Regresión Lineal

Métodos de aprendizaje para máquinas

Universidad del Valle
2024

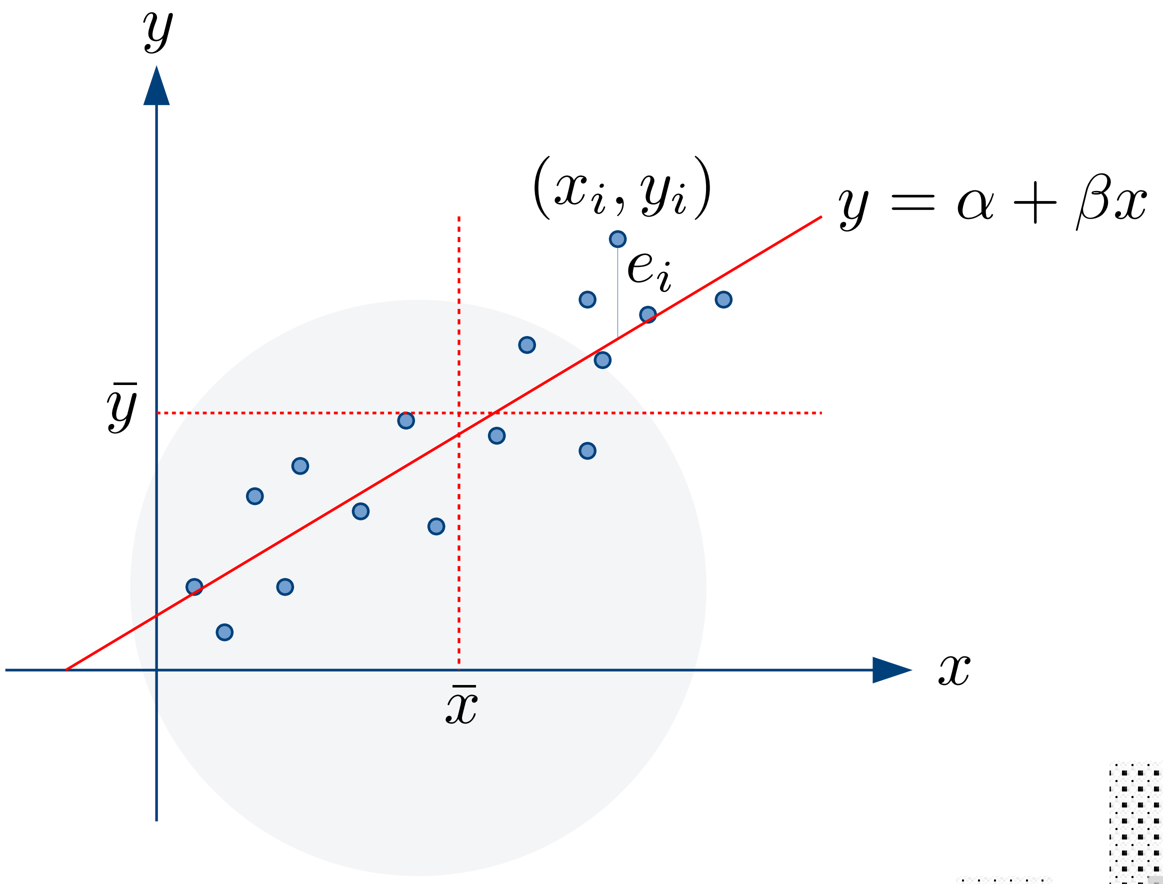
Regresión Lineal

- Predicción: Se emplea para ajustar un modelo predictivo conforme a un conjunto de datos $\{(x_k, y_k)\}$. Después de desarrollar el modelo, si se genera un valor adicional de x sin su pareja y, el modelo ajustado puede ser usado para predecir el valor de y.

- Dada una variable y y un conjunto de variables $x_1, \ldots, x_p$, que pueden estar relacionadas a y, el análisis de regresión lineal puede aplicarse para cuantificar la correlación entre y y las x_j.

- En la mayoría de los casos se emplea **Mínimos Cuadrados**.

Regresión Lineal



Mínimos Cuadrados

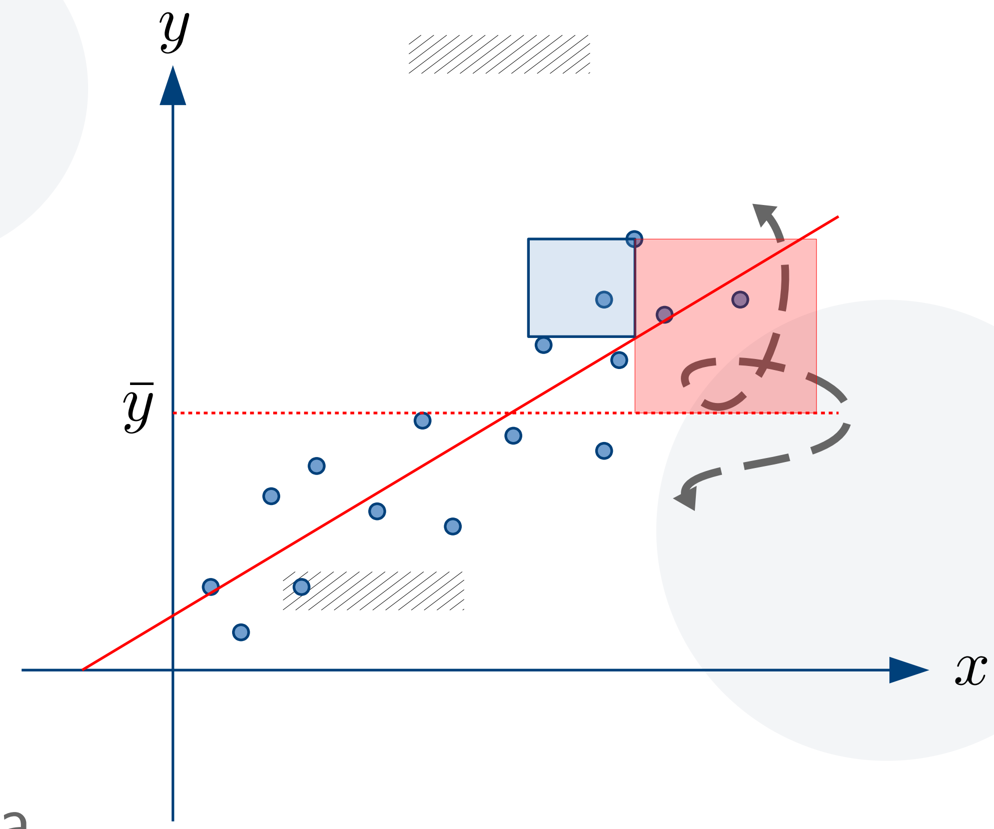
- El objetivo es minimizar la suma de los cuadrados de las diferencias entre las respuestas observadas y las predichas por una función lineal.
- $y = \alpha + \beta x$
- Los datos estan dados por $y_i = \alpha + \beta x_i + \epsilon_i$ y

$$\beta = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \alpha = \bar{y} - \beta \bar{x}$$

- Siendo $\bar{x}$ y $\bar{y}$ los promedios respectivos de los datos.

Coeficiente de Determinación

- Se denota por R^2.
- Es la proporción de la varianza de la variable dependiente que es predecible con la variable independiente.
- $$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$
- Mejor es la regresión cuanto más cerca esté de la unidad.



Coeficiente de Correlación

- La correlación r se define como la razón entre la covarianza y el producto de las desviaciones estándar de las variables dependiente e independiente.
- Si $r = 1$ indica una proporción directa creciente perfecta entre las variables, $r = -1$ indica una proporción directa decreciente perfecta (anticorrelación).
- Valores intermedios no nulos implican alguna dependencia lineal
- Cuando r se acerca a cero, la relación no es lineal, si la hay.
- $$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Ajuste no lineal

- Exponencial

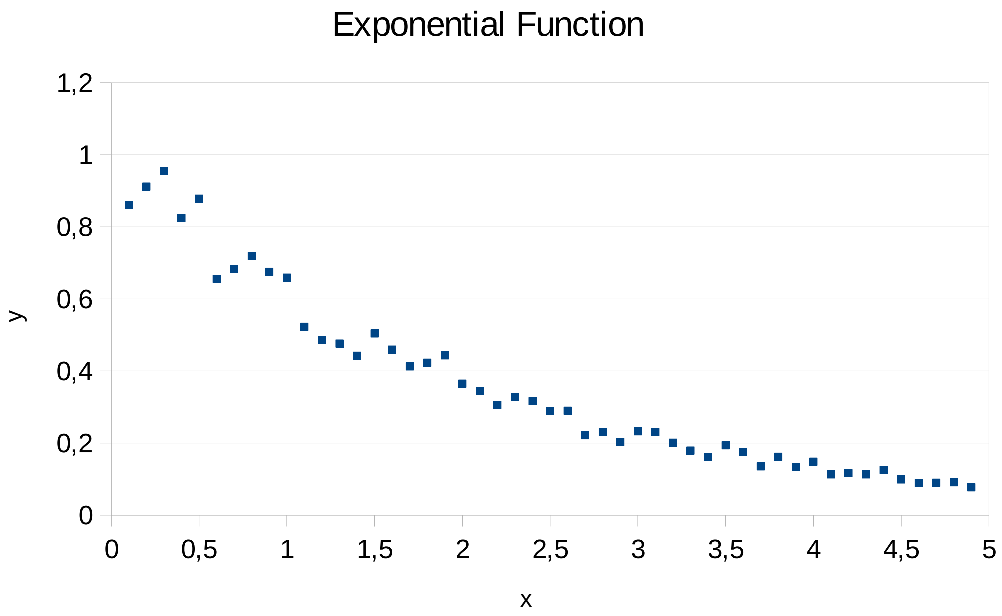
$$y = Ae^{Bx} \to \log y = \alpha x + \beta$$

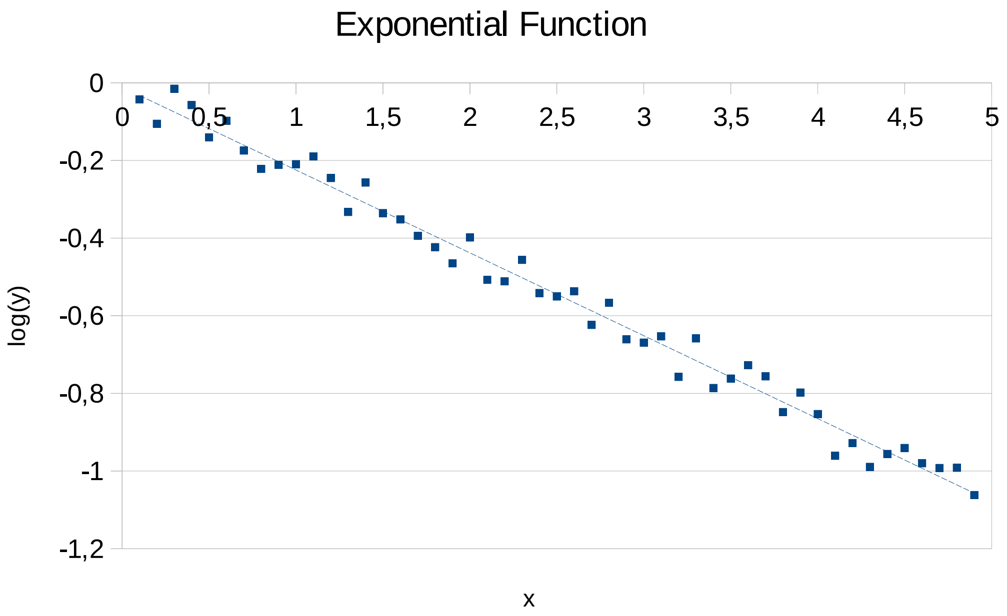
- Potencial

$$y = Ax^{B} \to \log y = \alpha \log x + \beta$$

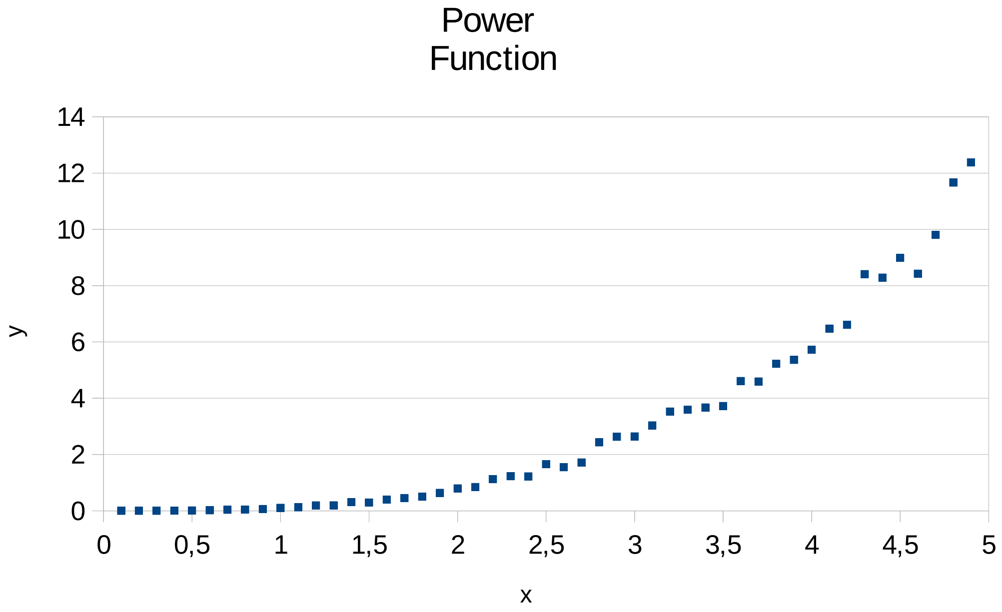
- Combinado

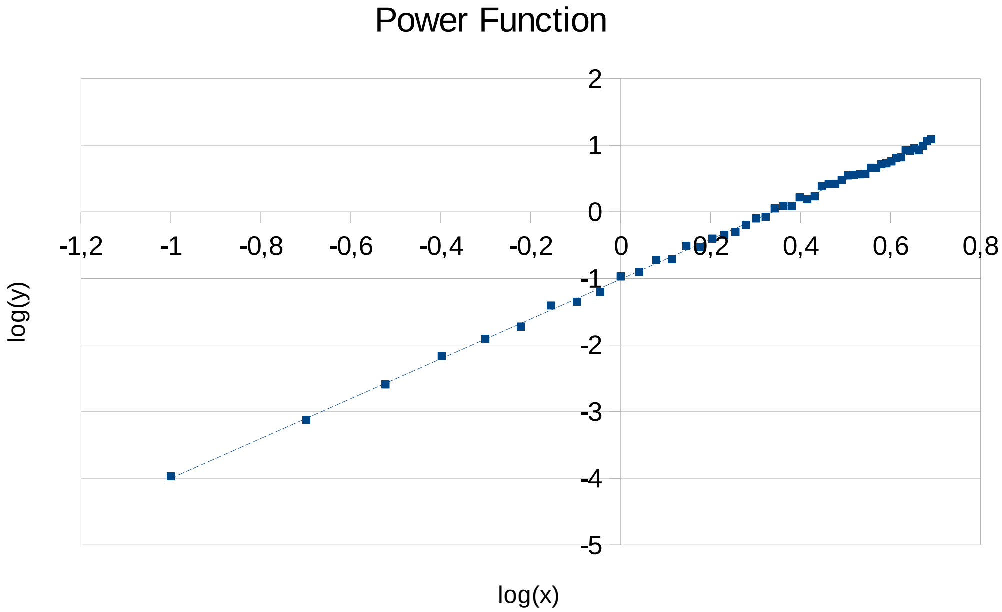
$$y = Axe^{Bx} \to \log\left(\frac{y}{x}\right) = \alpha x + \beta$$



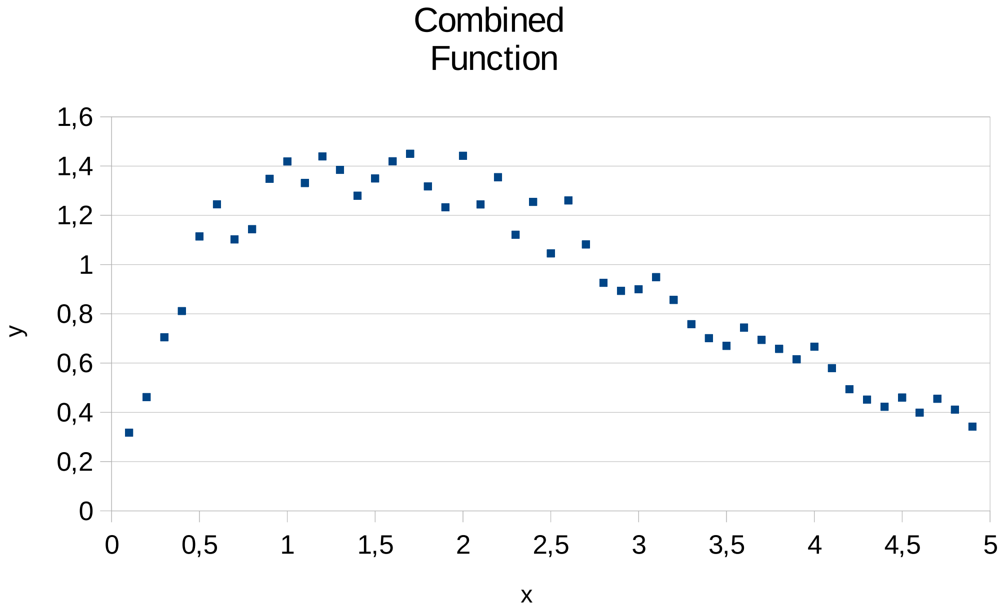


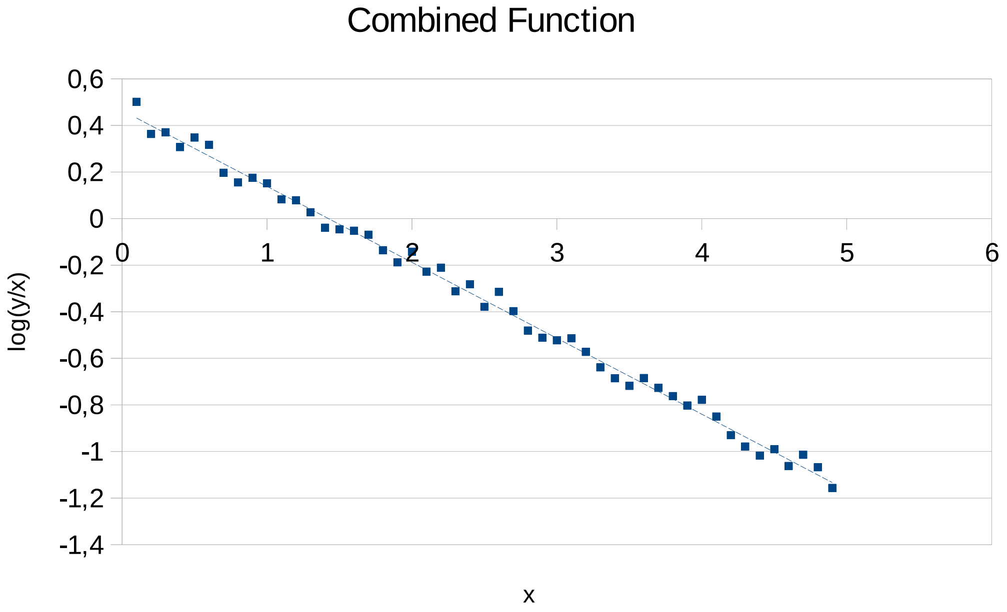
$$y = Ae^{Bx} \rightarrow \log y = \alpha x + \beta$$





$$y = Ax^B \rightarrow \log y = \alpha \log x + \beta$$





$$y = Axe^{Bx} \rightarrow \log\left(\frac{y}{x}\right) = \alpha x + \beta$$

“You can do linear regression without thinking about whether the phenomenon you’re modeling is actually close to linear. But you shouldn’t.”

- Jordan Ellenberg

Regresión no lineal

Métodos de aprendizaje para máquinas

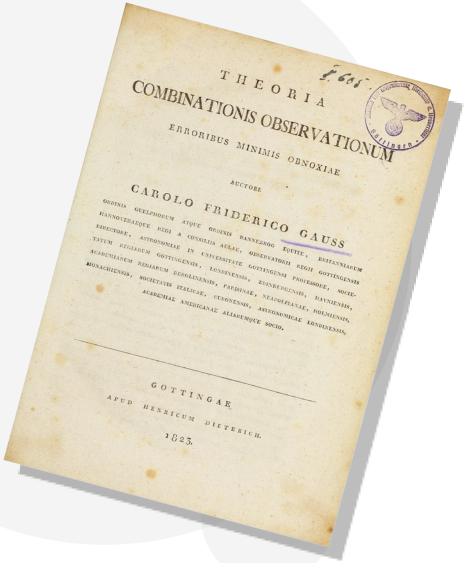
Universidad del Valle
Abril de 2023

Métodos analíticos

- Función lineal
- Función exponencal
- Funciones algebráicas

NOUVELLES MÉTHODES
POUR LA DÉTERMINATION
DES
ORBITES DES COMÈTES;
PAR A. M. LEGENDRE,
Membre de l'Institut et de la Légion d'honneur, de la Société royale de Londres, &c.

A PARIS,
Chez Firmin Didot, Libraire pour les Mathématiques, la Marine, l'Architecture, et les Éditions stéréotypes, rue de Thionville, n° 116.
AN XIII — 1805.



THEORIA
COMBINATIONIS OBSERVATIONUM
ERRORIBUS MINIMIS OBNOXIAE
AUCTORE
CAROLO FRIDERICO GAUSS
ORDINIS GUELPHORUM ATQUE ORDINIS DANNEBROG EQUITE, BRITANNIARUM HANNOVERAEQUE REGI A CONSILIIS AULAE, OBSERVATORII REGII GOTTINGENSIS DIRECTORE, ASTRONOMIAE IN UNIVERSITATE GOTTINGENSI PROFESSORE, SOCIETATUM REGIARUM GOTTINGENSIS, LONDINENSIS, EDINBURGENSIS, HAVNIENSIS, ACADEMIARUM REGIARUM BEROLINENSIS, PARISINAE, NEAPOLITANAE, HOLMIENSIS, MONACHIENSIS, SOCIETATIS ITALICAE, CURONENSIS, ASTRONOMICAE LONDINENSIS, ACADEMIAE AMERICANAE ALIARUMQUE SOCIO.

GOTTINGAE
APUD HENRICUM DIETERICH.
1823.

Funciones no lineales

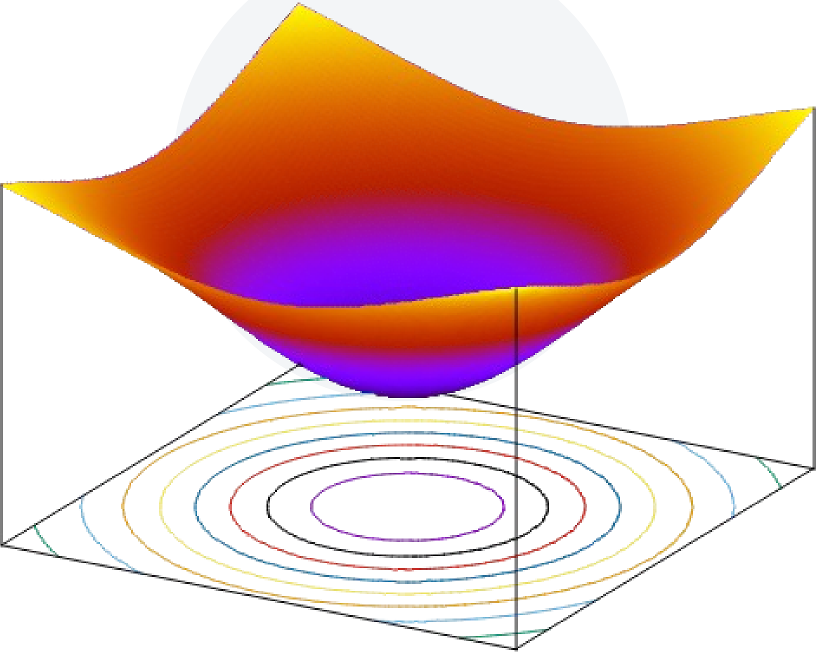
$$f(x) = \beta_1 x e^{-\beta_2 x} \quad \rightarrow \quad f(x) \equiv f(x, \vec{\beta})$$

$$E_2(\vec{\beta}) = \sum_{k=0}^{n} \left[f(x_k, \vec{\beta}) - y_k \right]^2 \quad \rightarrow \quad \frac{\partial E_2}{\partial \beta_i} = 0.$$

En general, el ajuste a funciones no lineales, conlleva encontrar la solución de un sistema de ecuaciones no lineal y, por tanto, cada problema debería tratarse de forma particular.

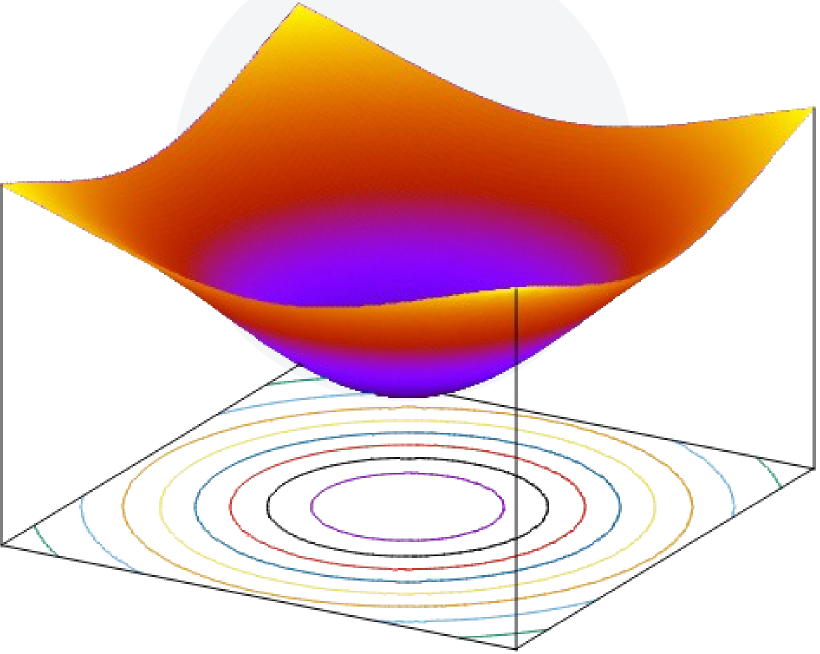
$$\sum_{k=1}^{n} \left[f(x_k, \vec{b}) - y_k \right] \frac{\partial f}{\partial \beta_i} = 0.$$

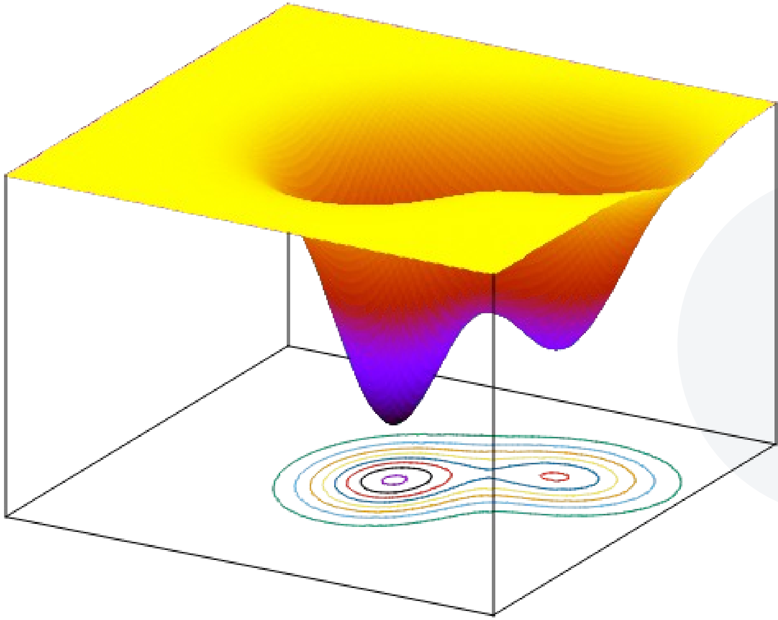
Para encontrar la solución, se deben emplear métodos de minimización.



El mínimo es breve de encontrar, simplemente desciendo por el gradiente.

$$\vec{\nabla}_{\beta} f(\vec{\beta}) = 0$$





- ¿Qué es el gradiente?
- ¿Qué hago en un mínimo local?
- ¿Qué hago en las regiones planas?

El mínimo es breve de encontrar, simplemente desciendo por el gradiente.

$$\vec{\nabla}_\beta f(\vec{\beta}) = 0$$

Descender por el gradiente

1. Seleccionar un punto inicial.

Es necesario contar con un amplio conjunto aleatorio de condiciones iniciales. Algunas no convergerán al mínimo global.

2. Calcular la dirección del gradiente

El gradiente indica el valor de la pendiente más fuerte en ascenso. Tomamos la dirección opuesta.

3. Calcular el paso a dar.

Si tomamos un paso demasiado grande, puede que nos pasemos de nuestro objetivo. Si es pequeño, tendremos que seguir en la misma dirección.

4. Criterio de parada

Si el gradiente es muy chico, podemos estar en una meseta o en un mínimo local.

“

Ejemplo:

$$f(x, y) = x^2 + 3y^2$$

”

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

El gradiente indica el valor de la pendiente más fuerte en ascenso. Tomamos la dirección opuesta.

3. Calcular el paso a dar.

Si tomamos un paso demasiado grande, puede que nos pasemos de nuestro objetivo. Si es pequeño, tendremos que seguir en la misma dirección.

4. Criterio de parada

Si el gradiente es muy chico, podemos estar en una meseta o en un mínimo local.

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla} f(x,y) = -2x\hat{e}_x - 6y\hat{e}_y$$

3. Calcular el paso a dar.

Si tomamos un paso demasiado grande, puede que nos pasemos de nuestro objetivo. Si es pequeño, tendremos que seguir en la misma dirección.

4. Criterio de parada

Si el gradiente es muy chico, podemos estar en una meseta o en un mínimo local.

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla} f(x, y) = -2x\hat{e}_x - 6y\hat{e}_y$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla} f(\vec{r}_k)$$

4. Criterio de parada

Si el gradiente es muy chico, podemos estar en una meseta o en un mínimo local.

Descender por el gradiente

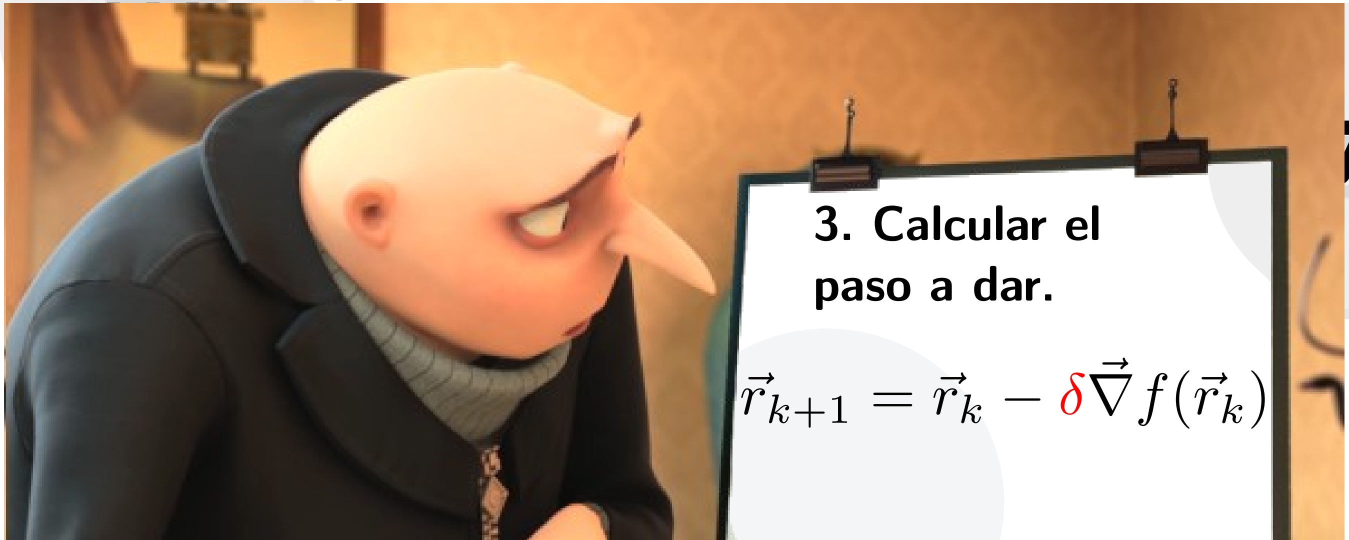
1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla} f(x, y) = -2x\hat{e}_x - 6y\hat{e}_y$$

4. Criterio de parada



3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla} f(\vec{r}_k)$$

Selección de δ

$$F_k(\delta_k) = f\left(\vec{r}_{k+1}(\delta_k)\right)$$

Se construye una nueva función que depende de δ y se optimiza.

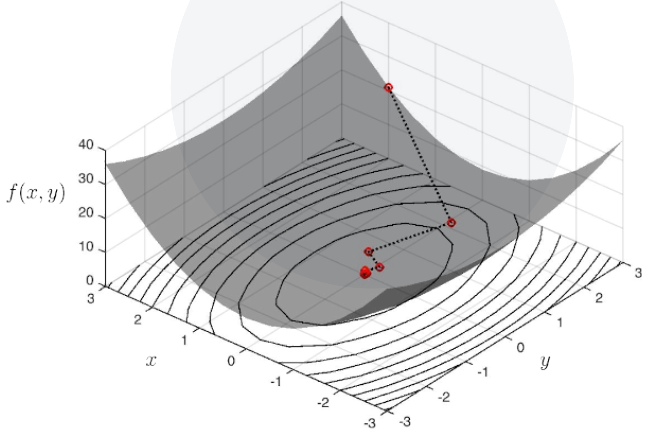
$$0 = \frac{dF_k}{d\delta_k} = -\left[\vec{\nabla} f(\vec{r}_{k+1})\right] \cdot \left[\vec{\nabla} f(\vec{r}_k)\right]$$

El criterio de optimización indica que se avanza en la dirección del gradiente hasta un punto en el que el gradiente sea cero u la dirección sea ortogonal.

$$F_{k+1}(\delta_{k+1})$$

Se actualizan los valores y se reitera el procedimiento. Los valores pueden ser analíticos o resultados numéricos.

Selección de δ



$$F(\delta) = (1-2\delta)^2 x^2 + 3(1-6\delta)^2 y^2 \qquad F'(\delta) = 0 \rightarrow \delta = \frac{x^2+9y^2}{2x^2+54y^2}$$

$$F_k(\delta_k) = f\left(\vec{r}_{k+1}(\delta_k)\right)$$

Se construye una nueva función que depende de δ y se optimiza.

$$0 = \frac{dF_k}{d\delta_k} = -\left[\vec{\nabla} f(\vec{r}_{k+1})\right] \cdot \left[\vec{\nabla} f(\vec{r}_k)\right]$$

El criterio de optimización indica que se avanza en la dirección del gradiente hasta un punto en el que el gradiente sea cero u la dirección sea ortogonal.

$$F_{k+1}(\delta_{k+1})$$

Se actualizan los valores y se reitera el procedimiento. Los valores pueden ser analíticos o resultados numéricos.

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla} f(x,y) = -2x\hat{e}_x - 6y\hat{e}_y$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla} f(\vec{r}_k)$$

4. Criterio de parada

$$|\vec{\nabla} f(x_k)| < \varepsilon$$

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = x_0\hat{e}_x + y_0\hat{e}_y$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla}f(x,y) = -2x\hat{e}_x - 6y\hat{e}_y$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla}f(\vec{r}_k)$$

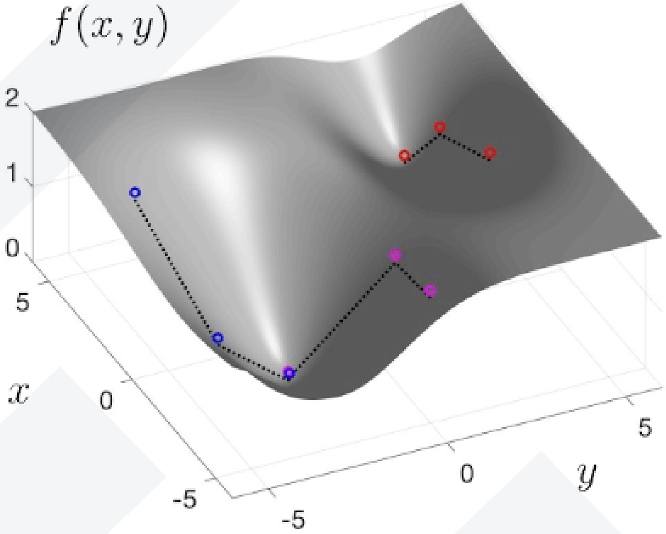
4. Criterio de parada

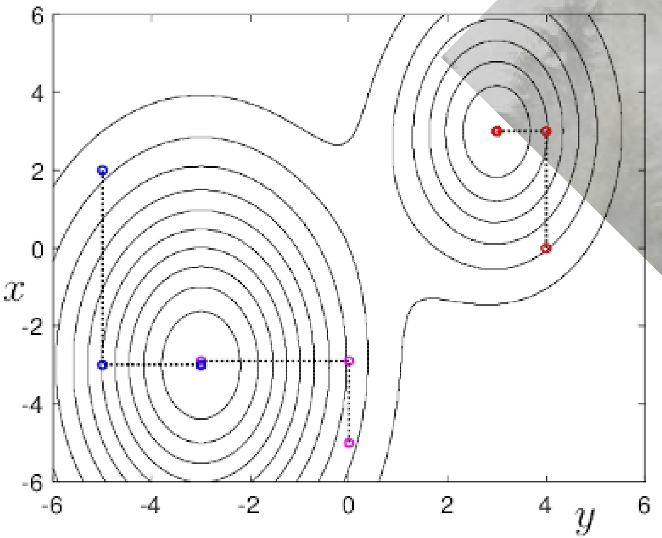
$$|\vec{\nabla}f(x_k)| < \varepsilon$$

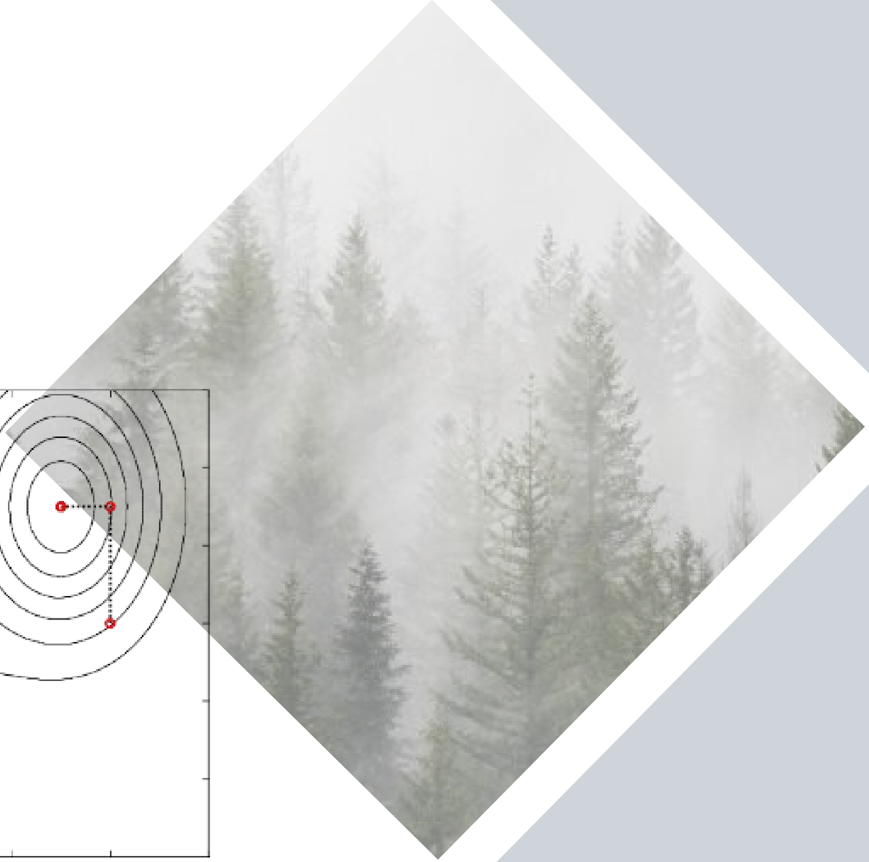
Alternativa: Descenso alternado

Otra posibilidad es tomar el sistema como unidimensional y minimizar la función para un parámetro.

Luego, proceder a minimizar para otro parámetro y así sucesivamente.







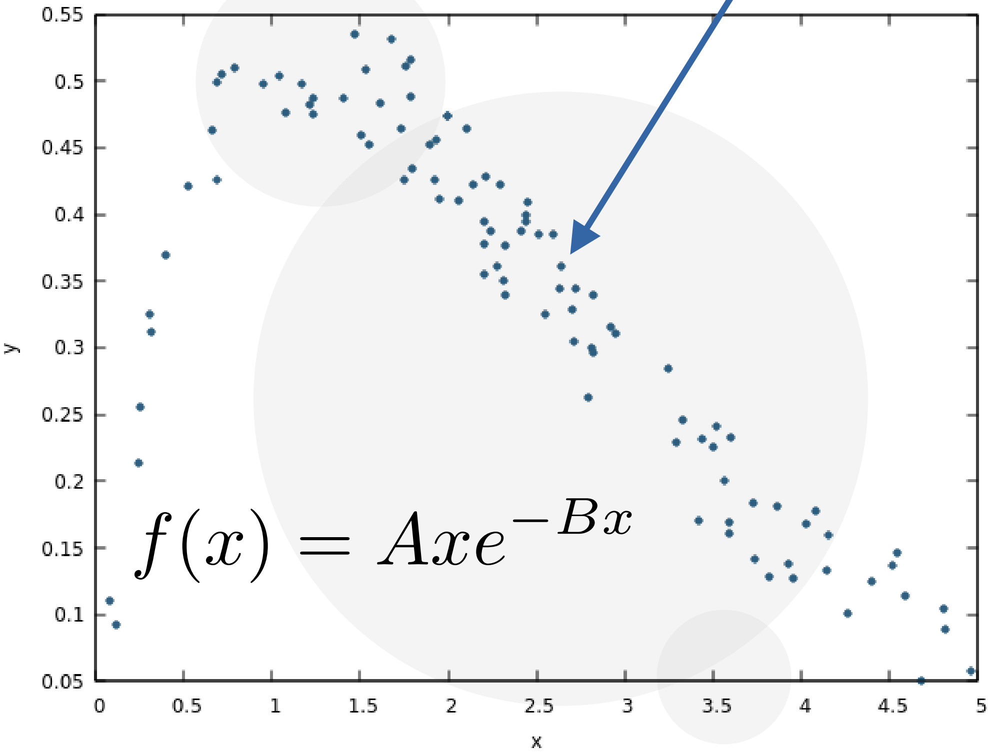
Caso particular

$$f(x) = Axe^{-Bx}$$

Machine Learning
Abril 2023

Datos

$$(x_k, y_k)$$

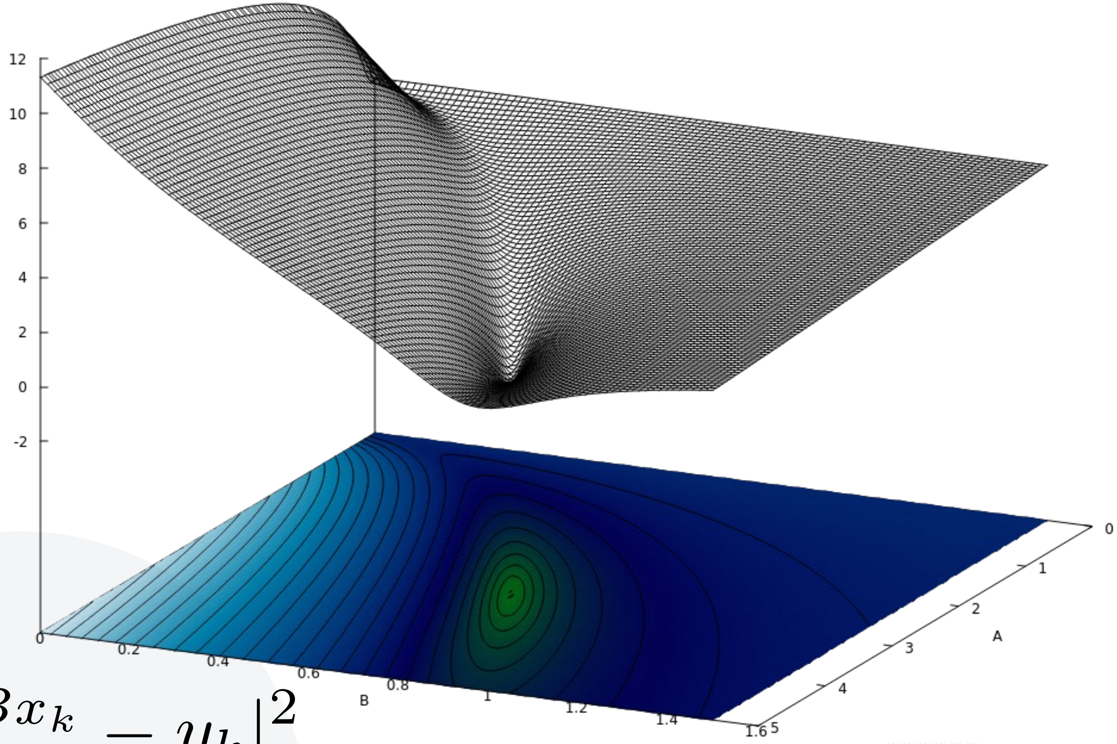


$$f(x) = Axe^{-Bx}$$

Función Error

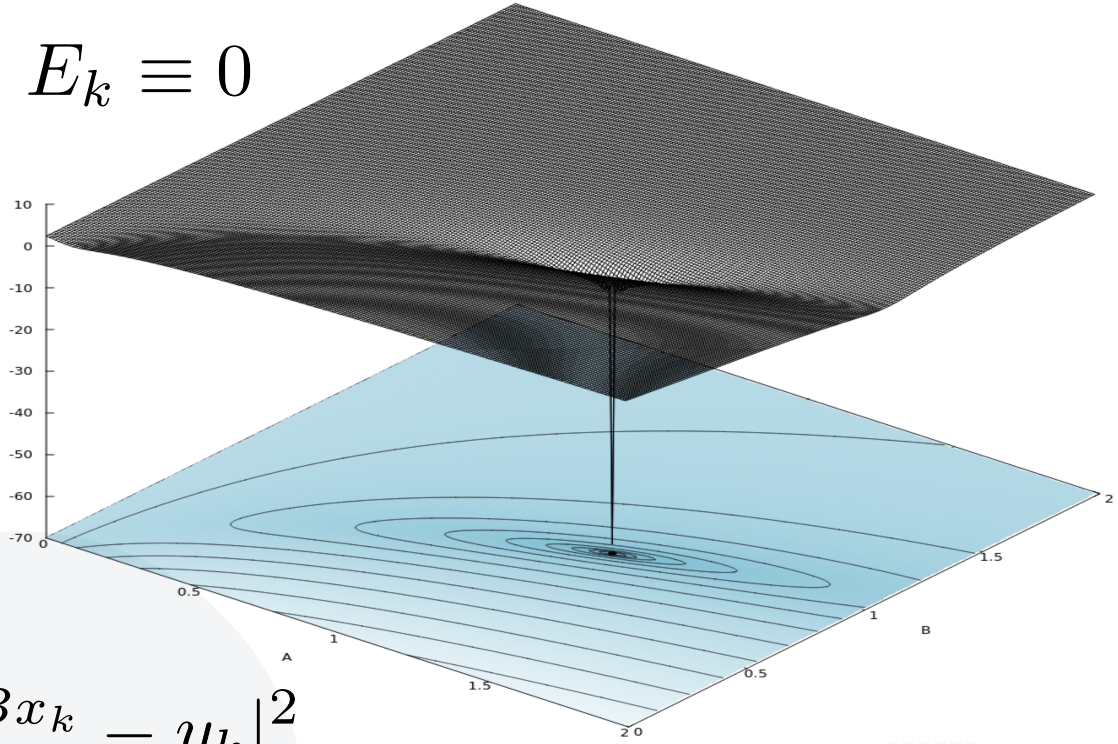
$$E(A,B) = \sum_k |Ax_k e^{-Bx_k} - y_k|^2$$

Función Error



$$E(A,B) = \sum_k |Ax_k e^{-Bx_k} - y_k|^2$$

Función Error



$$E(A,B) = \sum_k \left|A x_k e^{-B x_k} - y_k\right|^2$$

Descender por el gradiente

1. Seleccionar un punto inicial.

Es necesario contar con un amplio conjunto aleatorio de condiciones iniciales. Algunas no convergerán al mínimo global.

2. Calcular la dirección del gradiente

El gradiente indica el valor de la pendiente más fuerte en ascenso. Tomamos la dirección opuesta.

3. Calcular el paso a dar.

Si tomamos un paso demasiado grande, puede que nos pasemos de nuestro objetivo. Si es pequeño, tendremos que seguir en la misma dirección.

4. Criterio de parada

Si el gradiente es muy chico, podemos estar en una meseta o en un mínimo local.

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = A_0\hat{e}_A + B_0\hat{e}_B$$

2. Calcular la dirección del gradiente

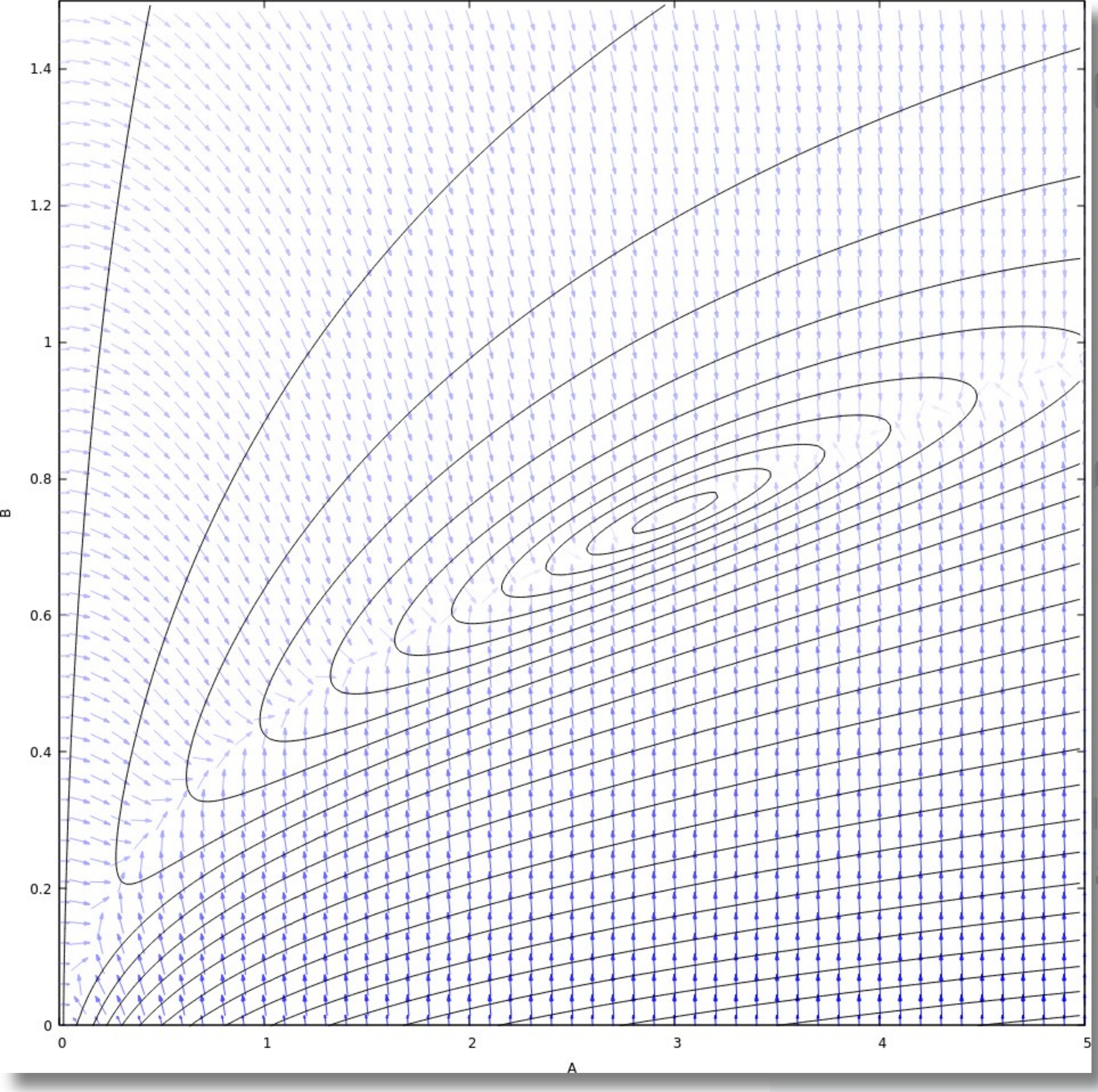
$$-\vec{\nabla}E(A,B)$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla}E(\vec{r}_k)$$

4. Criterio de parada

$$|\vec{\nabla}E(\vec{r}_k,)| < \varepsilon$$



ente

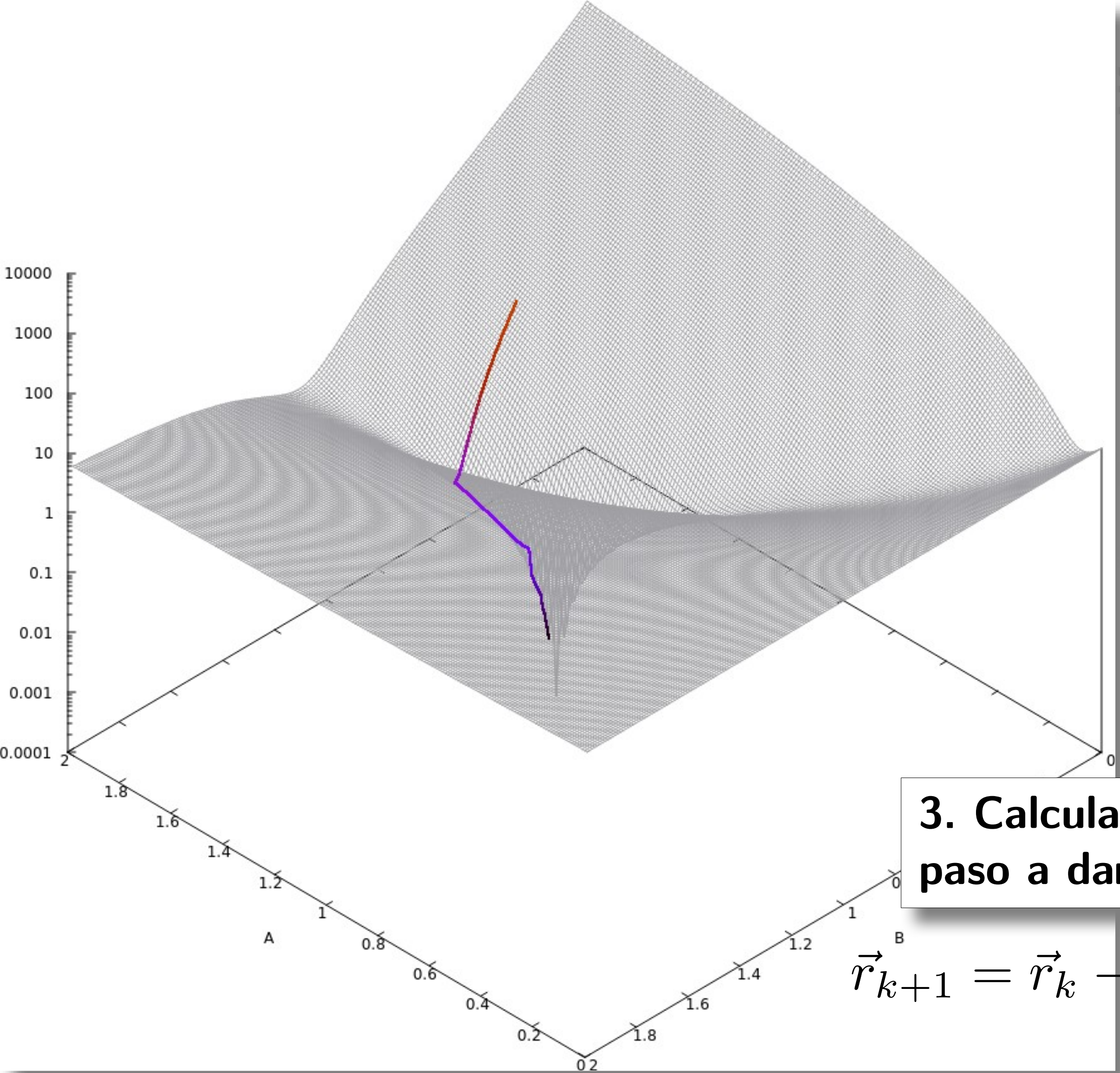
2. Calcular la dirección del gradiente

$$-\vec{\nabla}E(A,B)$$

lar el
ar.

$$- \delta\vec{\nabla}E(\vec{r}_k)$$

ente



2. Calcular la dirección del gradiente

$$-\vec{\nabla}E(A,B)$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla}E(\vec{r}_k)$$

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = A_0\hat{e}_A + B_0\hat{e}_B$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla}E(A,B)$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla}E(\vec{r}_k)$$

4. Criterio de parada

$$|\vec{\nabla}E(\vec{r}_k)| < \varepsilon$$

Descender por el gradiente

1. Seleccionar un punto inicial.

$$\vec{r}_0 = A_0\hat{e}_A + B_0\hat{e}_B$$

2. Calcular la dirección del gradiente

$$-\vec{\nabla}E(A,B)$$

3. Calcular el paso a dar.

$$\vec{r}_{k+1} = \vec{r}_k - \delta\vec{\nabla}E(\vec{r}_k)$$

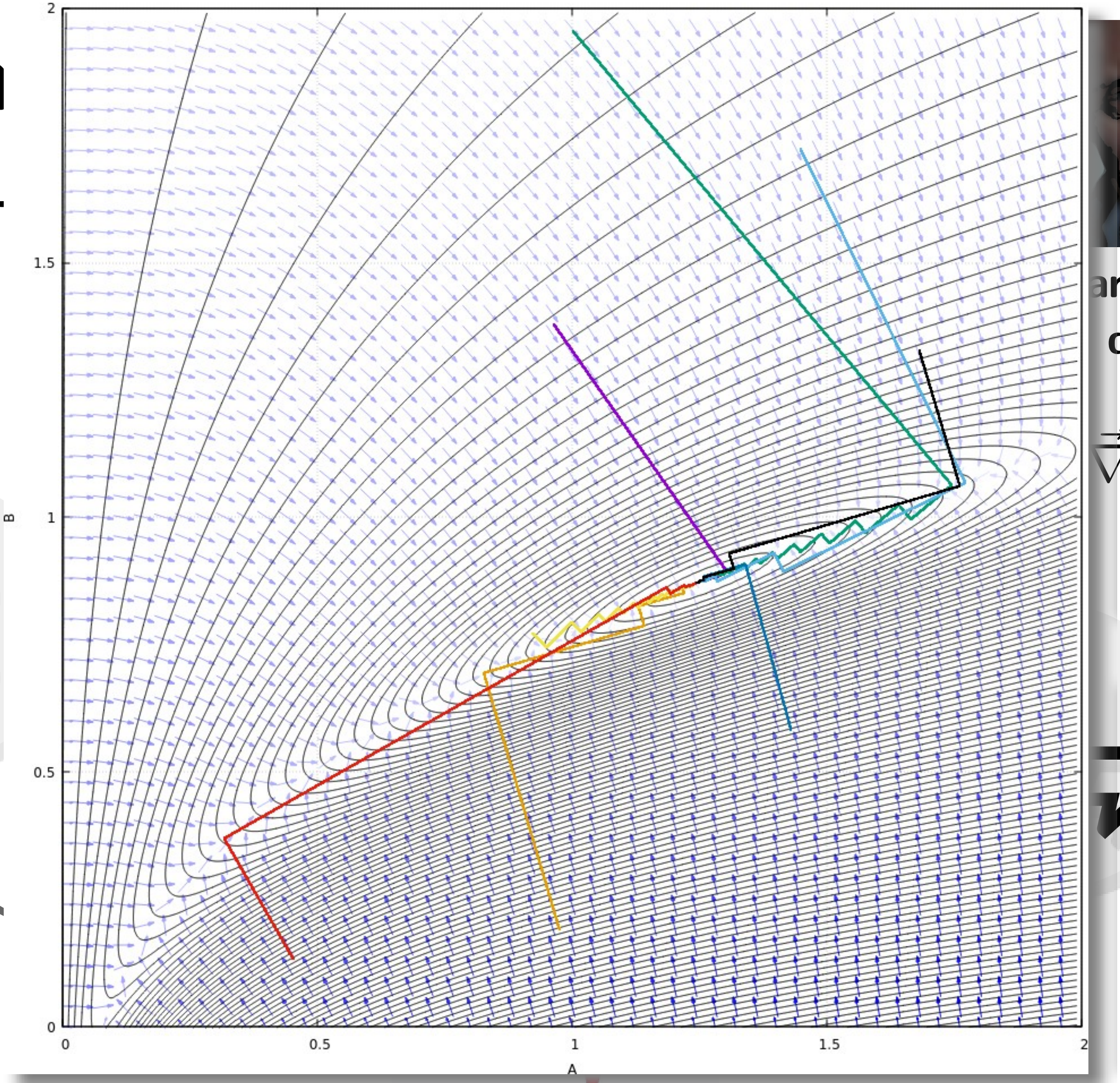
4. Criterio de parada

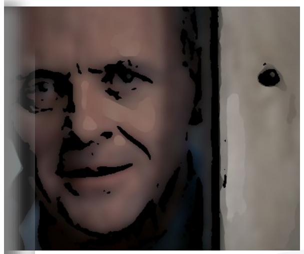
$$|\vec{\nabla}E(\vec{r}_k)| < \varepsilon$$

Descen

1.

$\vec{r}_0$

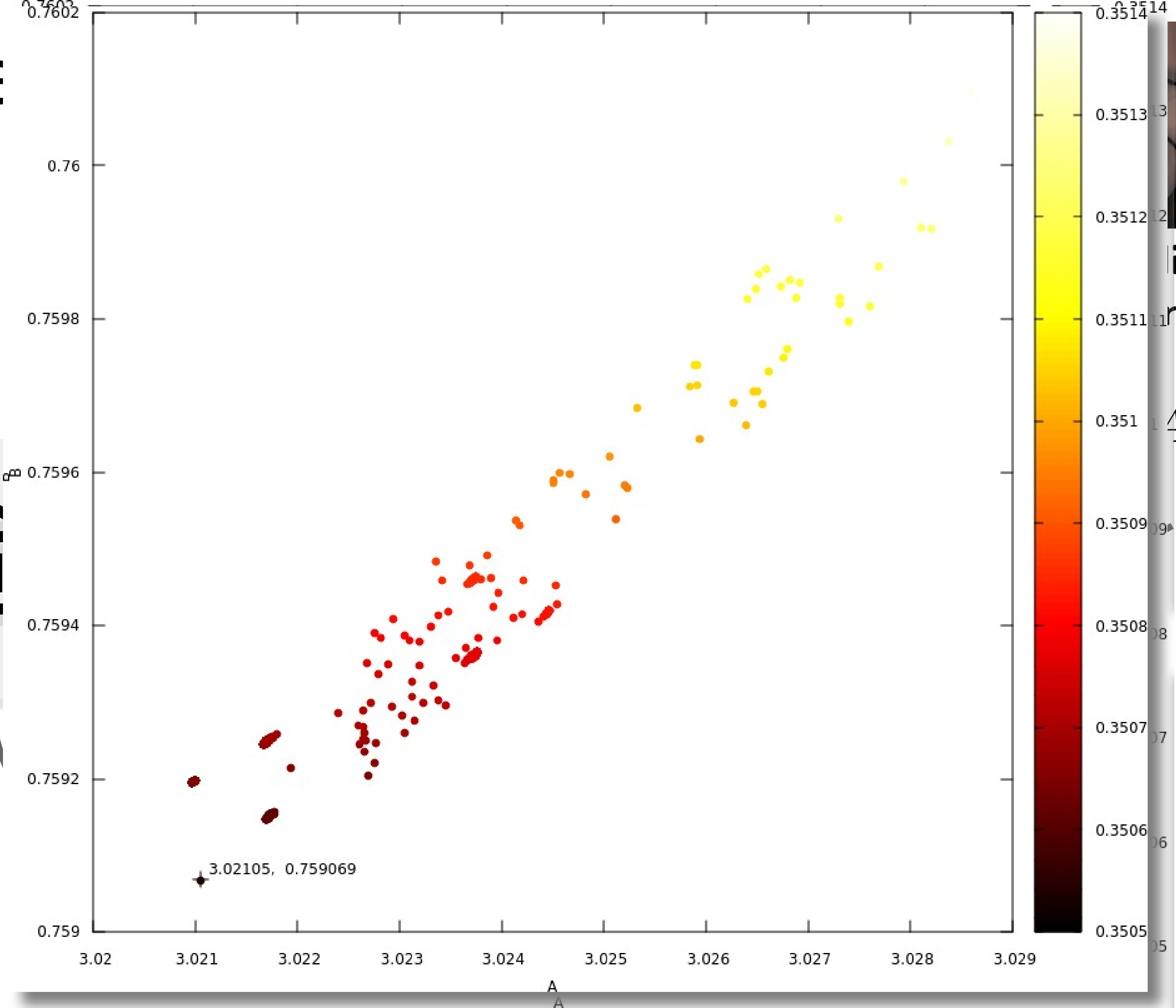


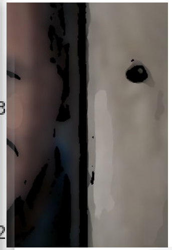


ar la dirección del gradiente

$\vec{\nabla} E(A,B)$

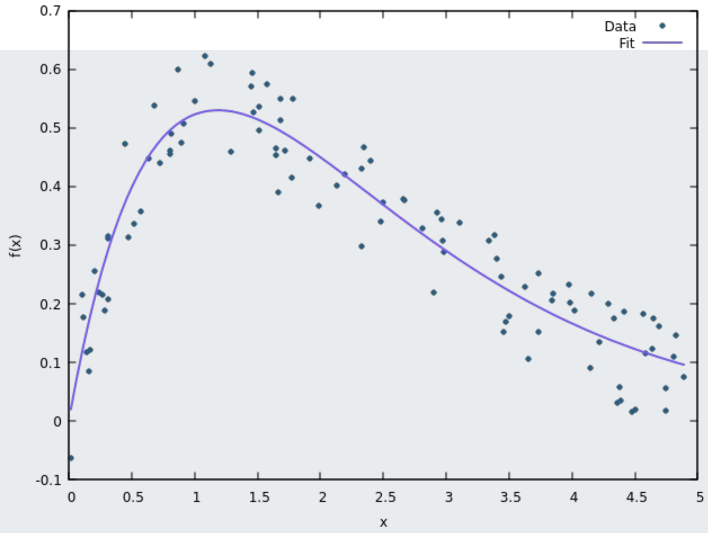
Desce





irección
radiente

$A, B)$


Data
Fit
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
f(x)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
x

Conclusión

- Método eficiente para todo tipo de problema
- No se llega a los valores exactos, debido a la amplitud de los errores y el criterio de parada.
- Se puede mejorar descartando regiones en las que la función es menor al error. (Supervisado)