REGULARIZACIÓN L2:

$$J_{L^2} = J + \lambda \sum_w w^2$$

$$\frac{\partial J_{L_2}}{\partial w} = \frac{\partial J}{\partial w} + 2\lambda w$$

$\frac{\partial J}{\partial w}$: GRADIENTE DEL LOSS

$2\lambda w$: NUEVO

$$w \leftarrow w - \eta \cdot \frac{\partial J_{L_2}}{\partial w}$$

$$w \leftarrow w - \eta \frac{\partial J}{\partial w} - 2\eta\lambda w$$

$$w \leftarrow w(1-2\eta\lambda) - \eta \cdot \frac{\partial J}{\partial w}$$

$0 < 1-2\eta\lambda < 1$

WEIGHT DECAY

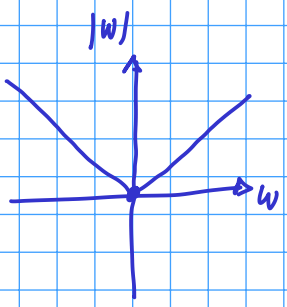
REGULARIZACIÓN L1:

$$J_{L^1} = J + \lambda \sum_w |w|$$

$$\frac{\partial J_{L_1}}{\partial w} = \frac{\partial J}{\partial w} + \lambda \frac{\partial |w|}{\partial w}$$

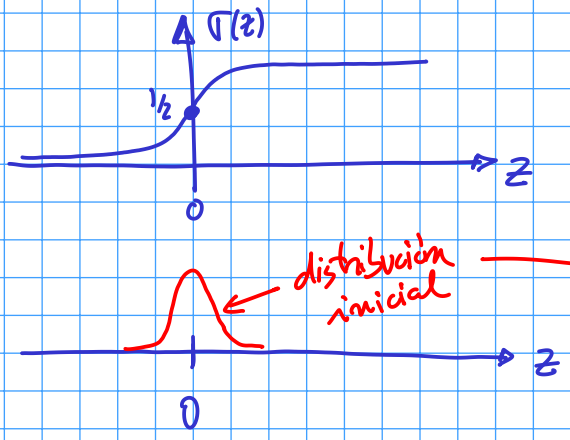
$$\frac{\partial J_{L_1}}{\partial w} = \frac{\partial J}{\partial w} + \lambda \, sgn(w)$$

$$|w| = \begin{cases} w & \text{si } w \geq 0 \\ -w & \text{si } w < 0 \end{cases}$$



$$\frac{\partial |w|}{\partial w} = \begin{cases} 1 & \text{si } w \geq 0 \\ -1 & \text{si } w < 0 \end{cases} = sgn(w)$$

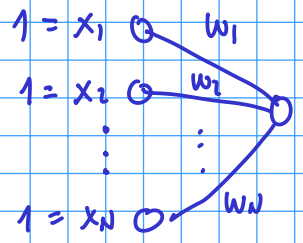
INICIALIZACIÓN DE PESOS:



¿Cómo elijo σ para que $z \sim N(0,1)$?

$z \sim N(0,1)$ $\quad$ $w \sim N(0, \sigma)$

$$\sigma = \frac{1}{\sqrt{N}}$$

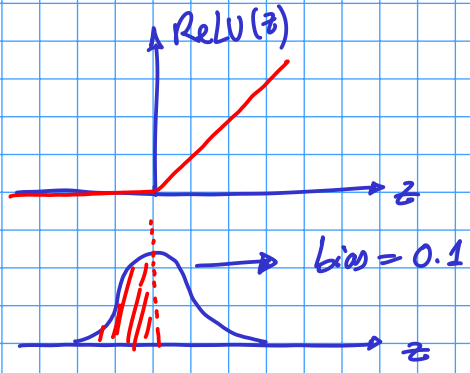


$$z = x_1 w_1 + x_2 w_2 + \cdots + x_N w_N = w_1 + w_2 + \cdots + w_N \sim N(0, \sqrt{N}\,\sigma)$$

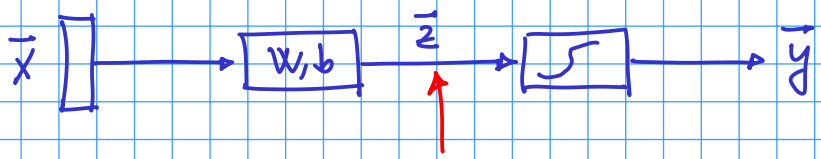
$x_i = 1$

$1 = 1^2 + 1^2 + \cdots + 1^2$ (N veces)

INICIALIZAR BIAS EN RELU:

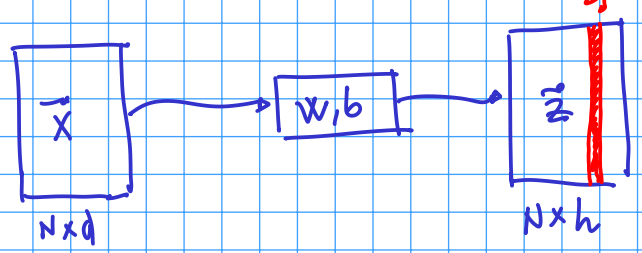


BATCH NORMALIZATION:



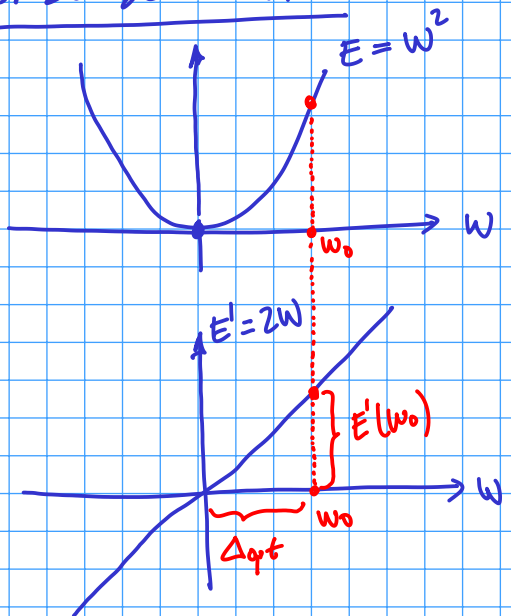
NORMALIZAR PARA QUE $z_j \sim N(0,1)$

batch size = N



$$z_{Norm} = \frac{z - np.mean(z, axis=0)}{np.std(z, axis=0)}$$

MÉTODO DE NEWTON:



¿PASO ÓPTIMO? Δ_{opt}

$$E''(w_0) = \frac{E'(w_0)}{\Delta_{opt}} \Rightarrow \Delta_{opt} = \frac{E'(w_0)}{E''(w_0)}$$

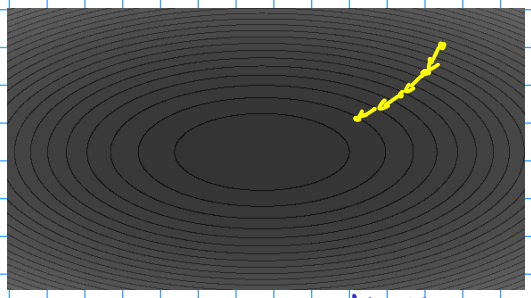
$$\boxed{w = w_0 - \frac{E'(w_0)}{E''(w_0)}}$$ PROBLEMA CUADRÁTICO

$$\boxed{w = w_0 - \eta \cdot \frac{E'(w_0)}{E''(w_0)}}$$ MÉTODO NEWTON

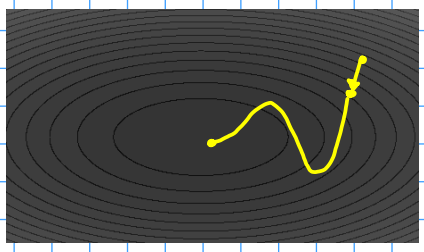
SGD CON MOMENTO:

COEFICIENTE DE MOMENTO $\mu < 1$

$$\mathbf{v} \leftarrow \mu\mathbf{v} - \eta\nabla J$$
$$\mathbf{w} \leftarrow \mathbf{w} + \mathbf{v}$$



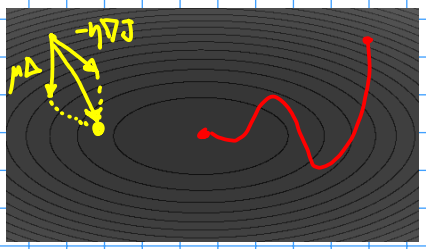
SIN MOMENTO



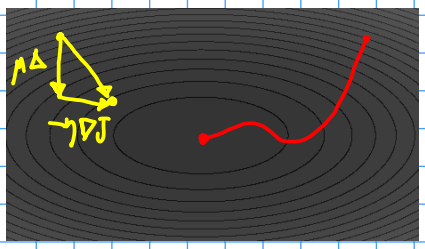
CON MOMENTO

$$w_t = w_{t-1} + \overbrace{v_t}^{\Delta t}$$
$$= w_{t-1} + \mu \underbrace{v_{t-1}}_{\Delta t-1} - \eta \nabla J$$

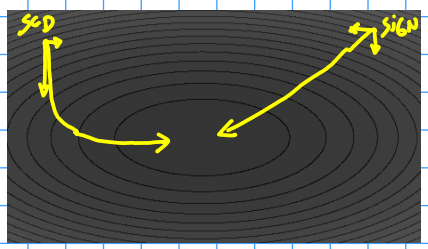
MOMENTO



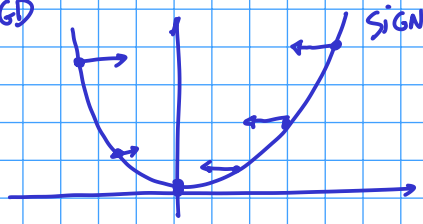
NESTEROV



UTILIZAR EL SIGNO DE LA DERIVADA:



SGD



SIGNO

$$sgn(x) = \frac{x}{\sqrt{x^2}} \Rightarrow sgn\frac{\partial J}{\partial w} = \frac{\frac{\partial J}{\partial w}}{\sqrt{\left(\frac{\partial J}{\partial w}\right)^2}}$$

PROBLEMA CON MINI-BATCHES: