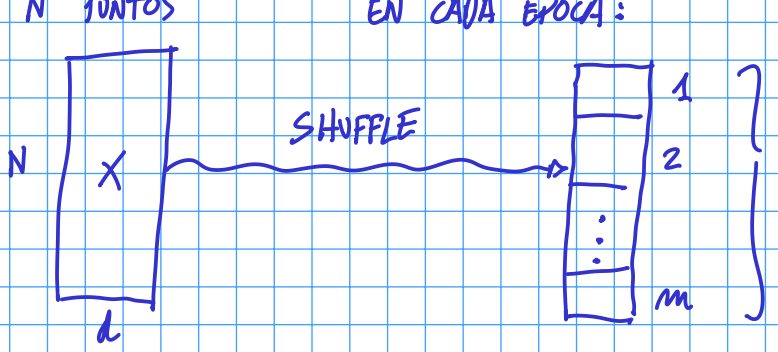
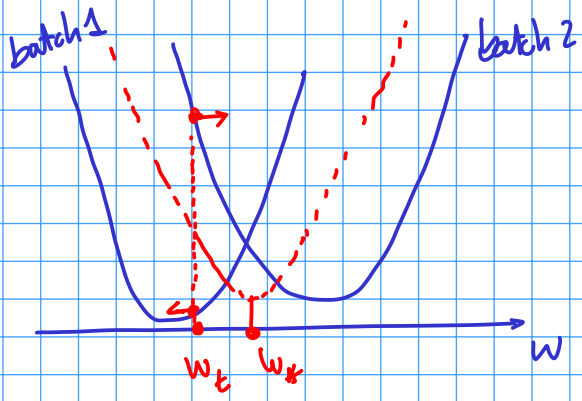
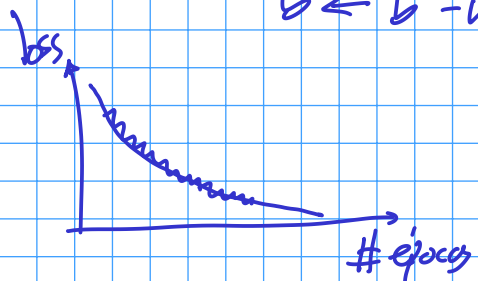
SGD



Con cada mini-batch → 1 actualización pesos

$$w \leftarrow w - \eta \cdot \nabla_w J$$
$$b \leftarrow b - \eta \cdot \nabla_b J$$





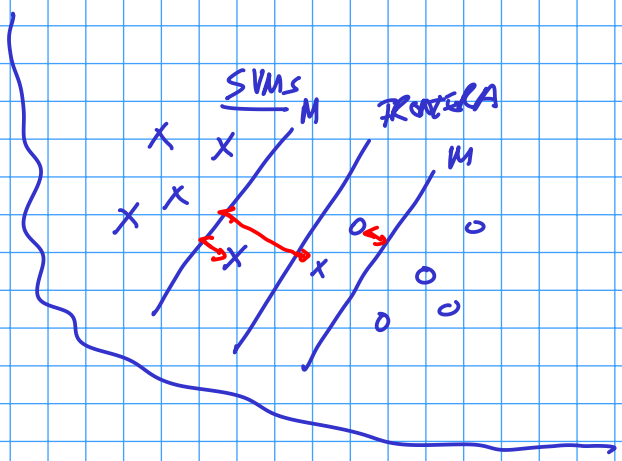
RECETA:

PROBLEMA	ACTIVACIÓN (SALIDA)	COSTE
REGRESIÓN	— (LINEAL)	$MSE = \frac{1}{2N}\sum_{i=1}^{N}(y_i - t_i)^2$
CLASIFICACIÓN 2 CLASES	SIGMOIDE $\sigma = \frac{1}{1+e^{-z}}$	CATEGORICAL CROSS ENTROPY $= \frac{-1}{N}\sum_{i=1}^{N}\left[t_i \log y_i + (1-t_i)\log(1-y_i)\right]$
CLASIFICACIÓN k CLASES ($k>2$)	SOFTMAX $y_j = \frac{e^{z_j}}{e^{z_1}+e^{z_2}+\cdots+e^{z_k}}$ (LINEAL)	CATEGORICAL CROSS-ENTROPY $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} t_{ij}\log y_{ij}$ (HINGE)
K SALIDAS SIGMOIDES	k SIGMOIDES	BINARY CROSS-ENTROPY $-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k}\left[t_{ij}\log y_{ij} + (1-t_{ij})\log(1-y_{ij})\right]$

HINGE LOSS:

$$C(\mathbf{y}(\mathbf{x}), \mathbf{t}) = C(\mathbf{z}^{(K)}, \mathbf{t}) = \sum_{j \neq o} \max\left(0, z_j^{(K)} - z_o^{(K)} + \Delta\right)$$

$z_o^{(K)}$ ← objetivo



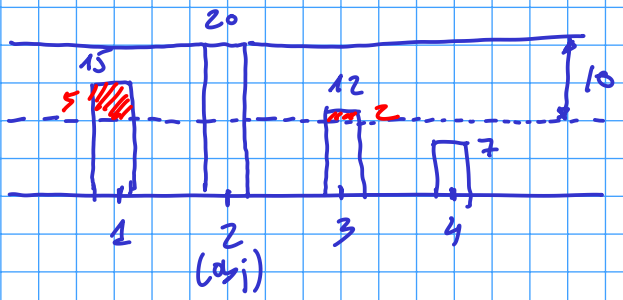
EJEMPLO:

$t =$ [0, 1, 0, 0]

$\vec{x} \rightarrow$ [ ] ⋯ [ ]

$\vec{z}$: 1 → 15, 2 → 20, 3 → 12, 4 → 7

$\Delta = 10$



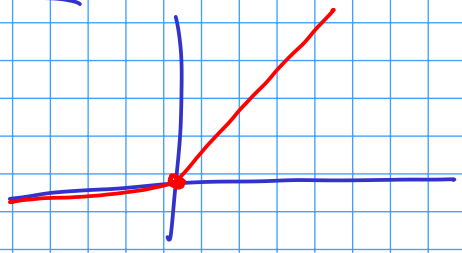
$C = 5 + 0 + 2 + 0 = 7$

$$\frac{dL}{dz_j} = \begin{cases} z_j = z_{obj} \rightarrow \text{— Nº puntos dentro del margen} \\ z_j \neq z_{obj} \rightarrow \begin{array}{l} 1 \text{ si dentro del margen} \\ 0 \text{ si fuera del margen} \end{array} \end{cases}$$

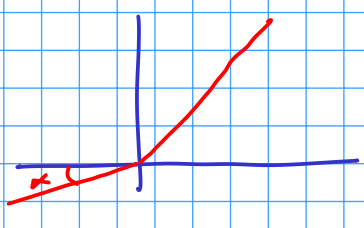
$\vec{x}$ [ ] ⋯ $\vec{h}$ [W, b] $\vec{z} = W\vec{h} + b \rightarrow \vec{y} = \sigma(\vec{z})$

$\vec{z}$: LOGITS ⇒ $\vec{y}$: PROBS.

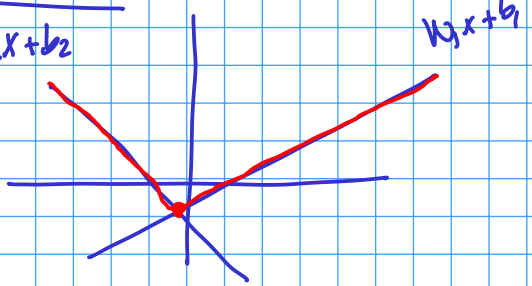
RELU



LEAKY RELU



MAXOUT



MODELO KERAS PARA MNIST