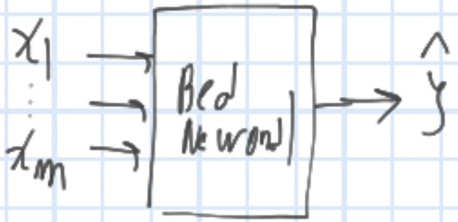
(1) Regresión



* Modelo

$$Y_i | \bar{X}_i = \bar{x}_i \sim \mathcal{N}(\mu_i, 1)$$

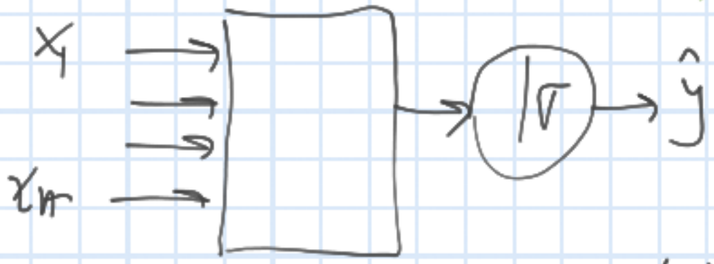
$$\mu_i = f(\bar{x}_i, \bar{\theta})$$

* Costo $J_\theta = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\bar{x}_i, \theta))^2$

* Optimizador
  - Mini-Batch

(1)

(2) Clasificación Bin



* Modelo $Y_i | \bar{X}_i = \bar{x}_i \sim Ber(p_i)$

$$p_i = \sigma(f(\bar{x}_i, \bar{\theta}))$$

* Aplicamos MV

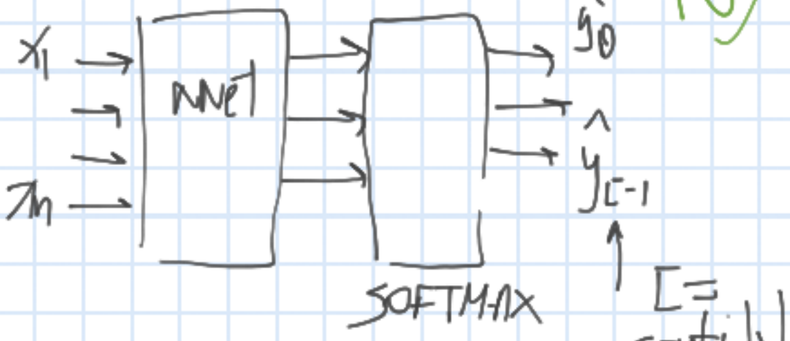
$$J_\theta = -\frac{1}{n} \sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

labels $\in \{0, 1\}$

* Optimizador → SGD / Mini-batch

(2)

(3) Clasificacion Multi Clase



* Modelo

$$Y_i | \bar{X}_i = \bar{x}_i \sim Multinomial$$

$$\mathbb{P}(Y_i = 0 | \bar{X}_i) = p_0$$

$$\mathbb{P}(Y_i = 1 | \bar{X}_i) = p_1$$

$$\mathbb{P}(Y_{L-1} = 1 | \bar{X}_i) = p_{L-1}$$

¿Costo?

(3)

(2) Probar que optimizar $D_{KL}(y_i \| \hat{y}_i)$ $\equiv$ optimizar MV

$$\arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} D_{KL}(y_i \| \hat{y}_i)$$ (clasificación binaria)

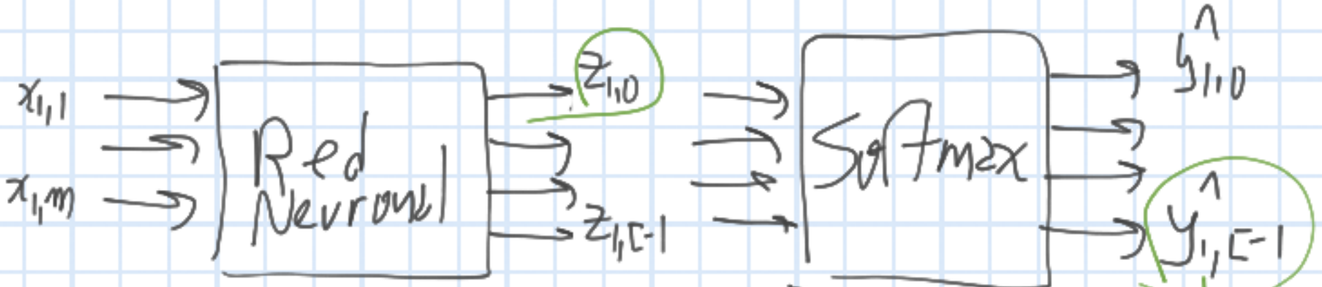
$$\frac{1}{n} \sum_{i=1}^{n} y_i \log\left(\frac{y_i}{p_i}\right) + (1-y_i) \log\left(\frac{1-y_i}{1-p_i}\right)$$ → Depende de θ

$$\frac{1}{n} \sum_{i=1}^{n} \left(-y_i \ln(p_i) - (1-y_i) \ln(1-p_i) \right) + \left(y_i \ln(y_i) + (1-y_i) \ln(1-y_i) \right)$$

cte para el problema de optimización

$$D_{KL}(y_i \| \hat{y}_i) = \underline{H(y_i, \hat{y}_i)} - \underline{H(y_i)}$$

(3) MV aplicada al modelo multinomial



$x_{1,1}$ ⋯	$x_{1,m}$	$y_1 = \{0, \dots, C-1\}$	$[0\ 0\ 1 \cdots 0]$
⋮		⋮	$[1\ 0\ 0 \cdots 0]$
$x_{n,1}$	$x_{n,m}$	$y_n = \{0, \dots, C-1\}$	$[0\ 0\ 0 \cdots 1]$

One hot encoding

y

VS

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$g(z_{1,0}) = e^{z_{1,0}} \Big/ \sum_{k=0}^{C-1} e^{z_{1,k}}$$

$$\vdots$$

$$g(z_{1,C-1}) = e^{z_{1,C-1}} \Big/ \sum_{k=0}^{C-1} e^{z_{1,k}}$$

$\text{sum}() = 1$

¿Que pasa cuando tengo muchas clases?

$\sum_{k=0}^{C-1} e^{z_k}$ → Tarda mucho en calcularse!?!?

↳ Negative sampling

¿Cómo queda la verosimilitud?

$$\arg\min_{W} \left(-\frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{C-1} \delta_{k,y_i} \log \left(\frac{e^{z_k}}{\sum_{w=0}^{C-1} e^{z_w}} \right) \right) \rightarrow J_{\bar{\theta}}$$

$$\delta_{k,y_i} = \begin{cases} 1 & y_i = k \\ 0 & y_i \neq k \end{cases}$$

$J_{\bar{\theta}}$ ↓

Cross Entropy