Regresión probabilística.

Sesión anterior:

- Sobreajuste.
- Distribuciones de probabilidad.

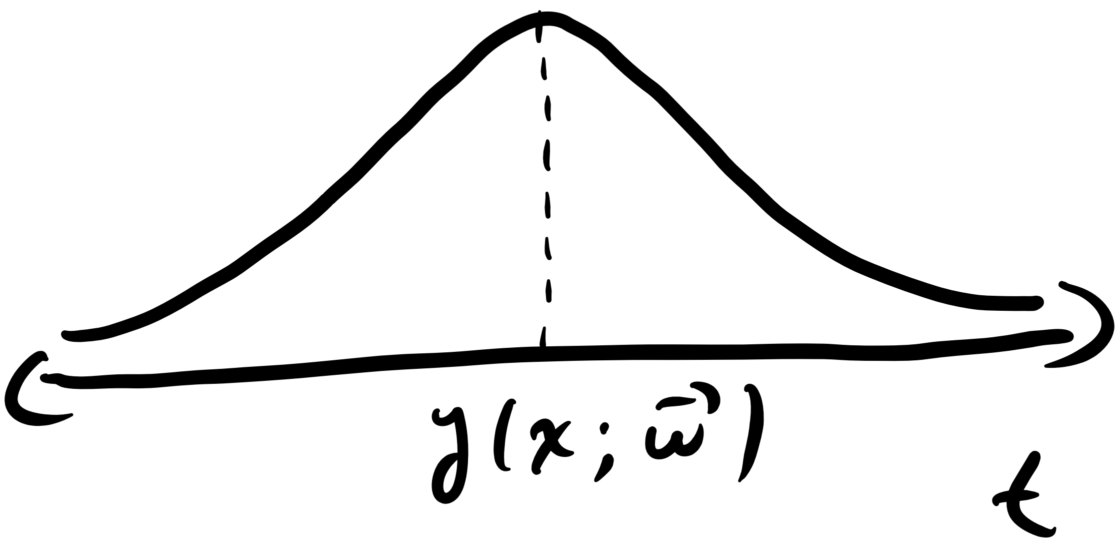
Tarea: re leer 1.2.5 del Bishop.

Hoy:

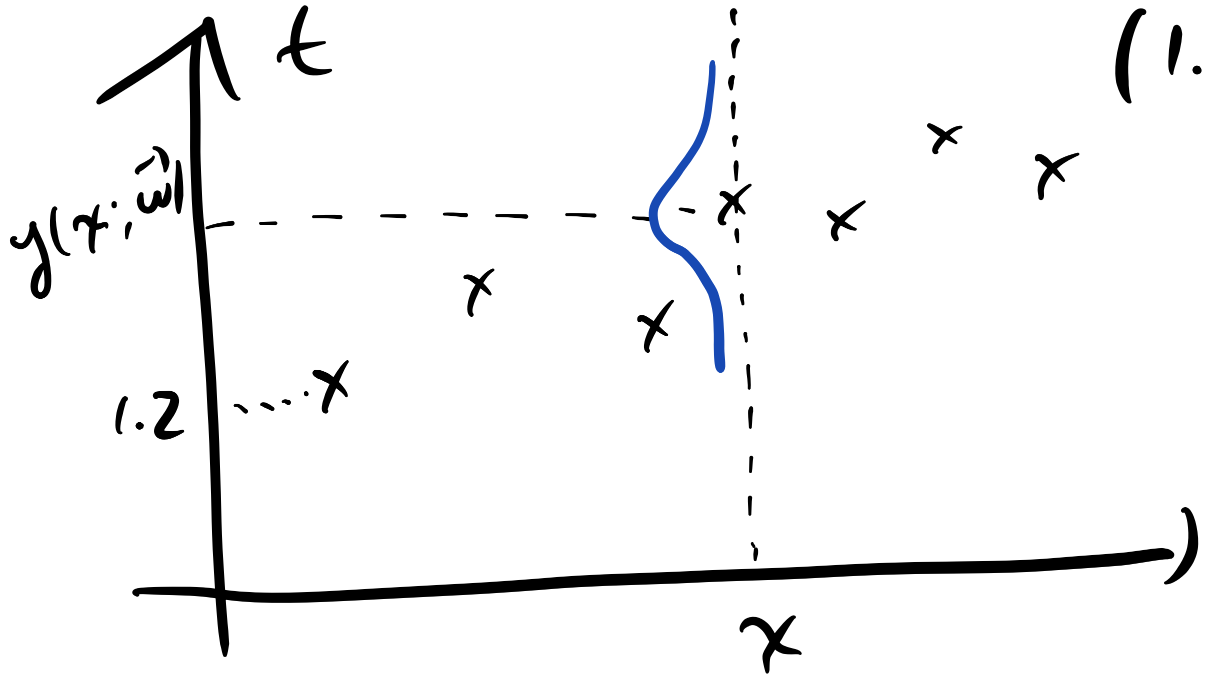
- Regresión probabilística
- código.

$$p(t \mid x, \vec{w}, \beta)$$

$$= \mathcal{N}(t \mid y(x;\vec{w}), \beta^{-1})$$



$$\{(x_n, t_n)\}_{n=1}^{N}$$



(1.60)

(1.60) → (1.61)

$$p(t \mid x, w, \beta)$$

$$p(t_1, \ldots, t_N \mid x_1, \ldots, x_N, \vec{w}, \beta)$$
$$= \prod_n \mathcal{N}(t_n \mid \vec{x}, \vec{w}, \beta)$$

$$p(a,b) \stackrel{?}{=} p(a)p(b)$$

si a y b son independientes

$t_1, \ldots, t_n$ son independientes

t_n son i.i.d

independientes
e
idénticamente
distribuidos.

1. $y(x;\vec{w})$ se toma como la media de la dist.

$$p(t \mid x, \vec{w}, \beta)$$

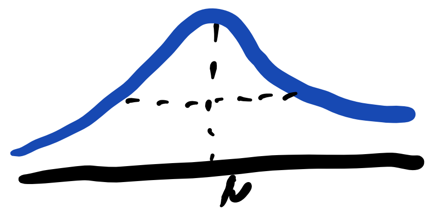
2. Si tenemos una base de datos

$$\{(x_n, t_n)\}_{n=1}^{N}$$

Suponemos que los t_n son i.i.d.

Verosimilitud
(likelihood)

↳ ≈ probable



$t \sim \mathcal{N}(t | \mu, \sigma)$

La función de verosim.

$$p(\vec{t} \,|\, \vec{x}, \vec{w}, \beta)$$

↳ es una función de $\vec{t}$

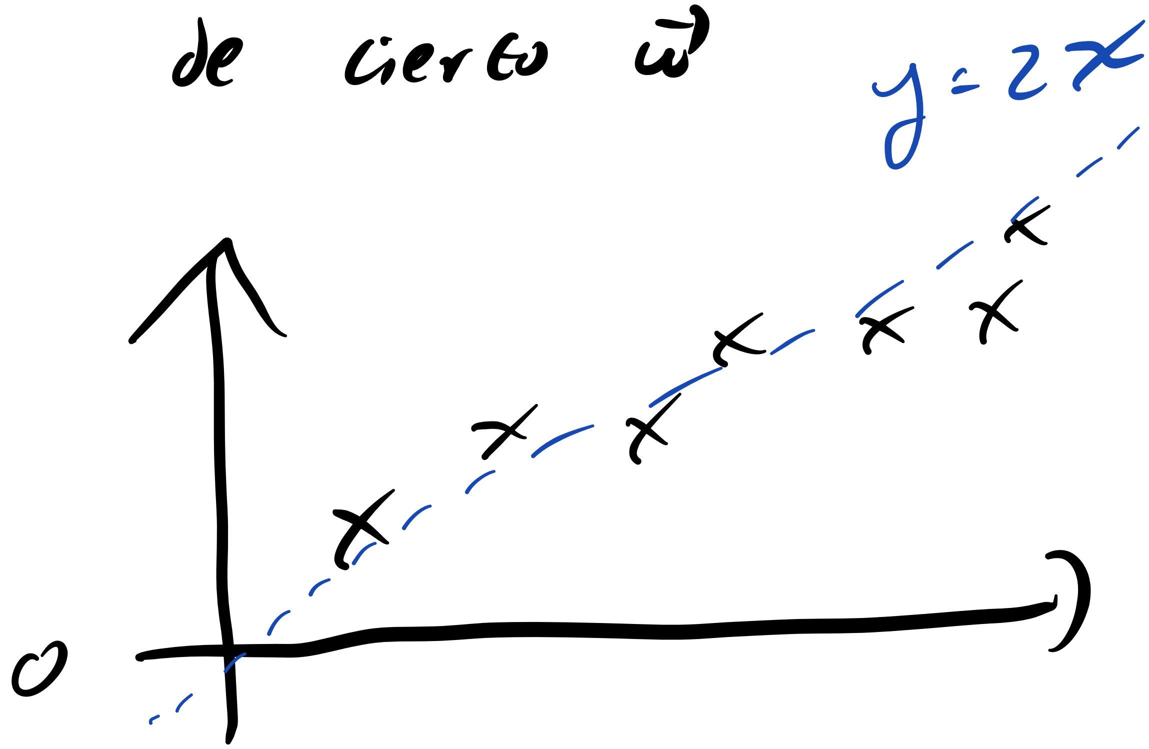
↳ nos dice "qué tan probable" es $\vec{t}$ dados $\vec{x}$, $\vec{w}$ y β

Definimos

$$L(\vec{w}) = p(\vec{t} \,|\, \vec{x}, \vec{w}, \beta)$$

↳ es una función de $\vec{\omega}$

↳ nos dice "qué tan prob. es $\vec{t}$" en términos de cierto $\vec{\omega}$



¿Cuál es el $\vec{\omega}$ que hace $\vec{t}$ más probable?

$$\vec{\omega} = (0, 2)$$

$L(\vec{w})$ se llama la "verosimilitud", y queremos maximizarla.

↳ Principio de máxima verosimilitud.

$$L(\vec{w}) = p(\vec{t} \mid \vec{x}, \vec{w}, \beta)$$

$$= \prod_n \mathcal{N}(t_n \mid y(x_n, \vec{w}), \beta)$$

$$= \prod_n \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma}(y_n - t_n)^2\right)$$

Un truco estándar

$$L(\vec{w}) \quad , \quad \log(L(\vec{w}))$$

$$\ell(\vec{w}) = \sum_n \left[\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \left(-\frac{1}{2\sigma}(y_n - t_n)^2\right)\right]$$

$$= -N\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma}\sum_n (y_n - t_n)^2$$

maximizar $l(\vec{w})$

$\Leftrightarrow$

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

///

Si suponemos

$p(t \mid x, w, \beta) = \mathcal{N}(\cdots)$,

entonces tenemos que

minimizar

$$\mathcal{E}(\vec{w}) = \frac{1}{2} \sum_n (y_n - t_n)^2$$