# Primer capítulo del Bishop.

## Regresión

- Probabilidad
- Decisión
- Teoría de la información.

- Entrenamiento
- Prueba
- generalización.

ML ---- semi-supervisado

- Supervisado
  - ↳ regresión ---->
  - ↳ Clasificación --->
- No Supervisado
- Reinforcement Learning

**Regresión**

$$(x_n, t_n)$$

$t_n \in \mathbb{R}$

**Clasificación**

$$(x_n, c_n)$$

$c_n \in \mathbb{N}$ clase.

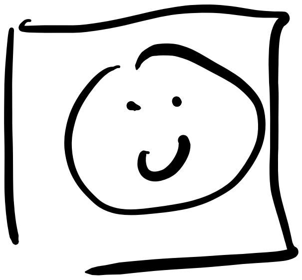
## Ejemplo de regresión:

$$x_n = \begin{bmatrix} \text{edad} \\ \text{peso} \\ \text{altura} \end{bmatrix} \qquad x_n = \begin{bmatrix} \text{densidad} \\ \text{peso m} \\ \\ \end{bmatrix}$$

$t_n$ = glucosa en la sangre.

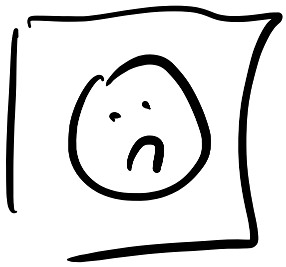
$t_n$ = temperatura de fusión

## Ejemplo de clasificación:



$\mapsto$ "feliz"

$$C = \begin{bmatrix} \text{"feliz"} \\ \text{"triste"} \\ \text{"con hambre"} \end{bmatrix}$$



$\mapsto$ "triste"

---

$n \mapsto$ "primo"

"no primo"

$$\boxed{\text{arte}} \longmapsto \begin{bmatrix} \text{dadaismo} \\ \text{impresionismo} \\ \vdots \end{bmatrix}$$

ML ----- semi-supervisado

- Supervisado
  - ↳ regresión ----→
  - ↳ clasificación ---→
- No Supervisado
- Reinforcement Learning

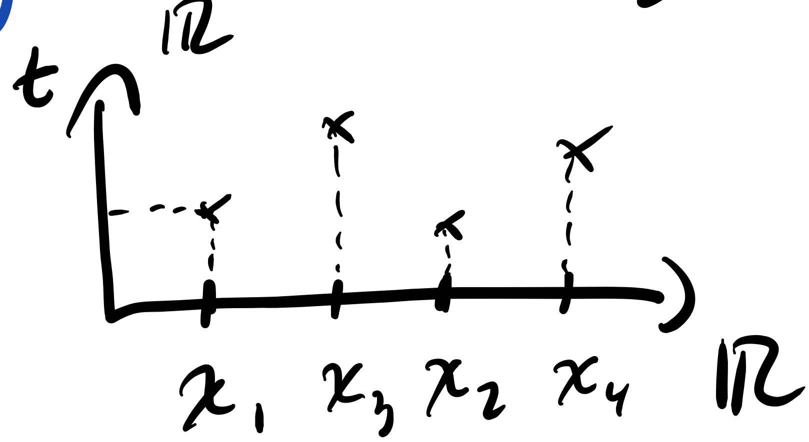
$$D = \{(x_n, t_n)\}$$

↳ "la verdad"

$$D = \{(x_n)\}$$

# Regresión

$D = \{(x_n, t_n)\}$




Un modelo ...

$$\underbrace{y(x; w_0, w_1) = w_0 + w_1 x}_{\text{regresión lineal}}$$

$$\underbrace{y(x; \vec{w}) = \sum_i w_i x^i}_{\text{regresión polinomial.}}$$

$$y(x; w_0, w_1) = \sin(w_0) + \tan(w_1 x)$$

$$y(x; w, a) = a\cos(wx)$$

$$y(x; W_1, W_2) = W_2 \phi(W_1 \phi(x))$$

"funciones de activación"

— /// —

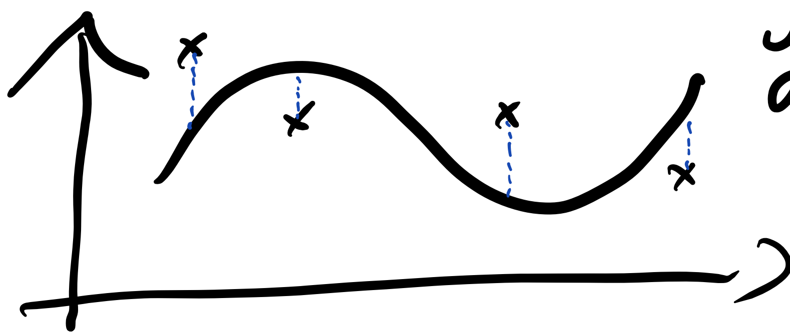
$$y(x; \vec{w}) = \sum_i w_i x^i$$

(1) Tenemos

$$D = \{(x_n, t_n)\}$$

(2) Definimos $y(x; \vec{w})$ y

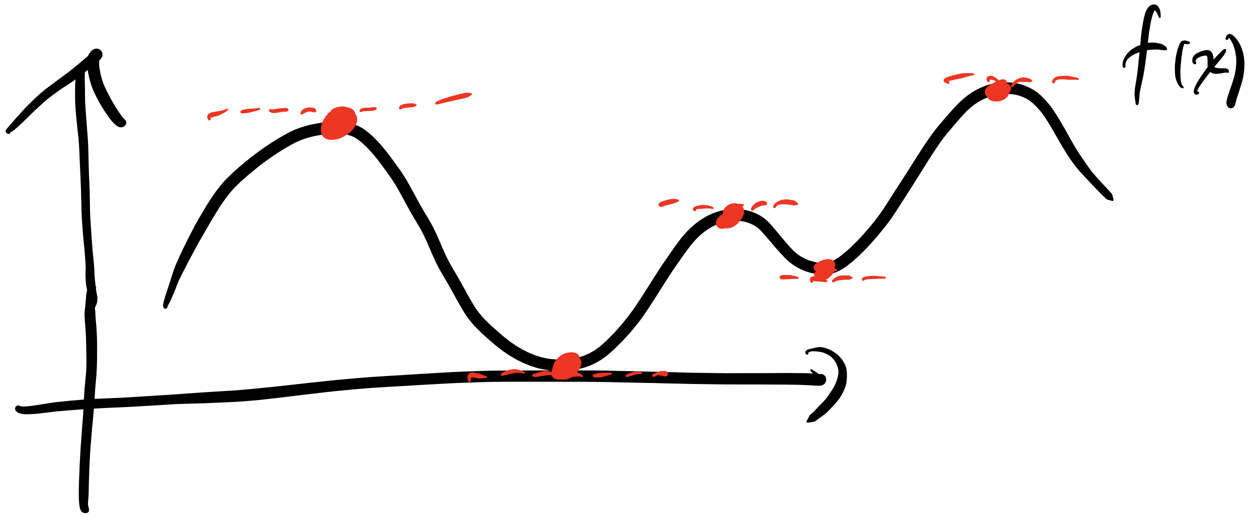
$$E(\vec{w}) = \frac{1}{2} \sum_n \left( y(x_n; \vec{w}) - t_n \right)^2$$



$y(x_n; \vec{w})$

Queremos minimizar $E(\vec{w})$ ...

Recordemos optimización.



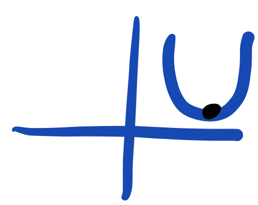

Encontrar $x$ t.q. $f'(x) = 0$

dependiendo de

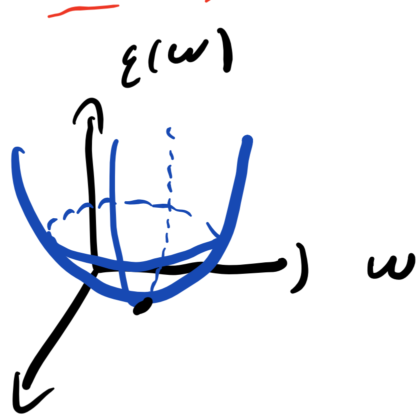
$f''(x) < 0$ max.

$f''(x) > 0$ min.

Convexas



$E(w)$



$$\mathcal{E}(\vec{w})$$

$$f'(x) = 0$$

$$\Updownarrow$$

$$\nabla_w \mathcal{E}(\vec{w}) = 0$$

$$\mathcal{E}(\vec{w}) = \frac{1}{2}\sum_n \left(y(x_n; \vec{w}) - t_n\right)^2$$

---

$$\nabla_{\vec{w}} \mathcal{E} = \left(\frac{\partial \mathcal{E}(\vec{w})}{\partial w_0}, \ \dots, \ \frac{\partial \mathcal{E}(\vec{w})}{\partial w_T}\right)$$

$$w_i = w_1$$

$$\mathcal{E}(\vec{w}) = \frac{1}{2}\left(\left(w_0 + w_1 x_n + \dots + w_T x_n^T\right) - t_n\right)^2$$

$$\frac{\partial \mathcal{E}}{\partial w_i} = \left(y(x_n; \vec{w}) - t_n\right) \cdot x_n^i$$

$$\frac{\partial y(x_n; \vec{w})}{\partial w_i} = \frac{\partial\left(w_0 + w_1 x_n^1 + \dots + w_T x_n^T\right)}{\partial w_i}$$

$$= \frac{\partial w_0}{\partial w_i} + \dots + \frac{\partial (w_i x_n^i)}{\partial w_i} + \dots$$

$$= x_n^i$$

——— /// ———

$$\nabla_w \mathcal{E} = 0$$

$$\frac{\partial \mathcal{E}}{\partial w_i} = 0$$

$$\left( y(x_n; \vec{w}) - t_n \right) \cdot x_n^i = 0$$

Tarea: 1. ¿Cuáles son los $\vec{w}$ que minimizan $\mathcal{E}(\vec{w})$?

2. Programar

```
def pesos([x_1, ..., x_N], [t_1, ..., t_N], T)
    ...
    return [w_0, ..., w_T]
```

3. Pensar en problemas de regresión o clasificación que te parezcan interesantes.