1) = 1^2 + 2\cdot 1 = 3$

Gradient descent:

$f(-4) = (-4)^2 + (-4)\cdot 2 = 16 - 8 = 8.$

Let $\alpha = 0.1 \Rightarrow$

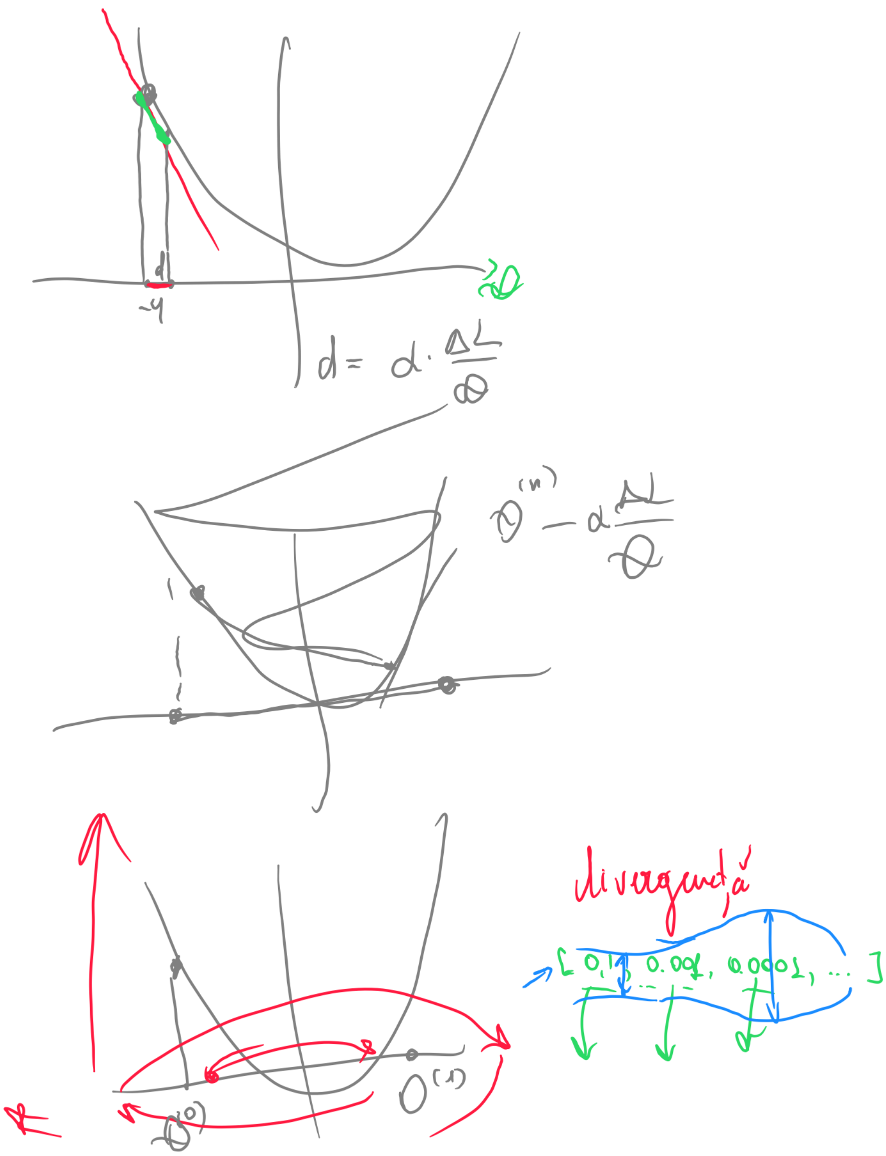
$\theta^{(1)} = \theta^{0} - \alpha \cdot \frac{\nabla L}{\theta} \Rightarrow 4 - 0.1(2\cdot 4 - 2) =$

$= 4 - 0.1\cdot 6 = 4 - 0.6 = 3.4$

$\theta^{(2)} = \theta^{(1)} - \alpha \frac{\Delta L}{\theta} = 3.4 - 0.1\cdot(2\cdot 3.4 - 2) =$

$= 3.4 - 0.48 = 2.92$

$\vdots$


θ
d
d = α · ΔL/Δθ
θ^(n) − α ΔL/Δθ
divergentă
[0.1, 0.01, 0.0001, ...]
θ^(0)
θ^(1)