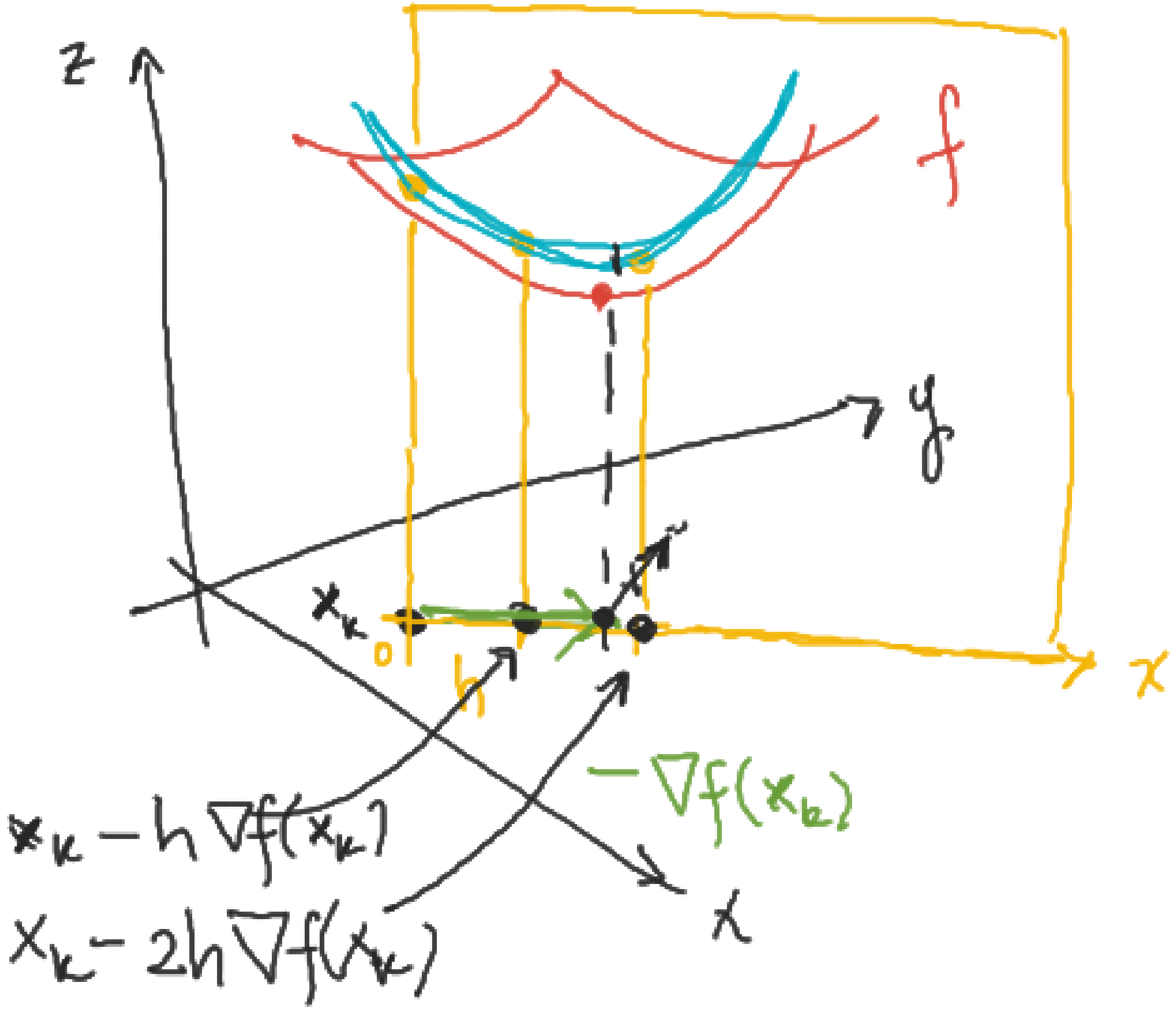
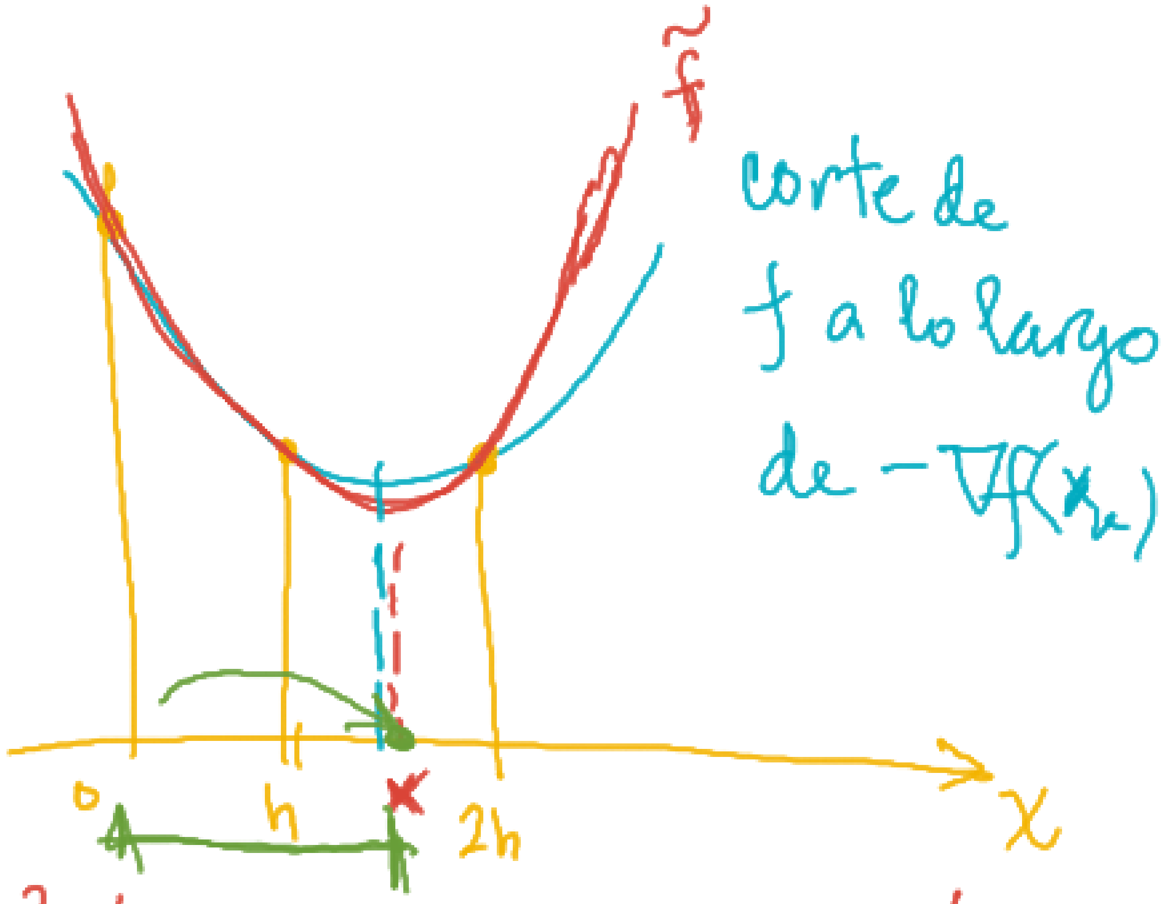
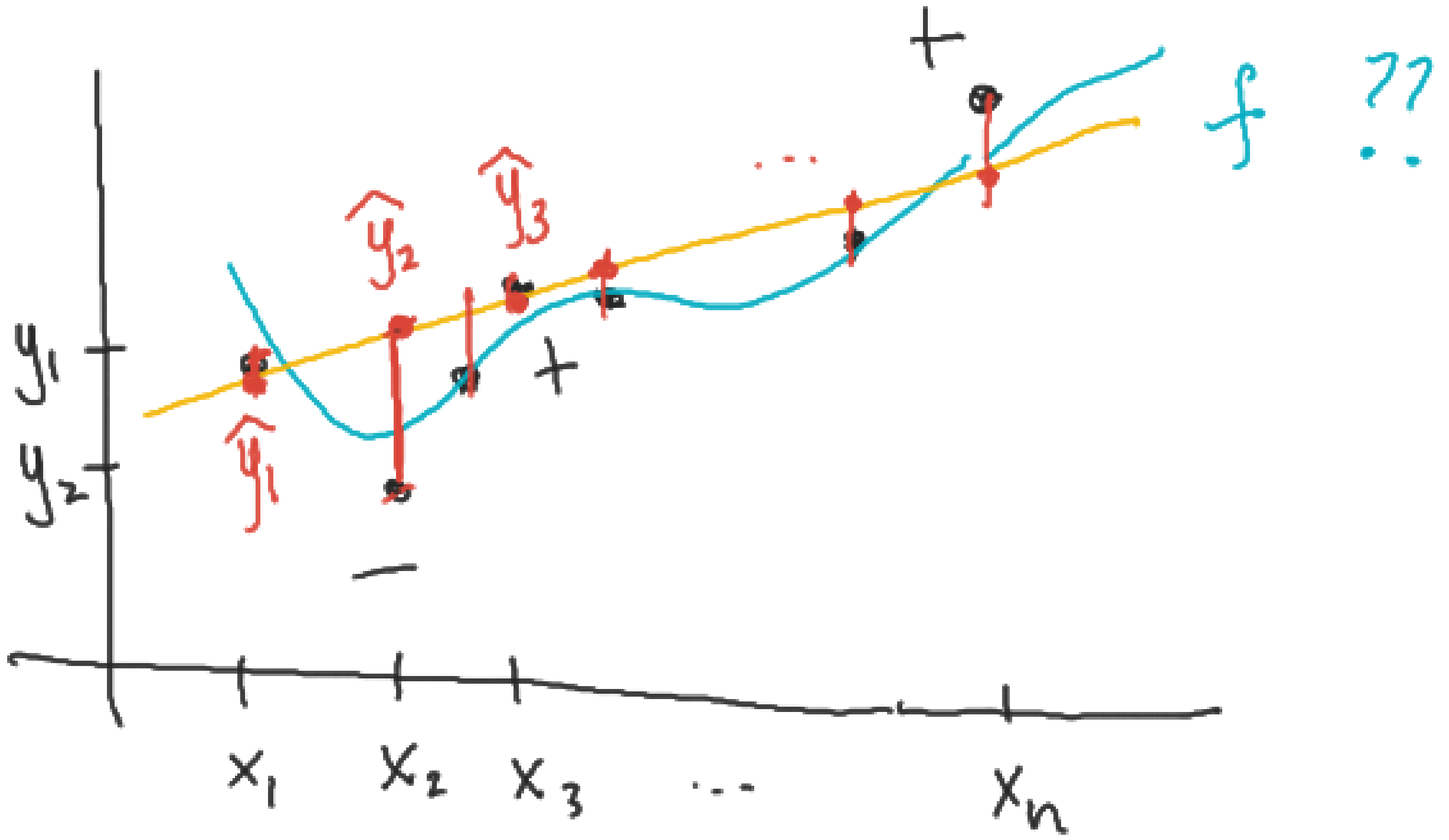




$$y = ax^2 + bx + c$$
$$y' = 2ax + b = 0 \Rightarrow x^* = -\frac{b}{2a}$$

Tamaño de paso es $\boxed{\alpha_k = x^*}$

Datos (x_1, y_1)
(x_2, y_2)
$\vdots$
(x_n, y_n)



$y = a_1 x + a_0$ $\quad$ $\hat{y}_i = y(x_i) = a_1 x_i + a_0$

error en la i-ésima observación $\quad y_i - \hat{y}_i$

función de error $\quad \boxed{E(a_1, a_0) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$ $\quad$ suma de errores cuadraticos (SSE)

Resolver $\underset{(a_1, a_0) \in \mathbb{R}^2}{\arg\min} E(a_1, a_0)$.

$x_0 = (a_{10}, a_{00})$

$x_k = x_{k-1} - \alpha \nabla E(x_k)$

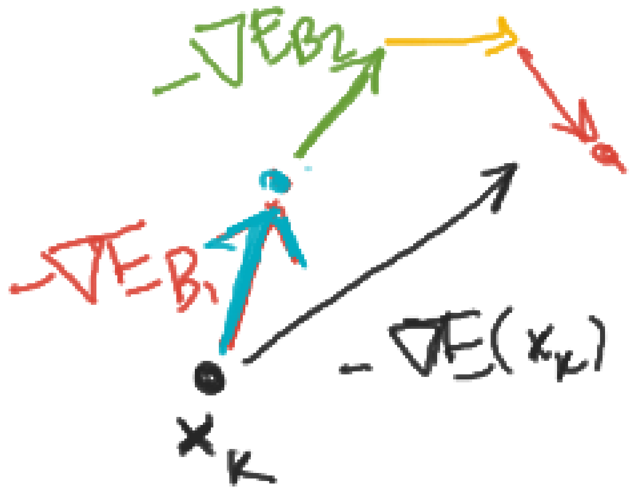
modelo 2: $y = \underline{a_3 x^3 + a_2 x^2 + a_1 x + a_0}$

(x_1, y_1),
(x_2, y_2)
$\vdots$
(x_n, y_n)

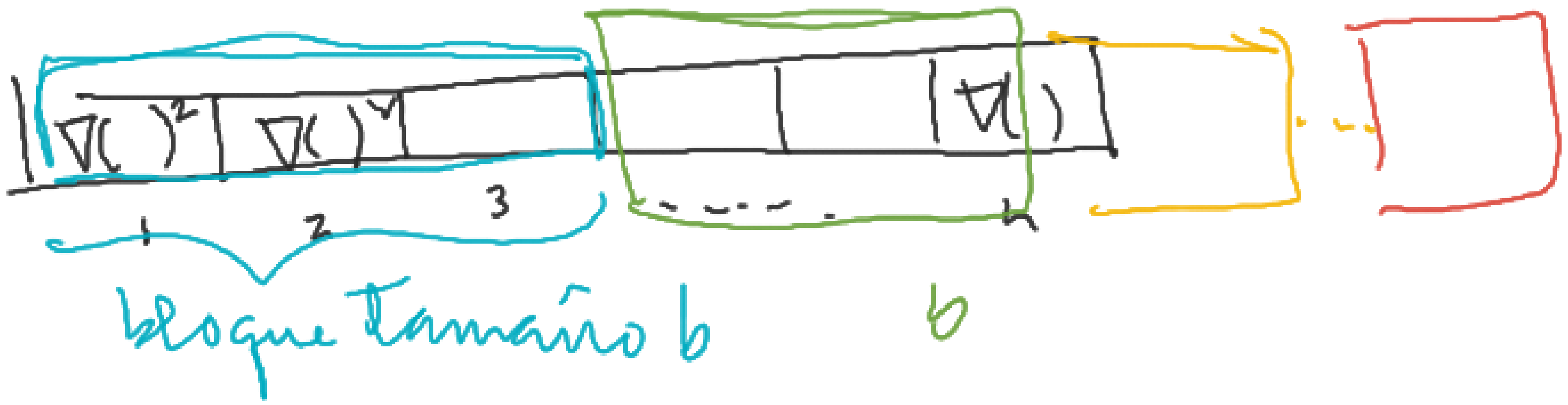
función objetivos

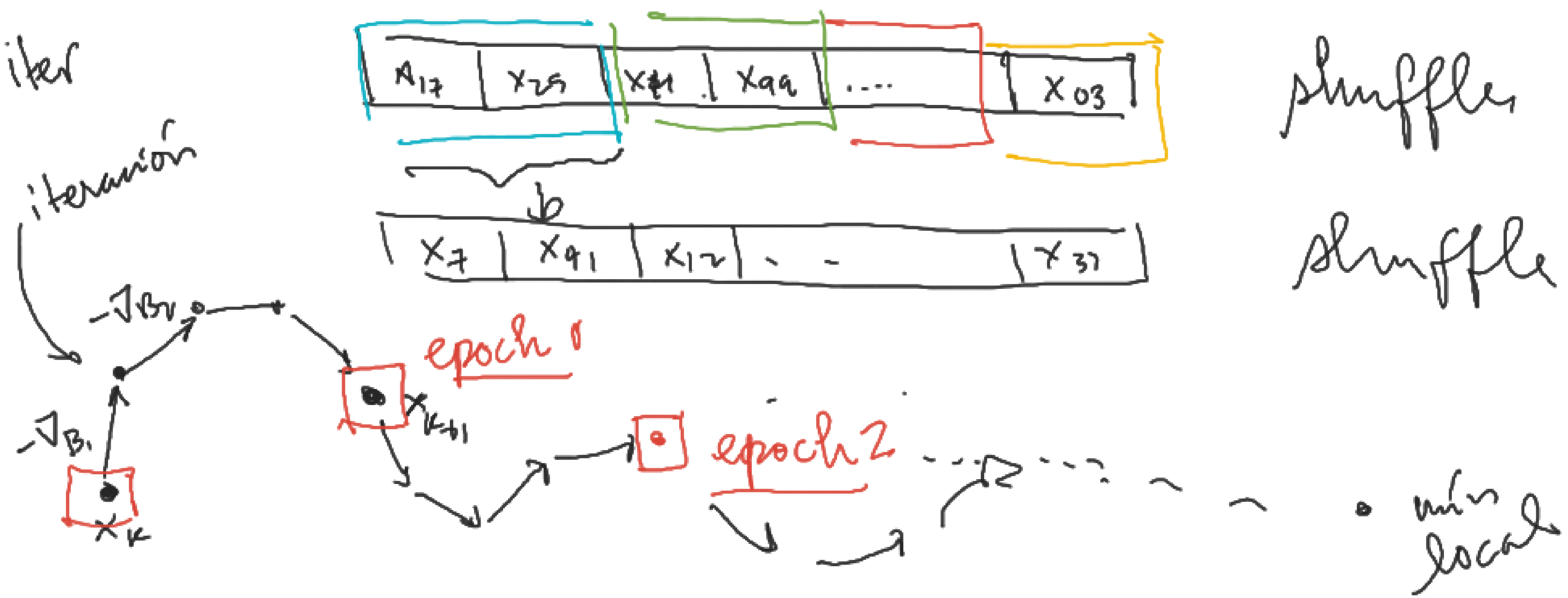
$$E(a_0, a_1, a_2, a_3) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\underline{\nabla_a E(a_0, a_1, a_2, a_3)} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\nabla_a (y_i - \hat{y}_i)^2}_{\text{gradientes en cada obs.}}$$



$S = 0$
for $i = 1$ to n
$\quad S = S + \nabla_a (y_i - \hat{y}_i)^2$





batch

batch size

Métodos de gradiente

Optimización de funciones { diferenciables
convexas

¿Qué hacer cuando f

- f no es diferenciable
- descenso gradiente no funciona (converge a un mínimo local, no converge)
- regions difíciles
- regiones con muchos mínimos