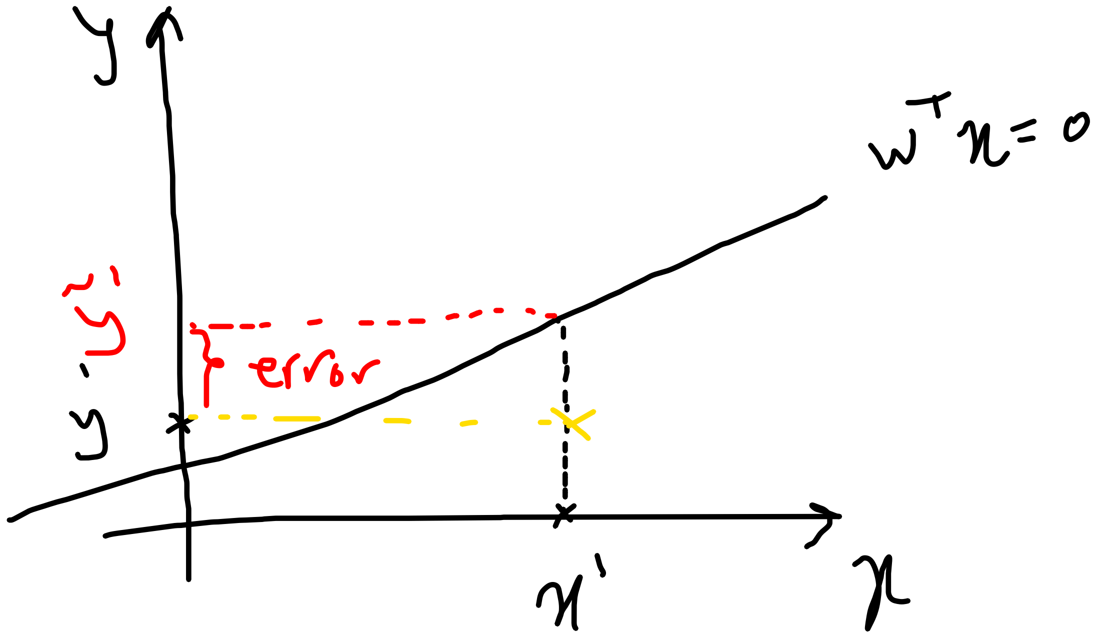


$$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$\langle x^1, y^1 \rangle$$
$$\langle x^2, y^2 \rangle$$
$$\vdots$$
$$\langle x^n, y^n \rangle$$

$n = 1000$

$$J(w) =$$

$$e_1 = |y^1 - \tilde{y}^1|$$

$$\boxed{J(w) = \sum_{i=1}^{n} |y^i - \tilde{y}^i|}$$

$$w^* \Rightarrow \frac{\partial J(w)}{\partial w} = 0$$

$$w^* = ?$$

$$x^i \longrightarrow \tilde{y}^i$$

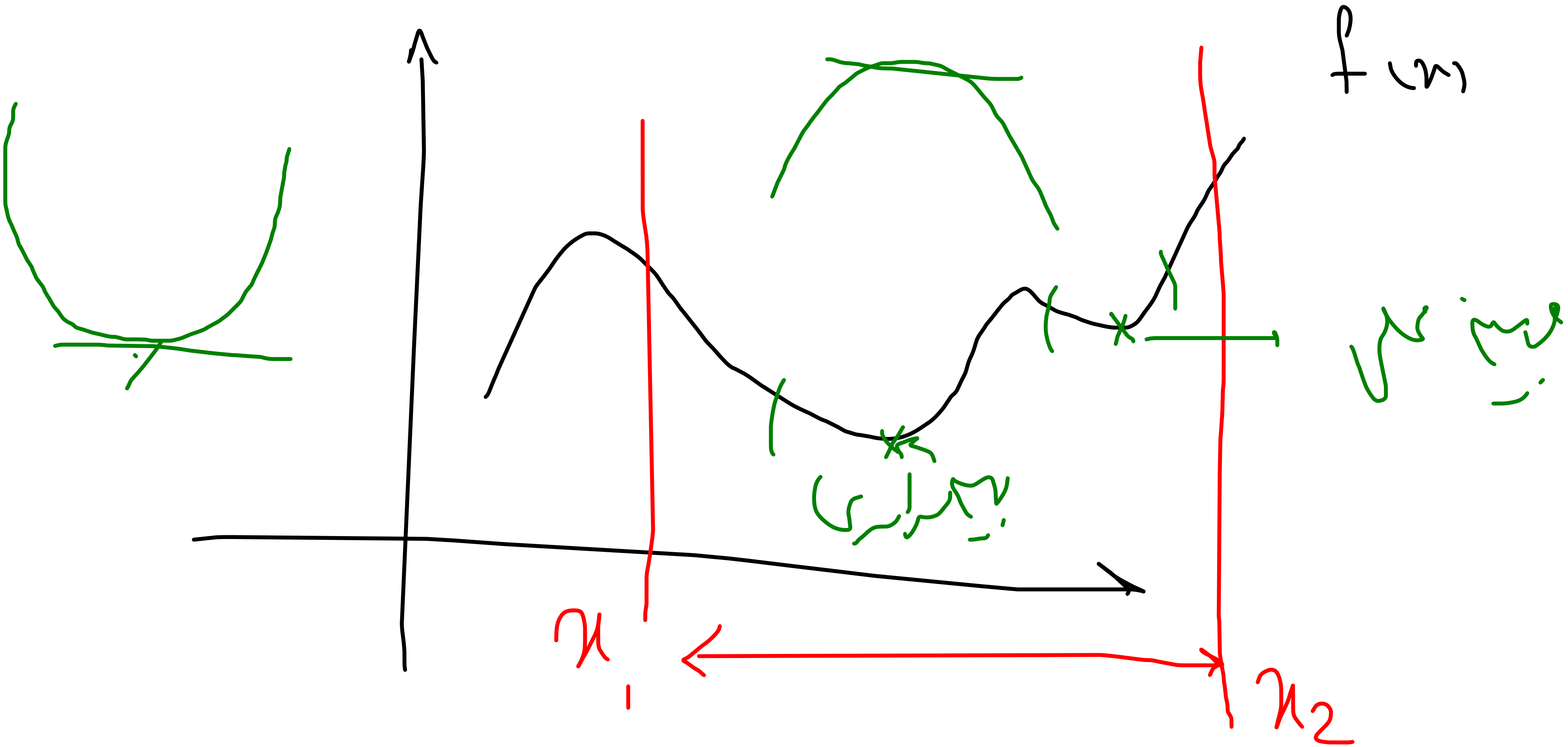
① تابع اگه $J(w)$ convex باشد تا بهینه سراسری داشته باشد

محدب

$$\boxed{\frac{\partial J(w)}{\partial w} = 0}$$

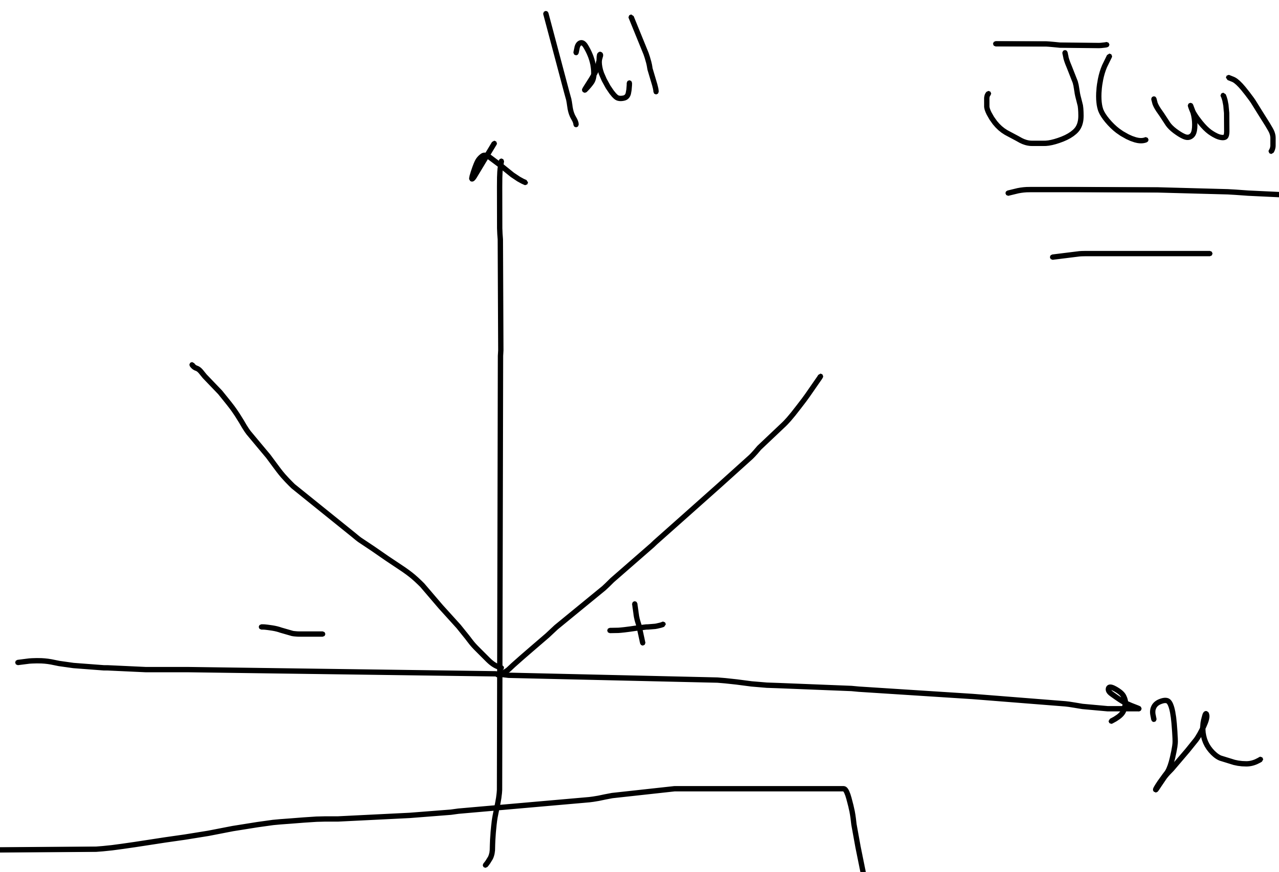
Concave

Convex

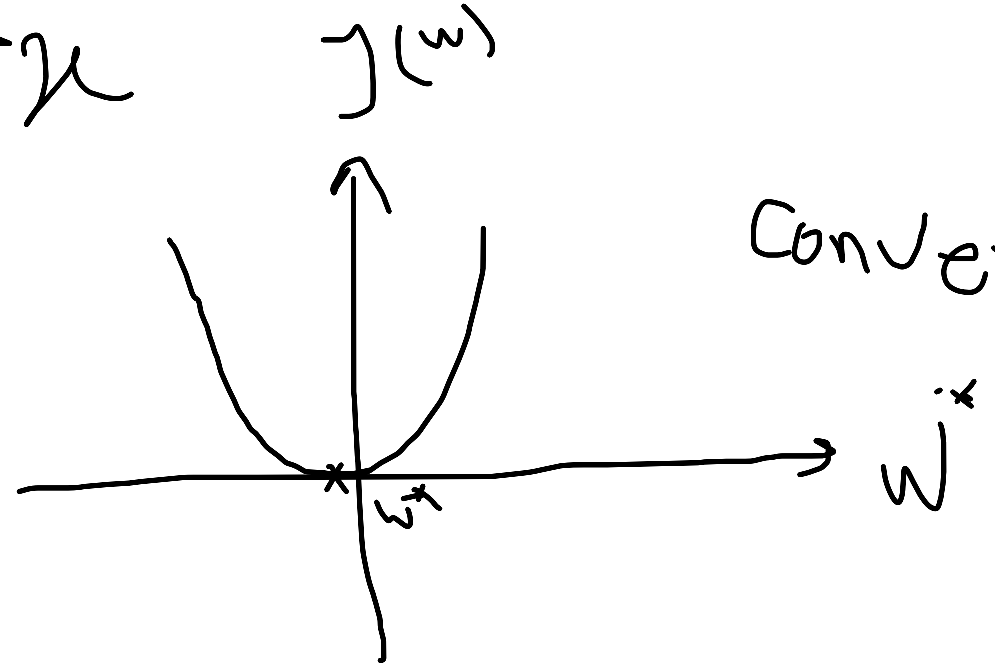


$$J(w) = \sum_{i=1}^{n} \text{error}_i = \sum_{i=1}^{n} (y^i - \tilde{y}^i)$$

$$= \sum_{i=1}^{n} |y^i - \tilde{y}^i|$$



$$\boxed{J(w) = \sum_{i=1}^{n} (y^i - \tilde{y}^i)^2}$$



Convex

$$\underline{\underline{J(w) = \sum_{i=1}^{n} (y^i - w^T x^i)^2}}$$

$$\Updownarrow$$

$$J(\vec{w}) = \|\vec{y} - w^T \vec{x}\|^2 = (y - w^T x)^T (y - w^T x)$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$

$$J(w) = (y^T - x^T w)(y - w^T x)$$

$$= y^T y - y^T w^T x - x^T w y + x^T w w^T x$$

$$\boxed{J(w) = y^T y - 2 x^T w y + x^T w w^T x}$$

کمینه خطا (w بهینه)

$$J(w) = y^T y - 2x^T w y + \underbrace{x^T w}\, w^T x$$

$$\frac{\partial J(w)}{\partial w} = 0 \Rightarrow -2x^T y + x^T w^T x + x^T w x = 0$$

$$-\cancel{2}x^T y + \cancel{2}\, x^T w x = 0 \Rightarrow x^T w x = x^T y$$

$$X \in \mathbb{R}^{n\times d}$$

$$y \in \mathbb{R}^{n}$$

$$\Rightarrow x^T x w = x^T y$$

$$\Longrightarrow \boxed{\underset{d\times 1}{w^*} = \underset{d\times d}{(x^T x)^{-1}}\, \underset{d\times n}{x^T}\, \underset{n\times 1}{y}} \rightarrow d\times n \rightarrow d\times 1$$

$X = $

$$(x^T x)^{-1} \Rightarrow d \times d$$

$d\times n \;\; n\times d$

حداقل سازی مربع خطاها $\Longleftarrow$ $\min_{w}$

$$J(w) = \frac{1}{N}\|y - \hat{y}\|^2$$

خطا

Mean Square Error

$\boxed{MSE}$

$X \in \mathbb{R}^{n \times d}$ $n \gg d$

$n = 60000$

$d = 786$

$$\frac{\partial J(w)}{\partial w} = 0$$

$$\boxed{w^* = (X^T X)^{-1} X^T y}$$

$\Longrightarrow$ بسیار زیاد

معکوس پذیر باشد

MNIST

نمونه 60,000

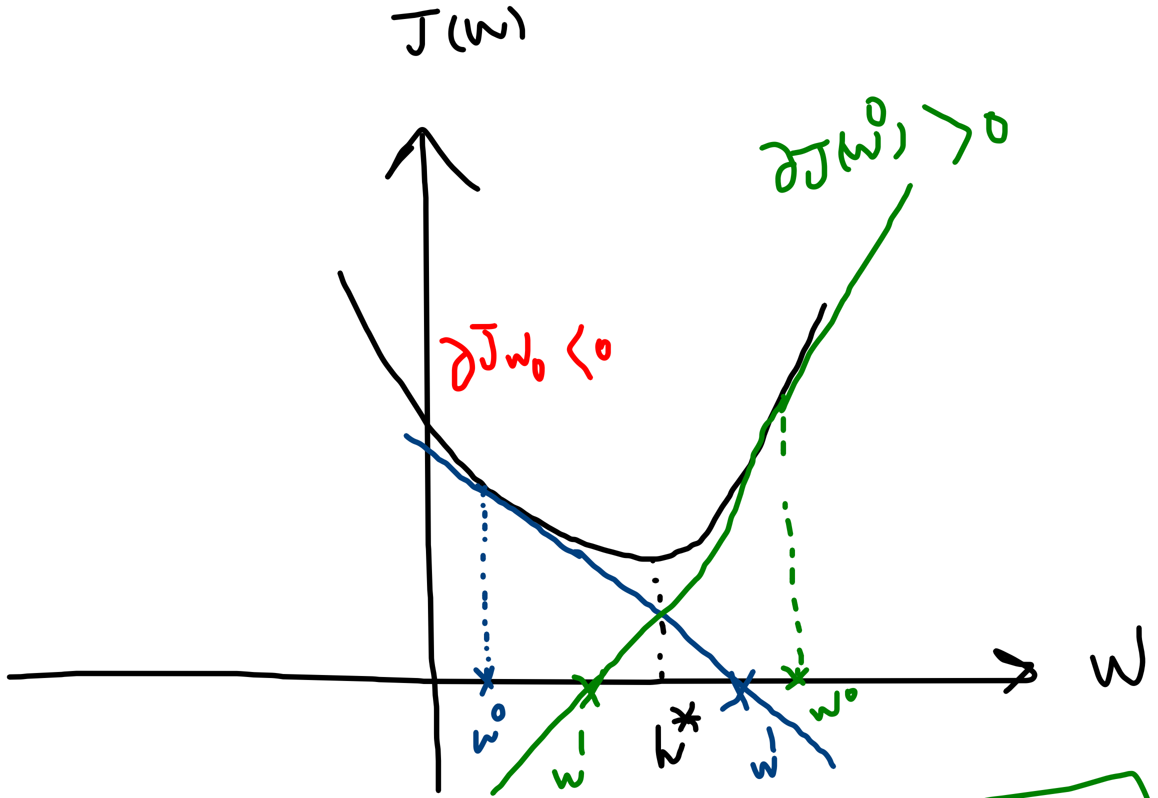
تصویر دست‌نویس از اعداد ۰-۹

25×25

روش‌های تقریبی

$$\boxed{w^* = (X^TX)^{-1}X^Ty}$$

$$\boxed{\tilde{w} \simeq w^*}$$



$w^0 =$ تصادفی

$w^1 =$ حل مقطع موقت

$\partial J(w^0) \gtrless 0$

$w^{i+1} = w^i \pm (\text{علامت شیب}) \times \alpha$

اگر $0 < \partial J(w^i)$ $\quad w^{i+1} < w^i$

اگر $0 > \partial J(w^i)$ $\quad w^{i+1} > w^i$

$$\begin{cases} w^{i+1} = w^i + \alpha & \partial J < 0 \\ w^{i+1} = w^i - \alpha & \partial J > 0 \end{cases}$$

$\Downarrow$

$$\boxed{w^{i+1} = w^i - \partial J * \alpha}$$

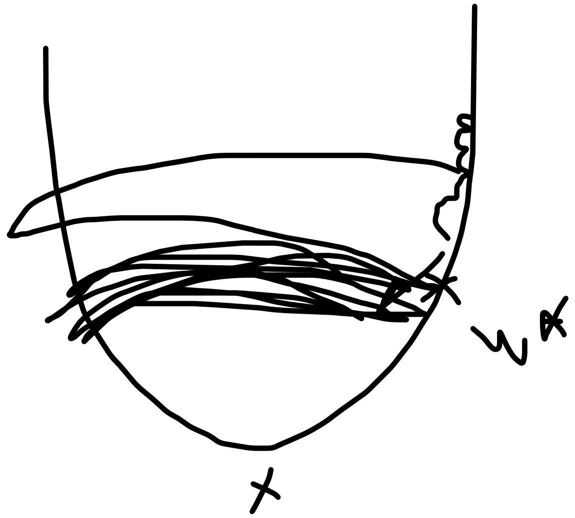
$w^0 =$ تصادفی

for $i=0$ to n:

$$\underline{w^{i+1}} = w^i - \partial J(w^i) * \underline{\underline{\alpha}}$$

$i = i+1$

اندازه گام

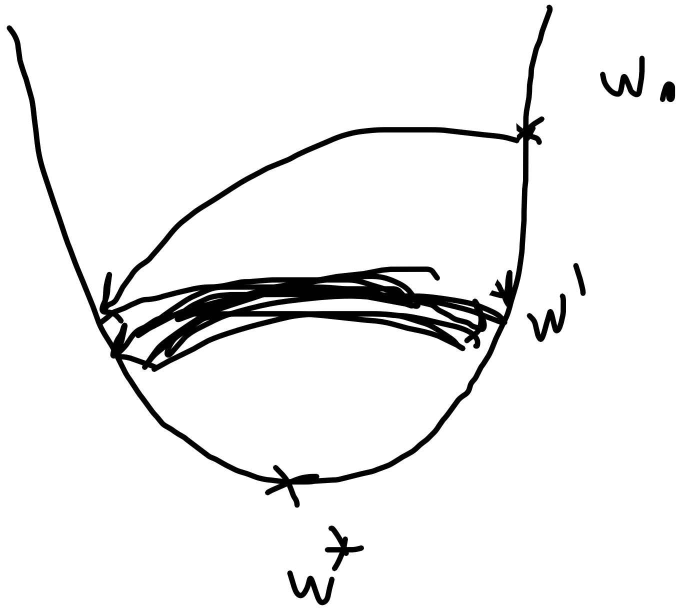


Gradient Descent (GD)

مشتق نزولی

$\alpha \swarrow \quad n \nearrow$

$n=100$



① گام بزرگ (نوسان زیاد)

تا محل w بهینه

کلی با سرعت کم

G.D.

$\tilde{w}$: $\begin{cases} w^0 = \text{تصادفی} \\ w^{i+1} = w^i - \alpha \dfrac{\partial J(w^i)}{\partial w} \end{cases}$ n

$|w^i - w^{i-1}| < eps$

$X = \begin{bmatrix} - \\ - \\ - \\ - \end{bmatrix}$

$w^* = (X^T X)^{-1} X^T y$

$\boxed{\tilde{w} \Rightarrow \text{G.D.} \quad \alpha, n}$