

Inteligencia

Artificial

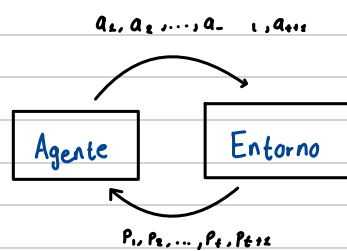
# Inteligencia Artificial

¿Que es realmente la IA?

- Sistemas que piensan como humanos.
- Sistemas que actúan como humanos.
- Sistemas que piensan racionalmente.
- Sistemas que actúan racionalmente.

## Agente racional

Es cualquier cosa que perciba su ambiente a través de sensores y actúe sobre él mediante actuadores.



Un agente se comunica con el entorno que usará para realizar una tarea a partir de una serie de pasos.

## Agente:

Para diseñar un agente, se debe definir:

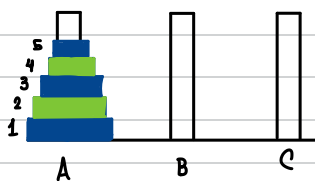
Performance

Entorno

Actuadores

Sensores

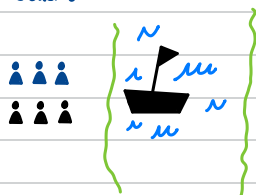
## Problema 1:



$S = (s_1, \dots, s_5) \leftarrow$  Numero de discos  
 $s_i \in \{A, B, C\} \leftarrow$  Torre asignada

$$|S| = 3^5 = 243 \text{ Estados.}$$

## Problema 2:



$S = (E, O, L)$   
 $E = (0, 1, 2, 3)$   
 $O = (0, 1, 2, 3)$   
 $L = (I, D)$

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

## ENTORNO

Entorno Determinista: Poder calcular el valor exacto.

Entorno Estático: Se va usar una única vez, no cambia en el tiempo.

Entorno Observable: Se puede observar todo el estado.

Entorno Discreto: Va a pertenecer a un grupo finito de estados.

Entorno Continuo: Va a pertenecer a un grupo real de estados.

$$S = (s_1, \dots, s_n) \in D_1 \times \dots \times D_n \quad (\text{ESTADO})$$

→ Determinista, Estático, Observable, discreto:

$$S = f(a) \quad \begin{array}{|c|} \hline a \in A \\ \hline \end{array} \quad \begin{array}{|c|} \hline a \in A(s) \quad (\text{acciones legales en } s) \\ \hline \end{array}$$

→ Determinista, dinámico, Observable, discreto:

$$S_{k+1} = f(s_k, a_k)$$

→ Determinista, Estático, Parcialmente Observable, discreto:

$$S_{k+1} = f(s_k, a_k)$$

$$P_k = g(s_k)$$

→ Determinista, Estático, Parcialmente Observable, Continuo:

$$\dot{X} = f(x_t, a_t)$$

$$P = g(x_t)$$

→ Estocástico, Estático, Observable, Discreto

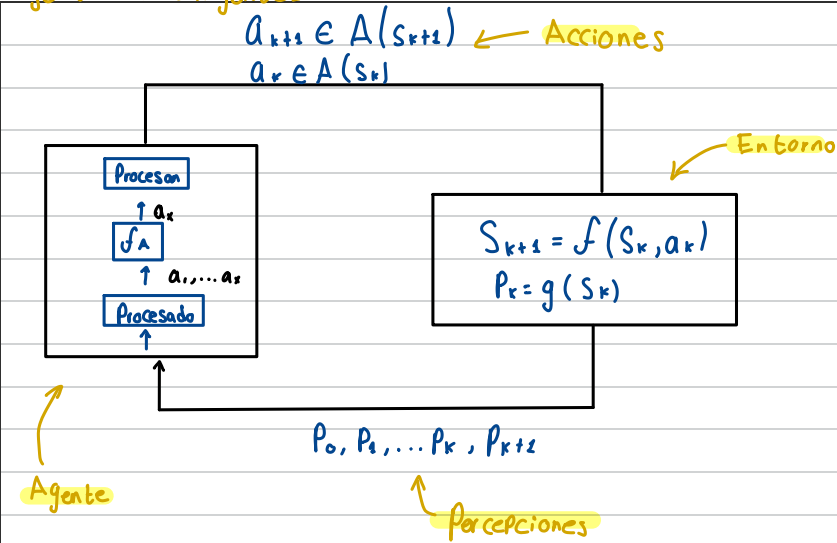
$$S = \{s^{(1)}, \dots, s^{(m)}\} \quad m = \text{Cardinalidad}(S)$$

\* Como es estocástico puede pasar una cosa o la otra entonces podemos sacar la probabilidad de estar en un estado \*

$$Pr[S/a] = \begin{bmatrix} Pr[S=s^{(1)}|a] \\ Pr[S=s^{(2)}|a] \\ \vdots \\ Pr[S=s^{(m)}|a] \end{bmatrix} \quad * \text{ La suma de las probabilidades es } 1 *$$

→ Podemos usar la función de desempeño para evaluar un agente racional.

## Agentes Inteligentes



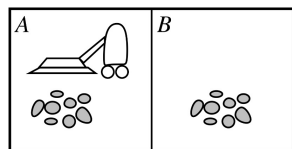
Los agentes incluyen humanos, robots, Softbot, termostatos, etc.

La función del agente se genera desde los historiales de percepción a las acciones.

$$f: p^* \rightarrow A$$

El Programa del agente se ejecuta en la arquitectura física para producir  $f$

### El Problema de la Aspiradora



**Percepciones:** Locación y contenido. Ej [A, Sucio]

**Acciones:** Izq, Der, Limpiar, Nada.

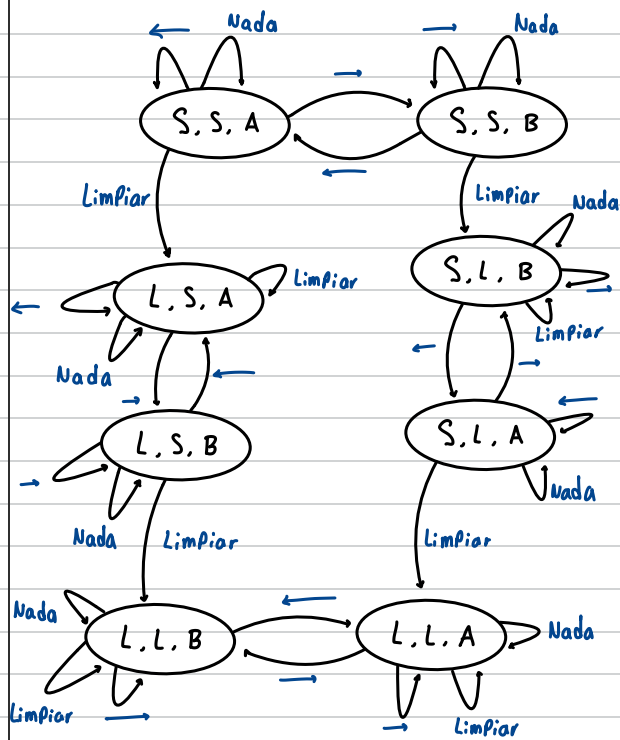
### Agente

- Parcialmente Observable.
- Estocástico.
- Determinista.
- Discreto.

### Estados:

- $S = \{\text{Cuarto 1, Cuarto 2, robot}\}$
- $\text{Cuarto 1} \in \{\text{Limpio, Sucio}\}$
- $\text{Cuarto 2} \in \{\text{Limpio, Sucio}\}$
- $\text{robot} \in \{\text{Cuarto 1, Cuarto 2}\}$

$$|S| = 2 \times 2 \times 2 = 8 //$$



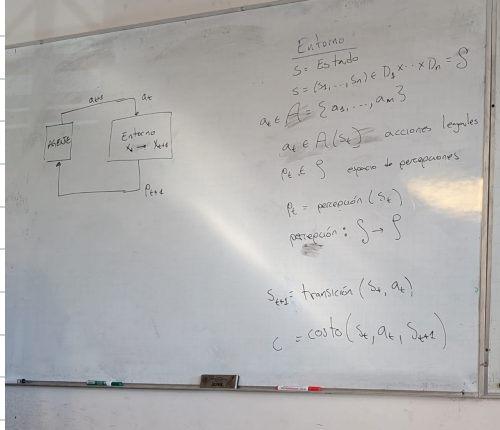
Acciones  $\in \{\leftarrow, \rightarrow, \text{Nada}, \text{Limpiar}\}$



## Entorno

$S = \text{Estado}$

$S = (s_1, s_2, \dots, s_n) \in (D_1, D_2, \dots, D_n)$



$f_A$  es desconocida

$$P \rightarrow \boxed{f_A(p)} \rightarrow a$$

$$f_A: \mathcal{X} \rightarrow \mathcal{Y}$$

tipicamente  $\mathcal{X} = \mathbb{R}^n$

$$x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

Si  $\mathcal{Y} = \mathbb{R}$  Regresión.

Si  $\mathcal{Y} = \{-1, 1\}$  Clasificación Binaria.

Si  $\mathcal{Y} = \{c_1, c_2, \dots, c_K\}$  Clasificación en Varias Clases.

→ Podemos encontrar nuestra función  $f$  se puede mediante datos del Pasado.

$\{x^{(1)}, \dots, x^{(m)}\}$  Una muestra de  $\mathcal{X}$  de dist. desconocida

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

$$\text{donde } y^{(i)} = f_A(x^{(i)}) + \hat{e} \quad \text{V.A. dist. Desc.}$$

El Problema es encontrar una función  $h^*: \mathcal{X} \rightarrow \mathcal{Y}$

tal que  $h^* \approx f_A$   $h^* \in \mathcal{H}$  hipótesis posibles

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

$$h: \mathcal{X} \times \Theta \rightarrow \mathcal{Y}, \quad \Theta = \mathbb{R}^p$$

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

$$h_\theta(x^{(i)}) = \hat{y}^{(i)} \quad \text{Si } \theta \text{ es un vector de parámetros fijos.}$$

$$h_\theta: \mathcal{X} \rightarrow \mathcal{Y}$$

## Hipotesis

1.  $f: X \rightarrow y$  existe y es desconocida
2. Tengo  $\{x^{(1)}, \dots, x^{(m)}\} \subseteq X$  Una muestra  $m$  de datos con 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 V.A.
4. Tenemos una función "Parametrizada"  $h: X \times \Theta \rightarrow y$ , típicamente  $\Theta = \mathbb{R}^p$
5. Para un valor específico de  $\theta$ ,  $h_\theta: X \rightarrow y$
6.  $h_\theta \in H$  Conjunto de hipótesis.
7. El aprendizaje supervisado consiste en encontrar  $h^* \in H$  tal que  $h^* \approx f$

## Error

$L: y \times y \rightarrow \mathbb{R}$  función de pérdida

$$L(y, \hat{y}) = L(f(x), h^*(x))$$

$$L(y, \hat{y}) = \frac{1}{2} (y, \hat{y})^2$$
$$= |y - \hat{y}| \quad y \in \mathbb{R}$$

$$L(y, \hat{y}) = \begin{cases} 0 & \text{si } y = \hat{y} \\ 1 & \text{si } y \neq \hat{y} \end{cases} = \vec{1}\{y \neq \hat{y}\} \quad y \in \{-1, 1\}$$

$$E_{out}(h^*) = \mathbb{E}_{x \in X} [L(f(x), h^*(x))]$$

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

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

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

## Desigualdad de Hoeffding

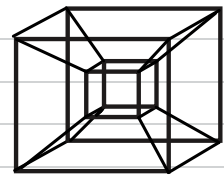
$$P_r(|E_{out}(h^*) - E_{in}(h^*)| > \epsilon) \leq 2e^{-2\epsilon^2 M}$$

donde  $M$  es el tamaño de la muestra

$$d_{vc}(H) \hat{=} \# \text{ Parametros Independientes}$$

El aprendizaje es posible si:

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



$$\begin{bmatrix} x^{(1)T} \\ x^{(2)T} \\ \vdots \\ x^{(m)T} \end{bmatrix} \rightarrow \begin{bmatrix} a_1, \dots, a_n \\ (x_1^{(1)}, \dots, x_n^{(1)}) \\ (x_1^{(2)}, \dots, x_n^{(2)}) \\ \vdots \\ (x_1^{(m)}, \dots, x_n^{(m)}) \end{bmatrix} \quad \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

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

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

$$X \quad y$$

$$(m, n) \quad (m, 1)$$

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

$$h_{\theta}(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b = \sum_{i=1}^n w_i x_i + b = w^T x + b = x^T w + b$$

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

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

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

$$h^* = \arg \min_{h \in \mathcal{H}} E_{in}(h) \longleftrightarrow \theta^* = w^*, b^* = \arg \min_{\substack{w \in \mathbb{R}^n \\ b \in \mathbb{R}}} E_{in}(h_{w,b})$$

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

MSE

Mean, Square, Error.

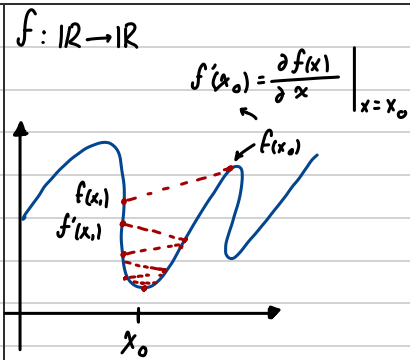
$$= \arg \min_{\substack{w \in \mathbb{R}^n \\ b \in \mathbb{R}}} \frac{1}{M} \sum_{i=1}^M \frac{1}{2} (y^{(i)} - x^{(i)T} w - b)^2$$

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

$$= \arg \min_{\substack{w \in \mathbb{R}^n \\ b \in \mathbb{R}}} \frac{1}{2M} \left[ y - X^T w - \vec{1} b \right]^T \left[ y - X^T w - \vec{1} b \right]$$

$$= \arg \min_{\theta \in \mathbb{R}} \frac{1}{2M} \left[ y - [X, \vec{1}] \theta \right]^T \left[ y - [X, \vec{1}] \theta \right]$$

$$\theta = [X^T X]^{\perp 1} X^T y$$



1.  $x_0$
  2.  $x_1 \leftarrow x_0 - \eta f'(x_0)$
  3.  $x_2 \leftarrow x_1 - \eta f'(x_1)$
  - $\vdots$
  - k.  $x_{k+1} \leftarrow x_k - \eta f'(x_k)$
- Descenso del gradiente

$$h_{\theta}(x) = w_1 x_1 + \dots + w_n x_n + b = w^T x + b = [x^T, 1] \theta \leftarrow \hat{y}$$

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

Para  $w_j$ :

$$\frac{\partial E_{in}(\theta)}{\partial w_j} = \frac{1}{2M} \sum_{i=1}^M -2 (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}$$

$$\frac{\partial E_{in}(\theta)}{\partial b} = \frac{1}{2M} \sum_{i=1}^M -2 (y^{(i)} - \hat{y}^{(i)})$$

$$E_{in}(\theta_k) = -\frac{1}{M} \begin{bmatrix} \sum_{i=1}^M (y^{(i)} - \hat{y}^{(i)}) x_1 \\ \sum_{i=1}^M (y^{(i)} - \hat{y}^{(i)}) x_2 \\ \vdots \\ \sum_{i=1}^M (y^{(i)} - \hat{y}^{(i)}) x_n \end{bmatrix} = -\frac{1}{M} \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(M)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(M)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(M)} \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y^{(1)} - \hat{y}^{(1)} \\ y^{(2)} - \hat{y}^{(2)} \\ \vdots \\ y^{(M)} - \hat{y}^{(M)} \end{bmatrix}$$

$x_e^T$        $y - x_e \theta$

## Programa 7.10.1

```
def Des_gr_F1(x, y, w_0, b_0, lr, max_epochs, e_tol):  
    M = x.shape[0]  
    w = w_0.copy()  
    b = b_0.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
```

$$h_{\theta} = \text{Sign}(w_1 x_1 + w_2 x_2 + b)$$

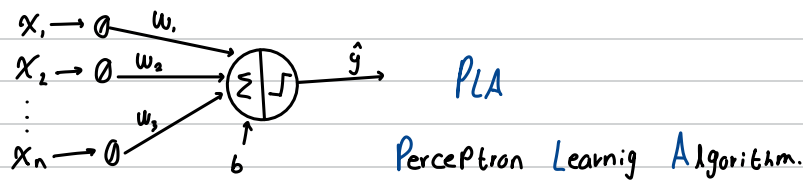
$$x \in \mathbb{R}^n$$

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

$$H = \{h_{\theta} \mid h_{\theta} = \text{sign}(w^T x + b), w \in \mathbb{R}^n, b \in \mathbb{R}, \theta = (w, b)\}$$

$$\text{loss}(y, \hat{y}) = \begin{cases} 1 & \text{si } \hat{y} \neq y \\ 0 & \text{en otro caso} \end{cases} = \max(-y\hat{y}, 0)$$

$$E_{\text{in}}(w, b) = \frac{1}{M} \sum_{i=1}^M \text{loss}(y^{(i)}, \text{sign}(w^{(T)} x^{(i)} + b))$$

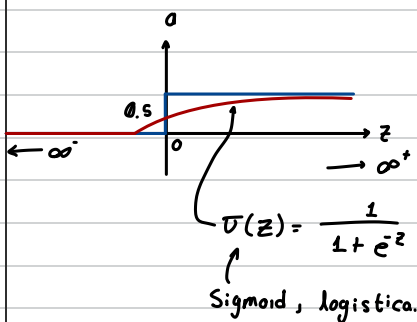


$$a = \Pr(y = 1 \mid X = x; \theta)$$

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

$$a = \int (w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b)$$

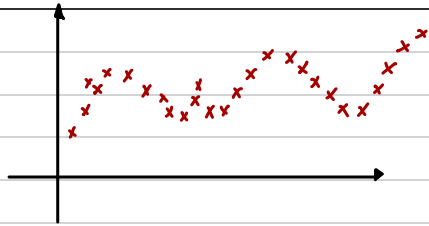
$$z = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b = x^T w + b = x_e^T \theta$$



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

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

$$= -a^{(i)} \log(\hat{a}^{(i)}) - (1 - a^{(i)}) \log(1 - \hat{a}^{(i)})$$



$$\phi: \mathbb{R}^n \rightarrow \mathbb{R}^{n'}$$

$$x \in \mathbb{R}^n$$

$$x' = \phi(x) = (x, x^2, x^3, x^4)$$

$$x' = \phi(x) = (\phi_1(x), \phi_2(x), \phi_3(x), \dots, \phi_{n'}(x))$$

### Expansion Polynomial

$$x = (x_1, x_2)$$

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

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

$$\frac{\partial E_{in}(w, b)}{\partial w_j} = \frac{1}{M} \sum_{i=1}^M \left( \frac{a^{(i)}}{\hat{a}^{(i)}} - \frac{1-a^{(i)}}{1-\hat{a}^{(i)}} \right) \hat{a}^{(i)} (1-\hat{a}^{(i)}) x_j^{(i)}$$

$$\frac{\partial E_{in}(w, b)}{\partial w_j} = \frac{\partial}{\partial w_j} \frac{1}{M} \sum_{i=1}^M [-a^{(i)} \ln |\hat{a}^{(i)}| - (1-a^{(i)}) \ln (1-\hat{a}^{(i)})]$$

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

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

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

such that

$$\sum_{j=1}^n w_j^2 \leq C$$

