Agentes

$a_1, a_2, \dots, a_t$ (acciones)

$P_1, P_2, \dots, P_t, P_{t+1}$ (Persepción)

Performance

Entorno

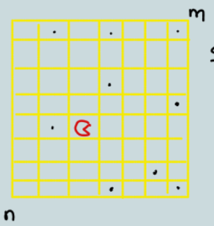
Actividades

Sensores

Estado s:

$$s = (s_1, s_2, \dots, s_n) \in D_1 \times D_2 \times \cdots \times D_n$$

Pacman



$$s = (fila, columna, i_1, \cdots, i_{n \times m})$$

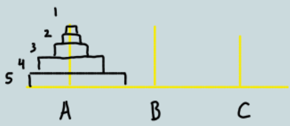
i_K = indicador si hay punto en la posición

$$c = K // m$$
$$f = K \div m$$

$$f \in \{0, 1, \cdots, n-1\}$$
$$c \in \{0, 1, \cdots, m-1\}$$
$$i_K \in \{0, 1\} \ \forall_K = 0, \dots, m \times n - 1$$

Dimensiondlidad:

$$|S| = n * m * 2^{n \times m} - (n * m)$$



1	0	0
2	0	0
3	0	0
4	0	0
5	0	0

$$S = [A, B, C]$$

$$A = \{0, 1, 2, 3, 4, 5\}$$
$$B = \{0, 1, 2, 3, 4, 5\}$$
$$C = \{0, 1, 2, 3, 4, 5\}$$

$$|S| = A * B * C$$

Problema de Los esclavistos

$$s = [E, O, L]$$

$$E = \{0, 1, 2, 3\}$$
$$O = \{0, 1, 2, 3\}$$
$$L = \{I, D\}$$

$$|S| = 4 \times 4 \times 2 = 32$$

Modelos estáticos/dinámicos
discretos/continuos
observable/Parcialmente obs./No obs.
Determinista/Estocastico
Episodicos/continuos

Determinista, estatico, observable, discreto

$$s = f(a)$$

$$s = (s_1, \dots, s_n) \in D_1 \times \cdots \times D_n = S$$

$a \in A$, $a \in A(s)$ acciones Legales en s

Determinista, dinámico, observable, discreto

$$s_{K+1} = f(s_K, a_K)$$

Determinista, dinámico, parcialmente observable, discreto

$$s_{K+1} = f(s_K, a_K)$$

$$P_K = g(s_K)$$

Determinista, dinámico, parcialmente observable, contiuo

$$\dot{x}_t = f(x_t, a_t)$$
$$P_t = g(x_t)$$

Estocastico, discreto, estatico, observable

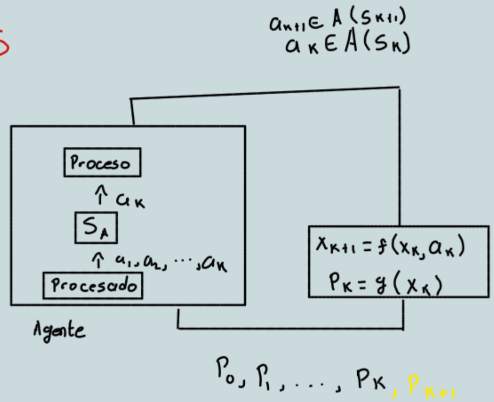
$$S = \{s^{(1)}, \dots, s^{(n)}\} \qquad M = card(S)$$

$$P(s|a) = \begin{bmatrix} P_r[S = s^{(1)} | a] \\ P_r[S = s^{(2)} | a] \\ \vdots \\ P_r[S = s^{(n)} | a] \end{bmatrix}$$

Estocastico, discreto, dinámico, observable

$$P_r[S_{K+1} | s_K, a_K] = \begin{bmatrix} P_r[S_{K+1} = s^{(1)} | s_K, a_K] \\ P_r[S_{K+1} = s^{(2)} | s_K, a_K] \\ \vdots \\ P_r[S_{K+1} = s^{(n)} | s_K, a_K] \end{bmatrix}$$

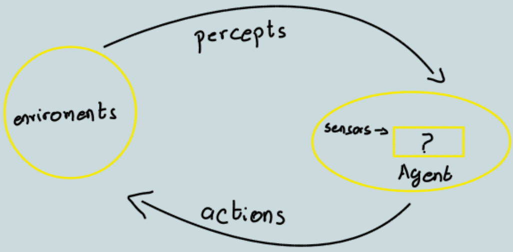
PEAS



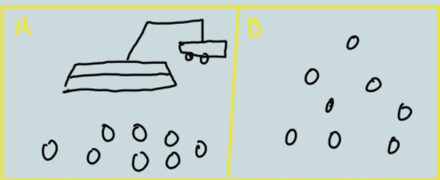
$f_A : P^* \to A$

Intelligent agent

Agent and Environments



Vacuum-cleaner world



Percepts: Location and contents

Actions: Left, Right, suck, NoOP

$S = (A, B, R^*)$

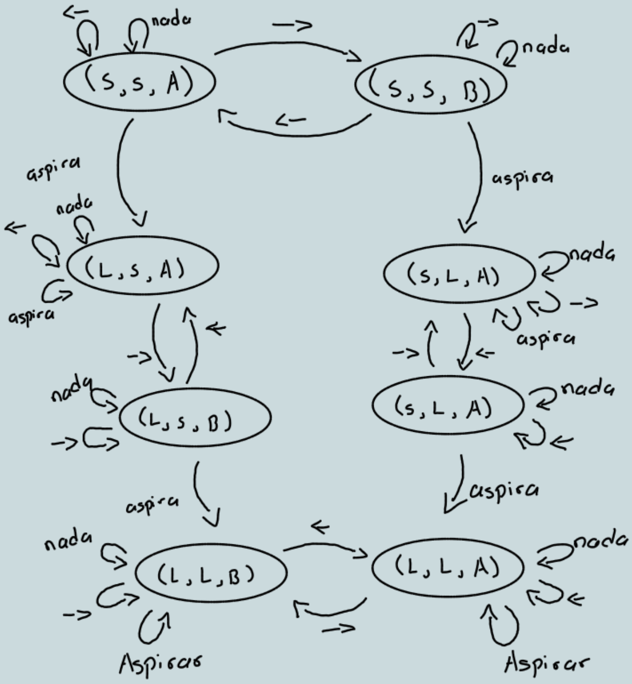
$A \in \{L, s\}$

$B \in \{L, s\}$

$R \in \{A, B\}$

$A = \{\leftarrow, \rightarrow, \text{aspira}, \text{nada}\}$

$p = (R, s[\text{val}(R)])$



- Reflex: $a = f(p)$
- History-based: $a = f([p_1, \ldots, p_n])$
- Model-based: $a = f_m(p)$ and may update m
- Goal-based agents $a = f_m(p, s)$ where a best archives goal
- Utility-based agents $a = f_m(p, s)$ where a maximizes

Goal-based agent

Simple reflex agent

Learning agent

Entornos

Est. / din.

Obs. / det.

desconocida

Agentes reactivos

Entornos

Din. / disc.

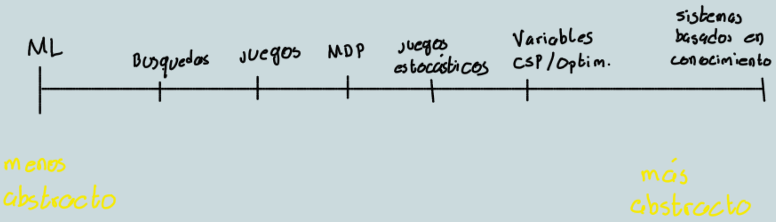
Obs. / Det.

conocidos

Agentes basados en metas

Estadísticos

modelos basados utilidad



Clonar el repositorio

https://github.com/IA-UNISON/1-Agentes-Inteligentes

Entorno

S = Estado

$S = (s_1, \dots, s_n) \in D_1 \times \cdots \times D_n = S$

$A = \{a_1, \dots, a_m\}$

$a_t \in A(s_t)$ acciones Legales

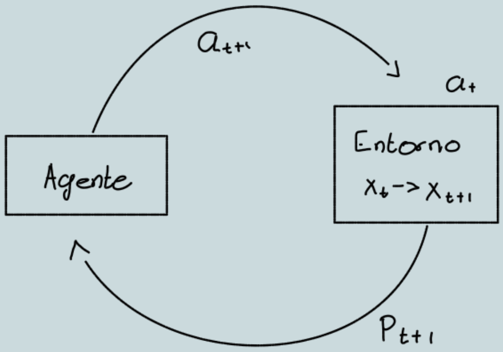
$P_t \in \mathcal{P}$ espacio de percepciones

$P_t = \text{percepción}(s_t)$

percepción: $S \to \mathcal{P}$

$S_{t+1} = \text{transición}(S_t, a_t)$

$C = \text{costo}(S_t, a_t, S_{t+1})$



Agente-reactivo

f_A es desconocida

$f_A : X \to y$

Tipicamente $X = \mathbb{R}^n$

$X = (x_1, \dots, x_n) \in \mathbb{R}^n$

si $y = \mathbb{R}$ Regresión

si $y = \{-1, 1\}$ Clasificación binaria

si $y = \{c_1, \dots, c_K\}$ Clasificación en varios clases

$\{x^{(1)}, \dots, x^{(m)}\}$ una muestra de X

$D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$

donde $y^{(i)} = f_A(x^{(i)}) + \hat{e}$ Error variable de distribución desconocida

El problema es encontrar una función $h^* : x \to y$

tal que $h^* \approx f_A$ $\quad h^* \in H$ hipótesis posibles

$H = \{h \mid h : \mathbb{R} \to \mathbb{R} \text{ donde } h_\alpha(x) = \alpha x, \forall \alpha \in \mathbb{R}\}$

$x = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}]$

$x^{(i)} \in \mathbb{R}^{12}$

$$\begin{bmatrix} x_1^{(1)} & \cdots & x_{12}^{(1)} \\ x_1^{(2)} & \cdots & x_{12}^{(2)} \\ \vdots & & \\ x_1^{(m)} & \cdots & x_{12}^{(m)} \end{bmatrix} \cdot \begin{bmatrix} \$ \\ \vdots \\ \$ \end{bmatrix}$$

$h : X \times H \to y$

$h(x^{(i)}; \theta) = \hat{y}^{(i)}$

$h_\theta(x^{(i)}) = \hat{y}^{(i)}$

Si θ es un vector de parámetros fijos

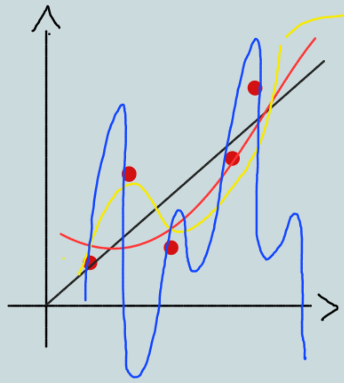
$h_\theta : X \to y$

$h(x) = \sum_{j=1}^{T} w_j x^j + w_0, \quad x^j \in \mathbb{R}$

$h_w : \mathbb{R} \to \mathbb{R}$

$w = (w_0, w_1, \dots, w_T) \in \mathbb{R}^{T+1}$

$\hat{y} = w_1 x + w, \quad \hat{y} = w_2 x^2 + w_1 x_1 + b$



Más optima

no muy generica

no demasiado representativa

$h^* \approx f$

Aprendizaje supervisado

Hipótesis:

1.- $f: X \to Y$ existe y es desconocida

2.- Tengo $\{x^{(1)}, \dots, x^{(m)}\} \subseteq X$ una distribución desconocida

3.- $D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, donde

$y^{(i)} = f(x^{(i)})\, e^{(i)}$ donde $e^{(i)}$ son variables aleatorias.

4.- Tenemos una función "parametrizada"

$h: X \cdot \Theta \to Y$, típicamente $\Theta = \mathbb{R}^p$

5.- Para un valor específico de Θ,

$h_\Theta: X \to Y$

6.- $h_\Theta \in H$ conjunto de hipótesis

7.- El aprendizaje supervisado consiste en encontrar

$h^* \in H$ tal que

$h^* \approx f$

$$E_{out}(h^*) = \underset{x \in X}{\text{Esperanza}}\left[L(f(x), h^*(x))\right]$$

$f \approx h^*$ si y solo si $E_{out}(h^*) \approx 0$

$$E_{in}(h^*) = \frac{1}{M}\sum_{i=1}^{M} L(y^{(i)}, h(x^{(i)}))$$

$$f \approx h^* \text{ si } \begin{cases} E_{in}(h^*) \approx 0 \\ E_{in}(h^*) \approx E_{out}(h^*) \end{cases}$$

Desigualdad de Hoeffding

$$P_\theta\left(|E_{out}(h^*) - E_{in}(h^*)| > \varepsilon\right) \leq \frac{1}{2} e^{-2\varepsilon^2 M}$$

donde M es el tamaño de la muestra

$d_{vc}(H) \approx$ # Parametros Independientes

El aprendizaje es posible si:

$10 * d_{vc}(H) << M \leftrightarrow E_{in}(h^*) \approx E_{out}(h^*)$

Problema:

Otorgar crédito

$$\begin{array}{ccc} & \begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} & y \\ \begin{array}{c} x^{(1)T} \to \\ x^{(2)T} \to \\ \vdots \\ x^{(m)T} \to \end{array} & \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_{(n)}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_{(n)}^{(2)} \\ \vdots & & & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_{(n)}^{(m)} \end{bmatrix} & \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix} \end{array}$$

$\underset{(m,n)}{X}$ $\qquad \underset{(m,1)}{y}$

$x^{(i)} \in \mathbb{R}^n$

$y^{(i)} \in \mathbb{R}$

$$D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$$

$$h_\Theta(x) = w_1 x_1 + \dots + w_n x_n + b = \sum_{j=1}^{n} w_j x_j + b = w^T x + b$$

$$= x^T w + b$$

$\Theta = (w_1, \dots, w_n, b) \in \mathbb{R}^{n+1}$

si $x = (x_1, \dots, x_n) \in \mathbb{R}^n$

$w = (w_1, \dots, w_n) \in \mathbb{R}^n$

Aprendizaje:

$E_{in}(h^*) \approx E_{out}(h^*)$

$E_{in}(h^*) \approx 0$

$$h^* = \underset{h \in H}{\arg\min}\ E_{in}(h) \leftrightarrow \Theta^* = w^*, b^*$$

$$= \underset{\substack{w \in \mathbb{R}^n \\ b \in \mathbb{R}}}{\arg\min}\ E_{in}(h_{w,b})$$

Libro

Pantaleon y las visitadoras

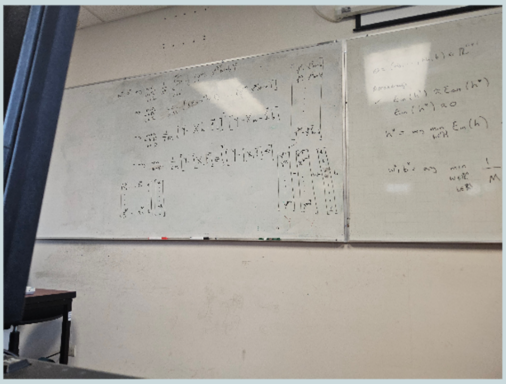
MSE = Mean, Square, Ersor

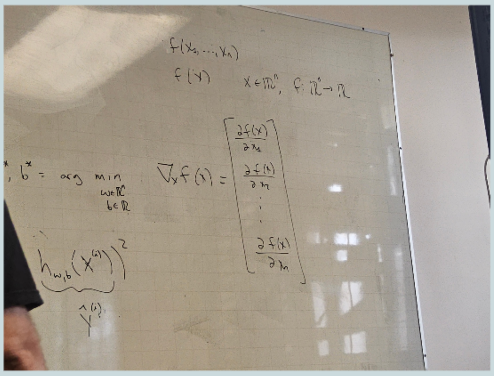
$$w_1^*,\ b^* = \arg\min_{\substack{w\in\mathbb{R}^n\\ b\in\mathbb{R}}} \frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\Big(y^{(i)} - \underbrace{h_{w,b}(x^{(i)})}_{\hat{y}^{(i)}}\Big)^2$$

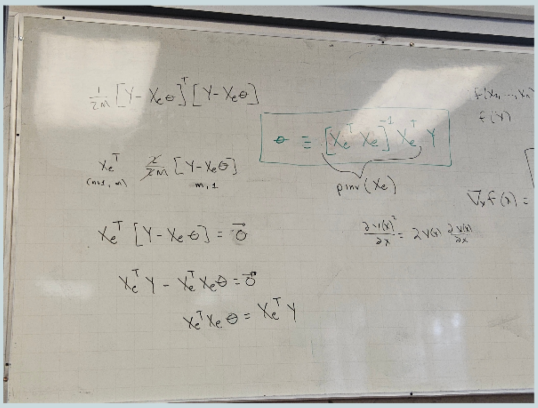
$$w^*, b^* = \arg\min_{\substack{w\in\mathbb{R}^n\\ b\in\mathbb{R}}} \frac{1}{M}\sum_{i=1}\frac{1}{2}\left(y^{(i)} - x^{(i)T}w - b\right)^2$$

$$w^*, b^* = \arg\min_{\substack{w\in\mathbb{R}^n\\ b\in\mathbb{R}}} \frac{1}{2M}\left[\left(y^{(1)} - x^{(1)T}w - b\right)\cdots\left(y^{(m)} - x^{(m)T}w - b\right)\right]$$

$$\begin{bmatrix} y^{(1)} - x^{(1)T}w - b \\ y^{(2)} - x^{(2)T}w - b \\ \vdots \\ y^{(m)} - x^{(m)T}w - b \end{bmatrix} \rightarrow \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \\ y^{(m)} \end{bmatrix} - \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix} w - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} b$$







MSR

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^{n+1}} \frac{1}{2M}\sum_{i=1}^{M}\left(Y^{(i)} - \left[x^{(i)}\ 1\right]\begin{bmatrix} w \\ b \end{bmatrix}\right)^2$$

$$X_e = \begin{bmatrix} x_1^{(1)} & \cdots & x_n^{(1)} & 1 \\ \vdots & & & \vdots \\ x_1^{(m)} & \cdots & x_n^{(m)} & 1 \end{bmatrix}$$

$$\underset{(m,1)}{\hat{Y}} = \underbrace{\underset{(m,n+1)}{X_e}\ \underset{(n+1,1)}{\theta}}_{(m,1)}$$

$$\underset{(m,1)}{E} = \underset{(m,1)}{Y} - \underset{(m-1)}{\hat{Y}}$$

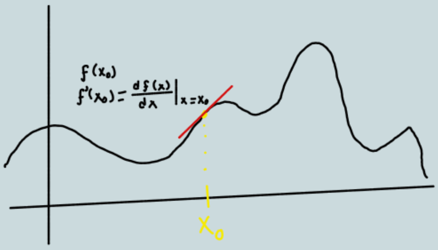
$$E_{in}(\theta) = \frac{1}{M} E^T E$$

Pseudo inversa (P inv)

$$\theta^* = \underbrace{\underbrace{\boxed{\underbrace{\left(\underset{(n+1,m)}{X_e^T}\ \underset{(m,n+1)}{X_e}\right)^{-1}}_{(n+1,n+1)}\ \underset{(n+1,m)}{X_e^T}}}_{(n+1,m)}\ \underset{(m,1)}{Y}}_{(n+1,1)}$$

$f : \mathbb{R} \to \mathbb{R}$

$f(x)$



$$x_0$$
$$x_1 = x_0 - \eta f'(x_0)$$
$$x_2 = x_1 - \eta f'(x_1)$$
$$\Rightarrow x_{k+1} \leftarrow x_k - \eta f'(x_k)$$

$$\Theta_{k+1} \leftarrow \Theta_k - \eta \nabla f(\Theta_k)$$

$$h_\Theta(x) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b = w^T x + b =$$
$$= [x_j^T, 1]\begin{bmatrix} w \\ b \end{bmatrix} = [x_j^T, 1] = \Theta$$

$$\Theta^* = w^*, b^* = \arg\min_{\Theta \in \mathbb{R}^{n+1}} \frac{1}{2M} \underbrace{\sum_{i=1}^{M} (Y^{(i)} - [x_j^T, 1]\Theta)^2}$$

Para w_i

$$\frac{\partial E_{in}(\Theta)}{\partial N_j} = \frac{1}{2M} \sum_{i=1}^{M} 2(Y^{(i)} - \hat{Y}^{(i)})\, x_j^{(i)}$$

$$\frac{\partial E_{in}(\Theta)}{\partial N_j} = \frac{1}{M} \sum_{i=1}^{M} -(Y^{(i)} - \hat{Y}^{(i)})$$

$$\nabla J(\Theta) = \begin{bmatrix} \frac{1}{M}\sum_{i=1}^{M} -(Y^{(i)} - \hat{Y}^{(i)})\, x_1^{(i)} \\ \frac{1}{M}\sum_{i=1}^{M} -(Y^{(i)} - \hat{Y}^{(i)})\, x_2^{(i)} \\ \vdots \\ \frac{1}{M}\sum_{i=1}^{M} -(Y^{(i)} - \hat{Y}^{(i)})\, x_n^{(i)} \\ \frac{1}{M}\sum_{i=1}^{M} -(Y^{(i)} - \hat{Y}^{(i)}) \end{bmatrix} = \frac{1}{M}$$

$$= \underbrace{\begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(m)} \\ \vdots & \vdots & & \vdots \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \end{bmatrix}}_{X_e^T} \underbrace{\begin{bmatrix} y^{(1)} - \hat{y}^{(1)} \\ y^{(2)} - \hat{y}^{(2)} \\ \vdots \\ y^{(m)} - \hat{y}^{(m)} \end{bmatrix}}_{Y - X_e\Theta}$$

$$\Theta \leftarrow \Theta - \eta \nabla E_{in}(\Theta)$$

$$\Theta \leftarrow \Theta + \frac{\eta}{M} X_e^T (Y - X_e \Theta)$$

```
def dg_lin(X, Y, w0, b0, lr, max_epochs, e_tol):
    M = X.shape[0]

    w = w0.copy()
    b = b0.copy()
    hist = []

    for _ in range(max_epochs):
        y_est = X @ w + b
        Err = y - y_est
        hist.append(np.square(Err).mean())
        grad_w = -(1/M) X.T @ Err
        d_b = Err.mean()
        w -= lr * grad_w
        b -= lr * d_b
        if np.abs(grad_w).max() < e_tol:
            break
    return w, b, hist
```

Probablemente Aproximadamente Correcto

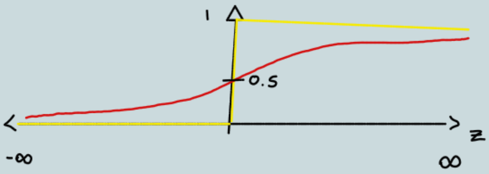
Perceptron

Regresión Lógistica

$\hat{a} = P_r(Y=1 \mid X=x;\Theta)$

$\hat{y} = \begin{cases} 1 & \text{si } \hat{a} > h = \text{umbral} \\ -1 & \text{en otro caso} \end{cases}$

$\hat{a} = f(\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + b) = x^T\omega + b = x_\circ^T\Theta$



$\hat{a} = \sigma(\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + b) = x^T\omega + b = x_\circ^T\Theta$

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

sigmoid, ℓogistic

$$\underbrace{\begin{bmatrix} x_1^{(1)} & \cdots & x_m^{(1)} \\ \vdots & & \\ x_1^{(n)} & \cdots & x_m^{(n)} \end{bmatrix}}_{X} \quad \underbrace{\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}}_{y} \quad \underbrace{\begin{bmatrix} a^{(1)} \\ \vdots \\ a^{(m)} \end{bmatrix}}_{A}$$

$y^{(i)} \in \{-1, 1\}$

$E_{in}(\omega, b) = \frac{1}{M}\sum_{i=1}^{M} Loss(a^{(i)}, \hat{a}^{(i)})$

$Loss(a^{(i)}, \hat{a}^{(i)}) = \begin{cases} -Log(\hat{a}^{(i)}) & \text{si } a^{(i)} = 1 \\ -Log(1-\hat{a}^{(i)}) & \text{si } a^{(i)} = 0 \end{cases}$

$Loss(a^{(i)}, \hat{a}^{(i)}) = -a^{(i)} Log(\hat{a}^{(i)}) - (1-a^{(i)}) Log(1-\hat{a}^{(i)})$

$\omega \leftarrow \omega - L_r \Delta_\omega E_{in}(\omega, b)$

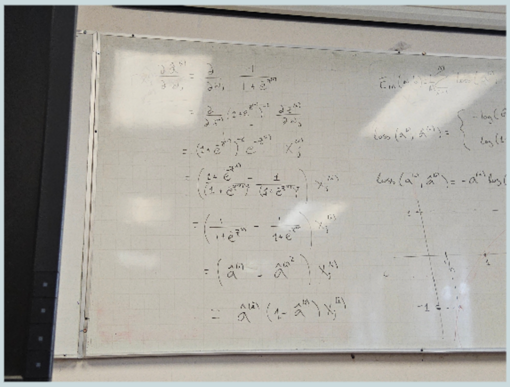
$b \leftarrow b - L_r \frac{\partial}{\partial b} E_{in}(\omega, b)$

$\frac{\partial}{\partial \omega_j} E_{in}(\omega, b) = \frac{\partial}{\partial \omega_j} \frac{1}{M}\sum_{i=1}^{M} -a^{(i)} Log(\hat{a}^{(i)}) - (1-a^{(i)}) Log(1-\hat{a}^{(i)})$

donde

$\hat{a}^{(i)} = \frac{1}{1-e^{-z^{(i)}}}$, y $z^{(i)} = \omega_1 x_1 + \cdots + \omega_n x_n + b$

$= \frac{1}{M}\sum_{i=1}^{M} -\frac{a^{(i)}}{\hat{a}^{(i)}} \frac{\partial a^{(i)}}{\partial \omega_j} + \frac{1-a^{(i)}}{1-\hat{a}^{(i)}} \frac{\partial a^{(i)}}{\partial \omega_j}$



$= \frac{1}{M}\sum_{i=1}^{M} [-a^{(i)}(1-a^{(i)}) + (1-a^{(i)}) a^{(i)}] x_j^{(i)}$

$= \frac{1}{M}\sum_{i=1}^{M} [-a^{(i)} + \cancel{a^{(i)}\hat{a}^{(i)}} + \hat{a}^{(i)} - \cancel{a^{(i)}\hat{a}^{(i)}}] x_j^{(i)}$

$\boxed{\frac{\partial}{\partial \omega_j} E(\omega, b) = -\frac{1}{M}\sum_{i=1}^{M} (a^{(i)} - \hat{a}^{(i)}) x_j^{(i)}}$

$\boxed{\frac{\partial}{\partial b} E(\omega, b) = -\frac{1}{M}\sum_{i=1}^{M} (a^{(i)} - \hat{a}^{(i)})}$

$\omega \leftarrow \omega - L_r X^T (A - \hat{A})$

$b \leftarrow b - \frac{L_r}{M}\sum_{i=1}^{M} (a^{(i)} - \hat{a}^{(i)})$

Función generadora de características

$\Theta : \mathbb{R}^n \to \mathbb{R}^{n'}$

$x \in \mathbb{R}$ $\quad x' = \Theta(x) = (x, x^2, x^3, x^4)$

$x' = \Theta(x) = (\Theta_1(x), \Theta_2(x), \cdots, \Theta_{n'}(x))$

Expansión polinomial

$x = (x_1, x_2)$

$\Theta(x) = (x_1, x_2, x_1^2, x_1 x_2, x_2^2)$

$\Theta(x) = (x_1, x_2, x_1^2, x_1 x_2, x_2^2,$

$H = \{h\}$ tal que

$h(x) = \omega_1 x_1 + \omega_2 x_1^2 + \omega_3 x_1^3 + \omega_4 x_1^4 + \omega_5 x_1^5 + \omega_6 x_1^6 + \omega_7^7 x_1 + b$

$\omega_j = 0$ si $j \geq 3$

$h(x) = \omega_1 x_1 + \omega_2 x_1^2 + b$

$h(x) = \omega^T \Theta(x) + b \qquad \omega, \Theta(x) \in \mathbb{R}^{n'}, \; b \in \mathbb{R}$

$\omega^*, b^* = \arg\min E_{in}(\omega, b)$

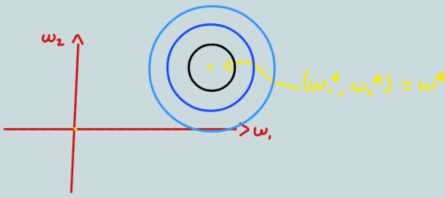
bajo

$\underbrace{\sum_{j=1}^{n'} \omega_j^2 \leq C}$ ← Regularización

$\omega^T \omega \leq C$

$\|\omega\|_2 = \sqrt{\omega^T\omega}$

$\|\omega\| = \sum_{i=1}^{n'} |\omega_i|$



$\lambda = \frac{1}{C}$

$\omega_{reg}^*, b_{reg}^* = \arg\min \left[E_{in}(\omega, b) + \frac{\lambda}{M}\sum_{j=1}^{n} \omega_j^2 \right]$

$\omega_{reg}^*, b_{reg}^* = \arg\min \left[E_{in}(\omega, b) + \frac{\lambda}{M} regu(\omega) \right]$

$\frac{\partial}{\partial \omega_j}\left[E(\omega, b) + \frac{\lambda}{M}\sum_{j=1}^{n} \right] = -\frac{1}{M}\sum_{i=1}^{M} (a^{(i)} - \hat{a}^{(i)}) x_j^{(i)} + \frac{2\lambda}{M}\omega_j$

$regu(\omega) = \begin{cases} \sum_{i=1}^{n} \omega_i^2 = \|\omega\|_2^2 & \ell_2 \\ \sum_{i=1}^{n} |\omega_i| & \ell_1 \end{cases}$

reg. lineal + $\begin{cases} \ell_2 - \text{Ridge} \\ \ell_1 - \text{Lasso} \\ \ell_1 + \ell_2 - \text{Elastic net} \end{cases}$

$$w^*, b^* = \arg\min E_{in}(w, b)$$

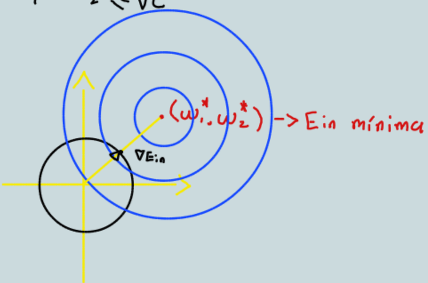
bajo

$$\underbrace{\sum_{j=1}^{n'} w_j^2 \leq C}_{w^T w \leq C} \longleftarrow \text{Regularización}$$

$$\|w\|_2 = \sqrt{w^T w}$$

$$\|w\| = \sum_{i=1}^{n'} |w_i|$$

$$w_1^2 + w_2^2 \leq \sqrt{C}^2$$



$$w_1^2 + w_2^2 = C$$

$$w_{regu} = -K \nabla E_{in}(w, b)$$

$$w_{regu} + K \nabla E_{in}(w, b) = 0$$

$$2\lambda w_{regu} + \nabla E_{in}(w, b) = 0$$

$$\nabla_w (\nabla E_{in}(w, b) + \lambda w^T w)$$

$$\nabla_w (\nabla E_{in}(w, b) + 2\lambda w)$$

multiplicadores de Lagrange

Árboles de desición

```
def generar_arbol(features, X, Y, nodo):
    Si todos los datos misma clase ó
    no hay datos o no hay features
    return nodo

    Var = escoge_feature(feature, X, Y):
      Quitar var de features
      por valor en valores(var):
        Xh, Yh = separa_dato(X, Y, var, valor)
        nh = crea_hijo(n, Yh)
        nh = genera_arbol(features, Xn, Yn, nh)
```

Entropia