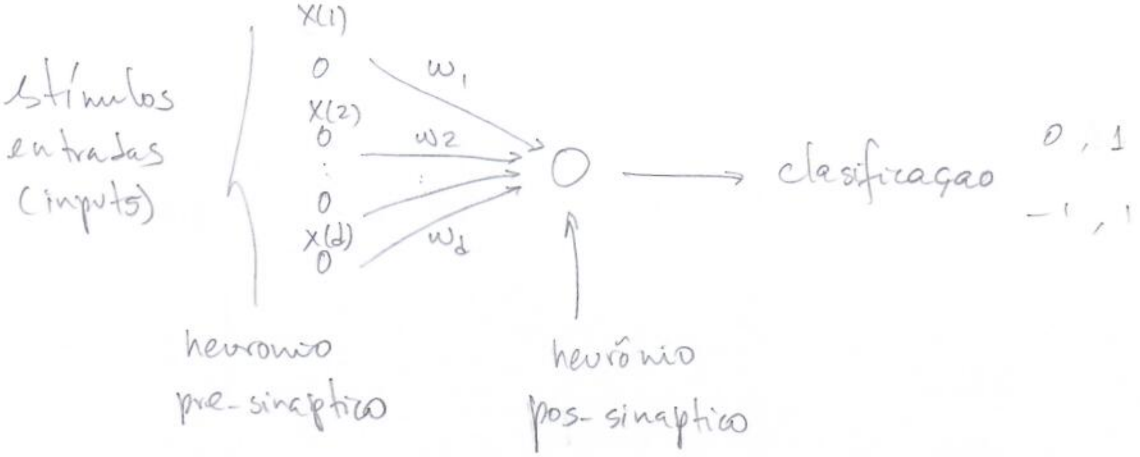
1963 Rosemblat (Perceptron)



$X = \big(X(1), X(2), \dots, X(d) \big) \quad , \quad X(i) \in \mathbb{R}$

$w_1, w_2, \dots, w_d$ pesos sinâpticos

Potencial de membrana

$$u = \sum_{i=1}^{d} w_i X(i) \quad , \quad \text{si } u > \theta \text{ disparo}$$

Vamos re-escrever:

$u - \theta = z$, então disparo se $z > 0 \Rightarrow Y = \mathbb{1}_{\{z>0\}}$

$$z = w_0 + \underbrace{\sum_{i=1}^{d} w_i X(i)}_{u}$$

Classificador $\quad Y = \mathbb{1}_{\{z>0\}}$, $\boxed{z = w_0 + \sum_{i=1}^{d} w_i X(i)}$

Regressor linear.

"Regressao lineal"

função : $X \longmapsto$ no. real

$X = (X(1), \dots, X(d))$

Dada uma amostra de treino $(X_1, Y_1), \dots, (X_n, Y_n)$ iid

onde $X_i = (X_i(1), \dots, X_i(d))$ com $X_i(j) \in \mathbb{R}$ e $Y_i \in \{0,1\}$

o objetivo da rede neural é calibrar automaticamente para fazer a classificação.

escolher os "pesos sinápticos".

Regresão linear:

$\mathcal{X} = \mathbb{R}^d$ — X : vector aleatorio
$\mathcal{Y} = \mathbb{R}$ — Y : variavel aleatória

definidas no mesmo espaço de probabilidade isto é a distribuçao conjunta de X e Y está bem definida.

Seja $f : \mathcal{X} \to \mathcal{Y}$

contínua por pedaço
"mesurável", isto é posso calcular a probabilidade de eventos do tipo $\{x \in \mathcal{X} : f(x) < r\}$.

Risco medio quadratico de f:

$$R(f) = \mathbb{E}\left[(f(X) - Y)^2\right]$$

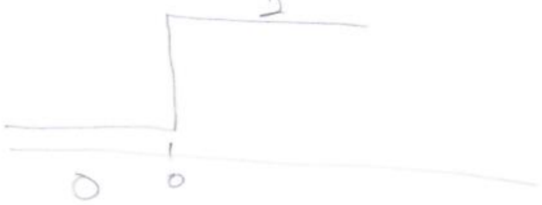
Desconhecido! na prática.

Dada uma amostra
$(X_1, Y_1), \dots, (X_n, Y_n)$ iid com mesma lei (X, Y)

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Y_i)^2$$ risco empírico.

Classificação

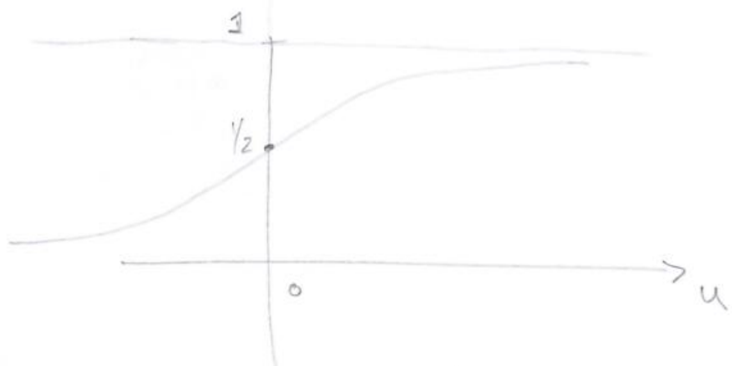
$$1_{\{w_0 + \sum_{i=1}^{d} w_i X(i) > 0\}} = H\left(w_0 + \sum_{i=1}^{d} w_i X(i)\right)$$



H (Heavyside)

Classificação mais suave

$$\sigma(u) = \frac{1}{1+e^{-u}}$$



$$\sigma(u) \to \begin{cases} 1, & \text{quando } u \to +\infty \\ 0, & \text{quando } u \to -\infty \end{cases}$$

$$\sigma(u) = 1 - \sigma(-u)$$

$y \in \{-1, 1\}$

$$\mathbb{P}(Y=y \mid X) = \frac{1}{1+e^{-y\,u(X)}} \quad \text{onde} \quad u(X) = w_0 + \sum_{i=1}^{d} w_i X(i)$$

$\mathbb{P}(Y=1 \mid X)$ é função crescente de $u(X)$

$\mathbb{P}(Y=-1 \mid X)$ é função decrescente de $u(X)$.

Dada uma amostra $(X_1, Y_1), \dots, (X_n, Y_n)$ iid com

$X_m \in \mathcal{X} = \mathbb{R}^d$, $Y_m \in \mathcal{Y} = \{-1, 1\}$ com

$$\mathbb{P}(Y_m = 1 \mid X) = \frac{1}{1+e^{-u(X)}}$$

Calcular a verossimilança da amostra.

$$\log \mathbb{P}\left((x_1,y_1), \ldots, (x_n,y_n) \mid x_1, \ldots, x_n \right)$$

$$\log \prod_{i=1}^{n} P(x_i,y_i) = \log \prod_{i=1}^{n} \frac{1}{1+e^{-y_i u(x_i)}}$$

$$= \sum_{i=1}^{n} \left[\log(1) - \log\left(1+e^{-y_i u(x_i)}\right) \right]$$

$$= \sum_{i=1}^{n} \log\left(1-e^{-y_i u(x_i)}\right) = \sum_{i=1}^{n} \log\left(1-e^{-y_i (a+b x_i)}\right)$$

$$u(x) = \underset{\substack{\uparrow \\ w_0}}{a} + \underset{\substack{\uparrow \\ w_1}}{b} x$$

Pergunta: Calcular $\hat{a}$ e $\hat{b}$ que maximizam a verosimilhança da amostra.