Curso de Métodos Numéricos

DEMAT, Universidad de Guanajuato

Clase 10: **Métodos de relajación para resolver SELs. Método de mínimos cuadradas**

- Métodos de relajación.
- Métodos de sobrerelajación sucesiva (SOR).
- Métodos de mínimos cuadrados.

MAT–251

Dr. Joaquín Peña Acevedo
CIMAT A.C.
e-mail: joaquin@cimat.mx

Métodos de relajación

En ciertos casos, hacer que $r_{i,i+1}^{(t)}$ sea cero puede ser una estrategia algo agresiva. Por ello se hace

$$x_i^{(t)} = x_i^{(t-1)} + \omega \frac{r_{ii}^{(t)}}{a_{ii}}$$

para algún $\omega > 0$.

Si $0 < \omega < 1$, se dice que el método es *sub-relajado*.

Si $1 < \omega$, se dice que el método es *sobrerelajado*, y su propósito es acelerar la convergencia.

Método de sobrerelajación sucesiva (SOR)

Definimos un factor $1 < \omega$. Entonces

$$x_i^{(t)} = \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(t)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(t-1)}\right) + (1-\omega)x_i^{(t-1)}.$$

Para ver la separación de la matriz, sólo hay que reordenar los términos de la ecuación

$$a_{ii}x_i^{(t)} + \omega\sum_{j=1}^{i-1} a_{ij}x_j^{(t)} = \omega b_i - \omega\sum_{j=i+1}^{n} a_{ij}x_j^{(t-1)} + (1-\omega)a_{ii}x_i^{(t-1)}.$$

$$(\boldsymbol{Q} - \omega\boldsymbol{L})\boldsymbol{x}^{(t)} = [\,(1-\omega)\boldsymbol{Q} + \omega\boldsymbol{U}]\boldsymbol{x}^{(t-1)} + \omega\boldsymbol{b}.$$

Nota: El método converge para $0 < \omega < 2$ ($|1-\omega| < 1$).

Ejemplo 1 (I)

Ejemplo. Consideremos el sistema con la matriz tridiagonal de tamaño $n = 100$ en la que el método de Jacobi tiene una convergencia lenta.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

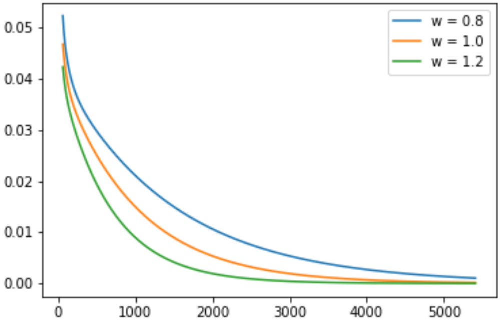
Fijando la tolerancia en $\tau = 10^{-5}$, el número de iteraciones que realiza el algoritmo depende del parámetro ω:

ω	$\|\boldsymbol{A}\boldsymbol{x}^{(t)} - \boldsymbol{b}\|_2$	**Iteraciones**
0.8	9.9957×10^{-6}	12172
1.0	9.9978×10^{-6}	8114
1.2	9.9920×10^{-6}	5409

En este caso el algoritmo diverge con $\omega = 1.5$.

Ejemplo 1 (II)

La siguiente gráfica muestra el valor del error $\|\boldsymbol{A}\boldsymbol{x}^{(t)} - \boldsymbol{b}\|_2$ dependiendo del parámetro ω.



Unidad 4:

Mínimos cuadrados lineales

Matrices rectangulares

Sea $\boldsymbol{A}$ una matriz $m \times n$ y $\boldsymbol{b}$ un vector de dimensión m. Tenemos que para

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- si $m > n$, el sistema es sobredeterminado.
- si $m < n$, el sistema es subdeterminado.

Para matrices cuadradas, $m = n$, si $\boldsymbol{A}$ es no singular, la solución es única,

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

Para matrices rectangulares, hay que analizar cada caso.

Gradientes de formas lineales y cuadráticas

Sean $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^n$, y $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.
Queremos calcular $\nabla(\boldsymbol{c}^\top \boldsymbol{x})$ y $\nabla(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x})$.

$$\nabla(\boldsymbol{c}^\top \boldsymbol{x}) = \boldsymbol{c}$$

$$\nabla(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}^\top)\boldsymbol{x}$$

De lo anterior se sigue que el gradiente del error

$$E(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$$

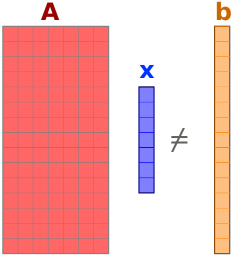
es

$$\nabla E(\boldsymbol{x}) = 2\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^\top \boldsymbol{b}$$

Introducción (I)

Consideramos una matriz $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$ con $m > n$.

- Lo usual es que el sistema sobredeterminado $\boldsymbol{Ax} = \boldsymbol{b}$ no tenga solución.



- En lugar de tratar de resolver el sistema de ecuaciones, planteamos el problema de optimización

$$\boldsymbol{x}^* = \min_{\mathbf{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 = \min_{\mathbf{x}} \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 .$$

Introducción (II)

- Si resulta que el valor mínimo de $\|\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}\|_2^2$ es cero, entonces $\boldsymbol{x}^*$ es solución del sistema de ecuaciones. En otro caso, sólo decimos que se obtuvo la mejor solución en el sentido de mínimos cuadrados.
- La solución del problema de mínimos cuadrados lineales es

$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b}.$$

- Para que $\boldsymbol{A}^\top \boldsymbol{A}$ sea no singular se require que $\boldsymbol{A}$ sea de rango (columna) completo.
- Para resolver este sistema conviene aplicar el método de factorización de Cholesky.

Sistemas sobredeteminados

Sea $\boldsymbol{A} = [\boldsymbol{A}_1 \ \cdots \ \boldsymbol{A}_n] \in \mathbb{R}^{m\times n}$ con $m > n$, formada por las columnas $\boldsymbol{A}_i$.

En general, el sistema $\boldsymbol{Ax} = \boldsymbol{b}$ puede no tener solución.

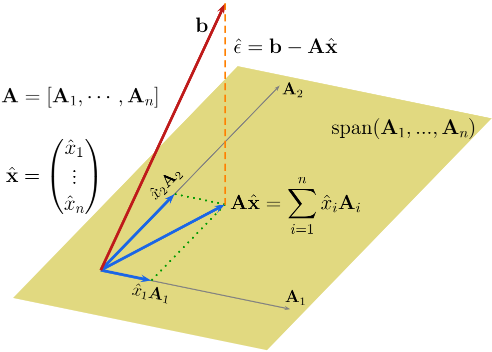
Por ello, tiene sentido calcular la solución de mínimos cuadrados,

$$\widehat{\boldsymbol{x}} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_m \end{pmatrix} = \arg\min_x \|\boldsymbol{Ax}-\boldsymbol{b}\|^2.$$

que está dada por

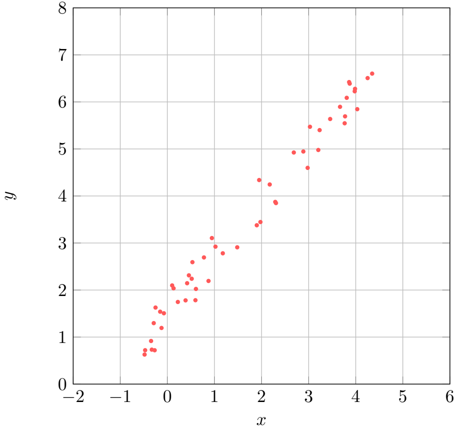
$$\boldsymbol{A}^\top \boldsymbol{Ax} = \boldsymbol{A}^\top \boldsymbol{b},$$

que tiene una interpretación geométrica.



Caso particular: Ajuste de una recta (I)

Tenemos m puntos $\{(x_i, y_i)\}$ y queremos hallar una recta que mejor los aproxime:



Caso particular: Ajuste de una recta (II)

Esto es, queremos determinar la pendiente a y la ordenada al origen b tales que

$$y_i = ax_i + b + \epsilon_i, \qquad i = 1, 2, \ldots, m$$

Si definimos

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}$$

Entonces $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}$ y podemos reescribir $E(a, b)$ como

$$E(a, b) = \sum_{i=1}^{m} \epsilon_i^2 = \|\boldsymbol{\epsilon}\|_2^2 = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

$$\begin{aligned} E(a, b) &= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) = (\boldsymbol{x}^\top \boldsymbol{A}^\top - \boldsymbol{y}^\top)(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \\ &= \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y} \end{aligned}$$

Caso particular: Ajuste de una recta (III)

Entonces

$$\nabla E = 0 \qquad \Longrightarrow \qquad \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{y}$$

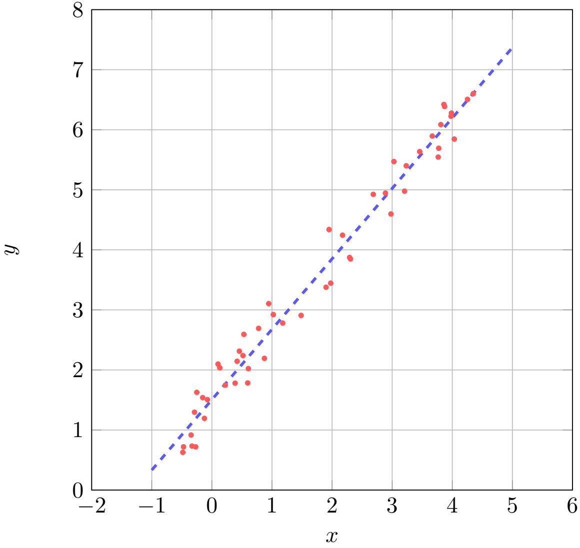
Para nuestro ejemplo, tenemos que $m = 50$, por lo que $\boldsymbol{A} \in \mathbb{R}^{50\times 2}$.
Además, $\det(\boldsymbol{A}^\top \boldsymbol{A}) = 6489.3$, por lo que el sistema tiene solución única:

$$\boldsymbol{x} = \begin{pmatrix} 1.173 \\ 1.505 \end{pmatrix}$$

En realidad, los datos fueron generados tomando puntos sobre la recta con $a = 1.2$ y $b = 1.5$ y se les agregó ruido a la ordenada.

En la siguiente gráfica se muestra en azul la recta obtenida por mínimos cuadrados

Caso particular: Ajuste de una recta (IV)



Planteamiento general del caso lineal (I)

Supongamos que tenemos un conjunto con m puntos en $\mathbb{R}^n$:

$$\{(x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})\}$$

A cada punto $(x_{i1}, x_{i2}, \ldots, x_{in})$ lo tenemos asociado a una observación y_i.

$$(x_{i1}, x_{i2}, \ldots, x_{in}) \mapsto y_i$$

El objetivo es determinar los coeficientes $c_0, c_1, \ldots, c_n$ un modelo que cumpla con:

$$y_i = c_0 + c_1 x_{i1} + \cdots + c_n x_{in} + \epsilon_i$$

donde ϵ_i es un error que tiene el modelo. Entonces

$$\epsilon_i = y_i - c_0 - \sum_{j=1}^{n} c_j x_{ij}$$

Queremos calcular los coeficientes c_j minimizando los errores ϵ_i. Una manera es resolver el problema

Planteamiento general del caso lineal (II)

$$\min_{c_0,\ldots,c_n} E(c_0,c_1,\ldots,c_n) = \sum_{i=1}^{m} \epsilon_i^2 = \sum_{i=1}^{m}\left(y_i - c_0 - \sum_{j=1}^{n} c_j x_{ij}\right)^2$$

Si definimos

$$\boldsymbol{A} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}$$

Entonces

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{\epsilon}$$

$$\min_{c_0,\ldots,c_n} E(\boldsymbol{c}) = \sum_{i=1}^{m} \epsilon_i^2 = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c})^\top (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c})$$

Planteamiento general del caso lineal (III)

Calculamos $\nabla E(\boldsymbol{c})$ y el vector $\boldsymbol{c}$ para el cual $\nabla E(\boldsymbol{c}) = \boldsymbol{0}$.

Se puede ver que la solución se obtiene resolviendo el sistema lineal

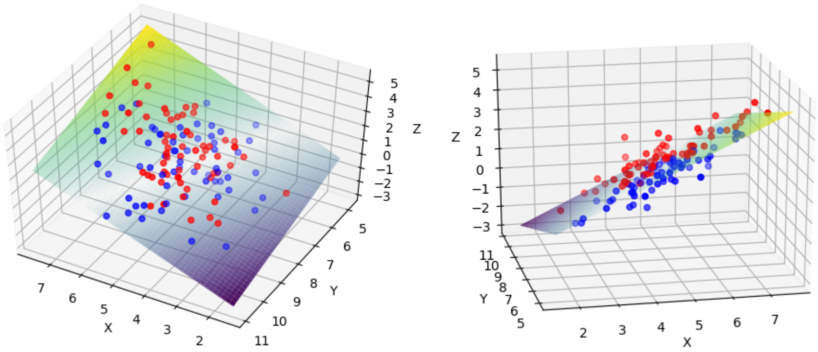
$$\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{c} = \boldsymbol{A}^\top \boldsymbol{y}.$$

Ejemplo. Consideramos un conjuntos de 125 puntos en el espacio 3D $\{(x_i, y_i, z_i)\}_{i=1}^{125}$. Se realiza la estimación de los coeficientes c_0, c_1 y c_2 del modelo:

$$z_i = c_0 + c_1 x_i + c_2 y_i + \epsilon_i, \quad i = 1, 2, \ldots, 125.$$

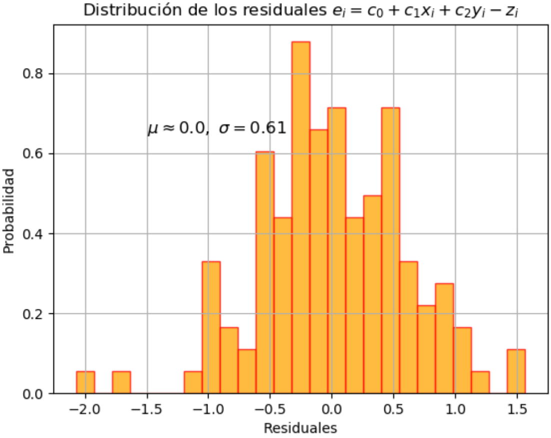
La siguiente gráfica muestra el resultado del ajuste:

Planteamiento general del caso lineal (IV)



El color que se asigna a los puntos depende del lado del semiplano en que está.

Planteamiento general del caso lineal (V)



Lo que NO estamos haciendo

No es que es estemos planteando resolver el sistema sobredeterminado

$$\boldsymbol{Ac} = \boldsymbol{y}$$

y para resolverlo multiplicamos por $\boldsymbol{A}^\top$ ambos miembros de la ecuación, pues el sistema puede ser inconsistente, por lo que podría no existir $\boldsymbol{c}$ que cumpla la igualdad.

Para el ejemplo anterior, tenemos que con

$$c_0^* = 0.9029, \qquad c_1^* = 0.8608, \qquad c_2^* = -0.4621.$$

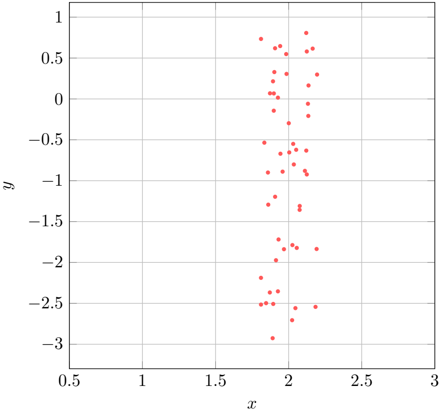
se tiene un error $E(c_0^*, c_1^*, c_2^*) = \sum_{i=1}^m e_i^2 = 6.7787.$

En ocasiones conviene reportar la raíz del error cuadrático medio (en inglés: *root mean square error*, RMSE):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{n} a_{ij}c_j\right)^2}, \qquad RMSE = \sqrt{\frac{1}{m}E(c_0^*, c_1^*, c_2^*)} \approx 0.37$$

Ajuste de una recta general (I)

Tenemos un conjunto de puntos $\{\boldsymbol{x}_i\}_{i=1}^{m}$, con $\boldsymbol{x}_i = (x_i, y_i)^\top$. El modelo $y = ax + b$ no sirve para el siguiente conjunto de datos.



Ajuste de una recta general (II)

Usamos la siguiente formulación:
Sea $\bar{\boldsymbol{x}} = (\bar{x}, \bar{y})^\top$ el centroide de los puntos,

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i, \qquad \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

Idealmente $\bar{\boldsymbol{x}}$ debería estar sobre la recta que vamos a ajustar, de modo que si $\boldsymbol{x}$ es otro punto sobre la recta, se debe tener que

$$\boldsymbol{n}^\top(\boldsymbol{x} - \bar{\boldsymbol{x}}) = 0.$$

Como los datos no están sobre una recta, debemos tener

$$\epsilon_i = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top \boldsymbol{n} = [\, x_i - \bar{x} \quad y_i - \bar{y} \,]\boldsymbol{n} \qquad i = 1, \ldots, m.$$

Si definimos

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix} = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ \vdots & \vdots \\ x_m - \bar{x} & y_m - \bar{y} \end{bmatrix} \boldsymbol{n} = \boldsymbol{A}\boldsymbol{n}$$

Ajuste de una recta general (III)

Entonces queremos minimizar

$$\min \sum_{i=1}^{m} \epsilon_i^2 = \|\boldsymbol{\epsilon}\|^2 = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = \boldsymbol{n}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{n}$$

Note que

- La matriz $\boldsymbol{A}^\top \boldsymbol{A}$ es semidefinida positiva.
- Falta agregar una condición al problema. Así como está planteado el problema, el vector $\boldsymbol{n} = \boldsymbol{0}$ es solución.

Pedimos que, por ejemplo, $\|\boldsymbol{n}\| = 1$. Esto equivale a pedir que $\boldsymbol{n}^\top \boldsymbol{n} = 1$. Entonces el problema que queremos resolver es

$$\min \boldsymbol{n}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{n} \qquad \text{sujeto a} \quad \boldsymbol{n}^\top \boldsymbol{n} = 1.$$

Para resolverlo, construimos la función Lagrangiana

$$\mathcal{L}(\boldsymbol{n}, \lambda) = \boldsymbol{n}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{n} - \lambda(\boldsymbol{n}^\top \boldsymbol{n} - 1).$$

Ajuste de una recta general (IV)

Calculamos su gradiente e igualamos a cero:

$$2\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{n} - 2\lambda\boldsymbol{n} = 0 \qquad \Longrightarrow \qquad \boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{n} = \lambda\boldsymbol{n}$$

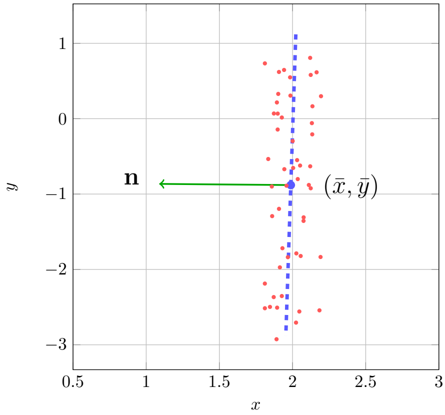
Esto es, la solución es un eigenvector de la matriz $\boldsymbol{A}^\top\boldsymbol{A}$.
Si sustituimos en la función de error tenemos

$$\boldsymbol{n}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{n} = \boldsymbol{n}^\top(\lambda\boldsymbol{n}) = \lambda\boldsymbol{n}^\top\boldsymbol{n} = \lambda$$

Por lo tanto, el error mínimo se obtiene si elegimos $\boldsymbol{n}$ como el eigenvector asociado al eigenvalor más pequeño.

Para cada conjunto de datos, se obtienen los resultados siguientes:

Ajuste de una recta general (V)



Eigenvalores:
$\lambda_1 = 0.6054$, $\lambda_2 = 61.324$

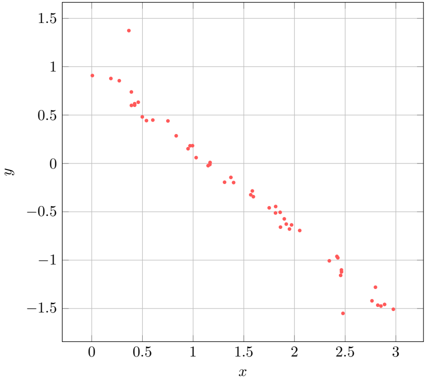
Eigenvectores

$$\boldsymbol{v}_1 = \begin{pmatrix} -0.9998 \\ 0.0171 \end{pmatrix}, \ \boldsymbol{v}_2 = \begin{pmatrix} 0.0171 \\ 0.9998 \end{pmatrix}$$

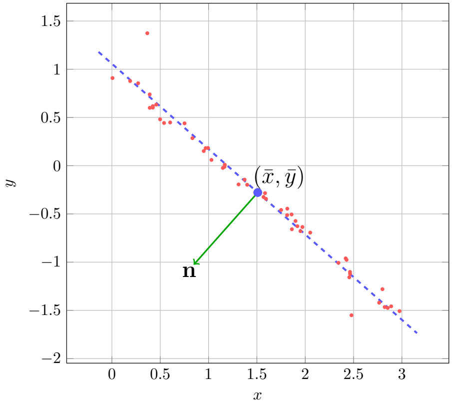
Así, $\boldsymbol{n} = \boldsymbol{v}_1$.

Este enfoque se puede aplicar a cualquier conjunto de datos que puedan ser aproximados por un modelo lineal, por ejemplo:

Ajuste de una recta general (VI)



Ajuste de una recta general (VII)



Eigenvalores:
$\lambda_1 = 0.4466$, $\lambda_2 = 64.785$

Eigenvectores

$$\boldsymbol{v}_1 = \begin{pmatrix} -0.663 \\ -0.749 \end{pmatrix}, \ \boldsymbol{v}_2 = \begin{pmatrix} -0.749 \\ 0.663 \end{pmatrix}$$

Así, $\boldsymbol{n} = \boldsymbol{v}_1$.

Observación (I)

Dado el conjunto de puntos $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ con $\boldsymbol{x}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$, resolver el problema

$$\min_{a,b} \sum_{i=1}^{m} (ax_i + b - y_i)^2$$

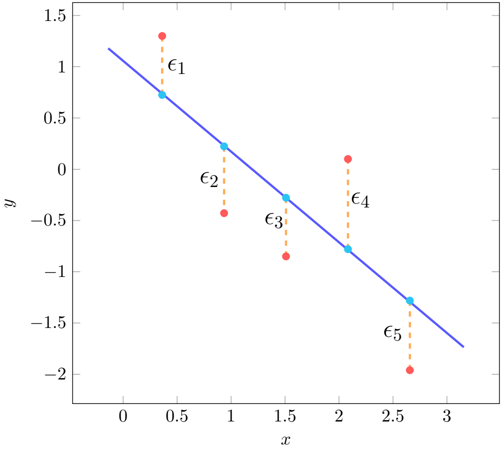
no da el mismo resultado que resolver el problema

$$\min_{\mathbf{n}} \sum_{i=1}^{m} \left[\boldsymbol{n}^\top (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\right]^2 \quad \text{sujeto a} \quad \boldsymbol{n}^\top \boldsymbol{n} = 1.$$

Observación (II)

En el primer caso la solución disminuye la suma de los cuadrados de las diferencias entre las ordenadas de los puntos (x_i, y_i) y $(x_i, ax_i + b)$, es decir, entre el dato y lo que predice el modelo.

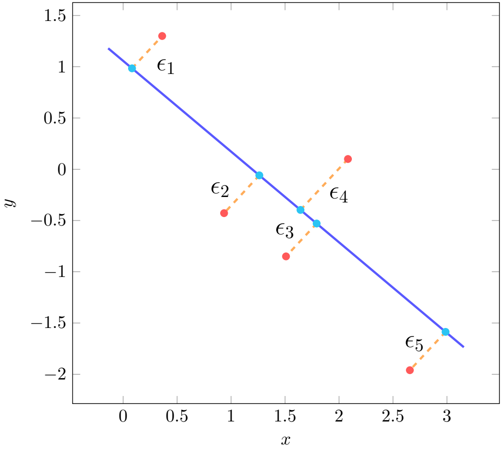
Observación (III)



Observación (IV)

En el segundo caso la solución disminuye la suma de los cuadrados de las proyecciones de los vectores $\boldsymbol{x}_i - \bar{\boldsymbol{x}}$ sobre la dirección $\boldsymbol{n}$, es decir, de las "distancias con signo" de los puntos $\boldsymbol{x}_i$ a la recta.

Observación (V)



Observación (VI)

Para la estimación de pendiente a y la ordenada al origen b se obtiene la recta

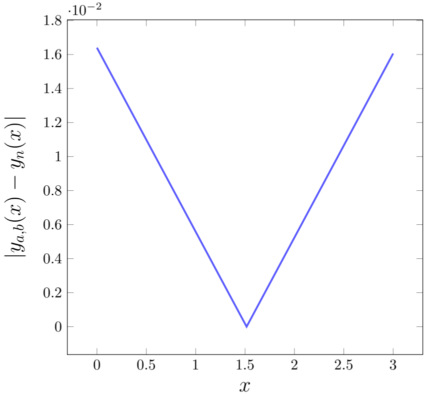
$$y_{a,b}(x) = -0.87417x + 1.0409,$$

con un error $\sum_{i=1}^{m} \epsilon_i^2 = 0.792$, donde $\epsilon_i = ax_i + b - y_i$.

Para la estimación del vector ortogonal $\boldsymbol{n}$ se obtiene la recta

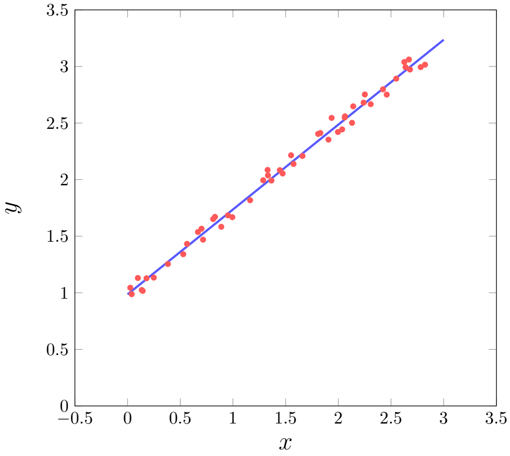
$$y_{\mathbf{n}}(x) = -0.88499x + 1.0573,$$

con un error $\sum_{i=1}^{m} \epsilon_i^2 = 0.446$, donde $\epsilon_i = \boldsymbol{n}^\top(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$.



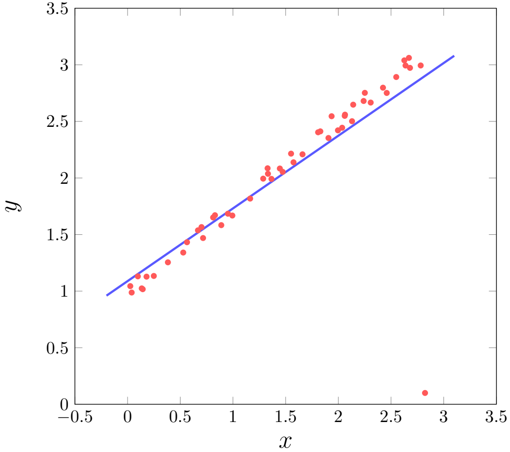
Sensibilidad de la solución por datos atípicos (I)

Sin datos atípicos, la solución de mínimos cuadrados proporciona un buen ajuste a los datos:



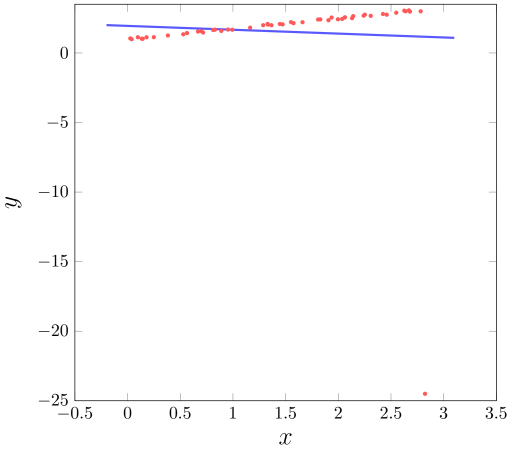
Sensibilidad de la solución por datos atípicos (II)

Sin embargo, un solo dato atípico puede modificar la solución del problema de mínimos cuadrados:



Sensibilidad de la solución por datos atípicos (III)

Entre más atípico sea el dato, mayor es el cambio en la solución del problema de mínimos cuadrados:



Sensibilidad de la solución por datos atípicos (IV)

Para reducir el efecto que tienen los datos atípicos, se puede agregar un peso $w_i > 0$ a cada termino de la suma de cuadrados:

$$E(a,b) = \sum_{i=1}^{m} w_i (a x_i + b - y_i)^2$$

de modo que a los términos que introducen errores grandes se les da un menor peso.

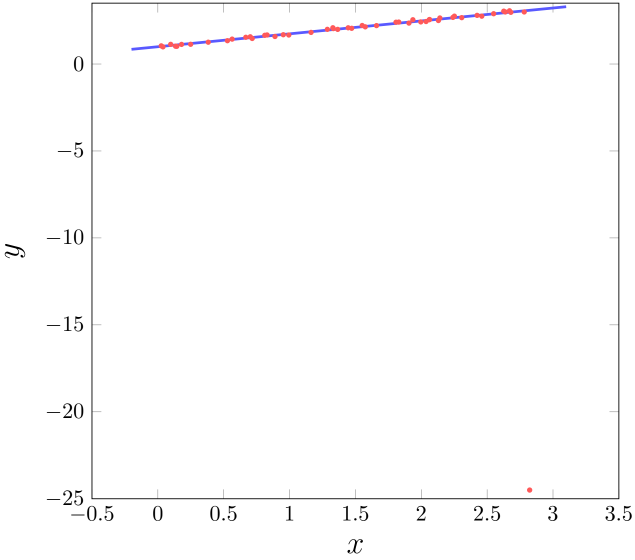
Esto nos lleva a un problema de *mínimos cuadrados pesados*.
En el ejemplo, cuando todos los pesos son $w_i = 1$, se tiene

$$a^* = -0.277, \qquad b^* = 1.945$$

Si cambiamos el peso del dato atípico a $w_{49} = 0.01$, se obtiene

$$a^* = 0.742, \qquad b^* = 0.994$$

Sensibilidad de la solución por datos atípicos (V)



Mínimos cuadrados pesados (I)

Se puede escribir el error

$$E(a,b)=\sum_{i=1}^{m} w_i(ax_i+b-y_i)^2$$

introduciendo las matrices y vectores

$$\boldsymbol{W}=\mathrm{diag}(w_1,\ldots,w_m),\quad \boldsymbol{A}=\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix},\quad \boldsymbol{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},\quad \boldsymbol{c}=\begin{pmatrix} a \\ b \end{pmatrix}.$$

$$\begin{aligned} E(\boldsymbol{c}) &= (\boldsymbol{A}\boldsymbol{c}-\boldsymbol{y})^\top \boldsymbol{W}(\boldsymbol{A}\boldsymbol{c}-\boldsymbol{y}) = (\boldsymbol{c}^\top\boldsymbol{A}^\top-\boldsymbol{y}^\top)\boldsymbol{W}(\boldsymbol{A}\boldsymbol{c}-\boldsymbol{y}) \\ &= \boldsymbol{c}^\top\boldsymbol{A}^\top\boldsymbol{W}\boldsymbol{A}\boldsymbol{c} - 2\boldsymbol{y}^\top\boldsymbol{W}\boldsymbol{A}\boldsymbol{c} + \boldsymbol{b}^\top\boldsymbol{W}\boldsymbol{b}. \end{aligned}$$

Por tanto, la solución de mínimos cuadrados está dada al resolver el sistema

$$\boldsymbol{A}^\top\boldsymbol{W}\boldsymbol{A}\boldsymbol{c}=\boldsymbol{A}^\top\boldsymbol{W}\boldsymbol{y}.$$

El problema es como fijar los pesos $w_1,\ldots,w_m$ asociados a cada ecuación. El método de mínimos cuadrados pesados hace la siguiente propuesta:

Mínimos cuadrados pesados (II)

1. Inicializar $k = 0$, los pesos $w_i = 1$ y formar la matriz $\boldsymbol{W}^{(0)}$.
2. Calcular la solución del problema de mínimos cuadrados
$$\boldsymbol{A}^\top \boldsymbol{W}^{(k)} \boldsymbol{A} \boldsymbol{c}^{(k)} = \boldsymbol{A}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}.$$
3. Calcular el vector de error $\boldsymbol{\epsilon} = \boldsymbol{A}\boldsymbol{c}^{(k)} - \boldsymbol{y}$.
4. Si $\|\boldsymbol{c}^{(k)} - \boldsymbol{c}^{(k-1)}\|$ o el RMSE $\sqrt{\dfrac{\boldsymbol{\epsilon}^\top \boldsymbol{W} \boldsymbol{\epsilon}}{m}}$ es menor que la tolerancia dada, terminar el algoritmo devolviendo la solución $\boldsymbol{c}^{(k)}$.
5. En caso contrario, modificar el valor del peso w_i de acuerdo a la magnitud de la componente i-ésima del vector de error, $\epsilon_i = (\boldsymbol{A}\boldsymbol{c})_i - y_i$, de modo que si ϵ_i es grande, el peso w_i debe ser pequeño, y viceversa. Construir la matriz $\boldsymbol{W}^{(k+1)}$. Hacer $k = k + 1$ y volver al paso 2.

Ejemplo de mínimos cuadrados pesados (I)

Consideramos el problema de ajustar una recta a un conjunto de puntos $\{(x_1, y_1), \dots, (x_m, y_m)\}$ en el plano aplicando el método de mínimos cuadrados pesados.

Una vez que se tiene una solución $\boldsymbol{c}^{(k)}$ en la iteración k, actualizamos el valor de los pesos $w_i^{(k+1)}$ se actualiza para la siguiente iteración mediante:

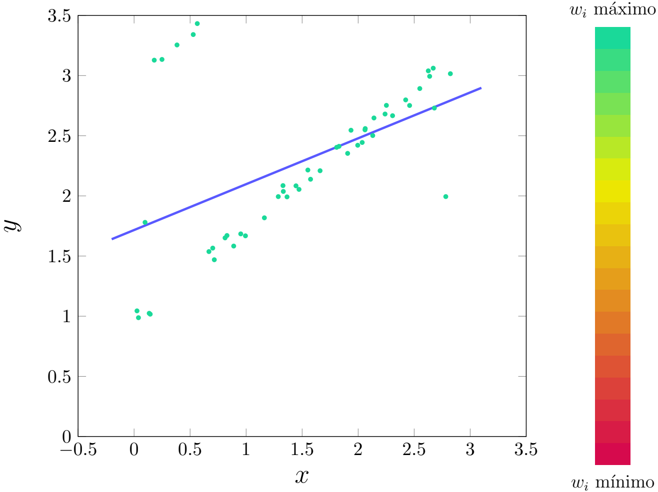
$$w_i^{(k+1)} = \exp(-5\epsilon_i^2), \qquad \text{con } \epsilon_i = \left(\boldsymbol{A}\boldsymbol{c}^{(k)}\right)_i - y_i.$$

La tolerancia usada es $\tau = \sqrt{\epsilon_m}$.

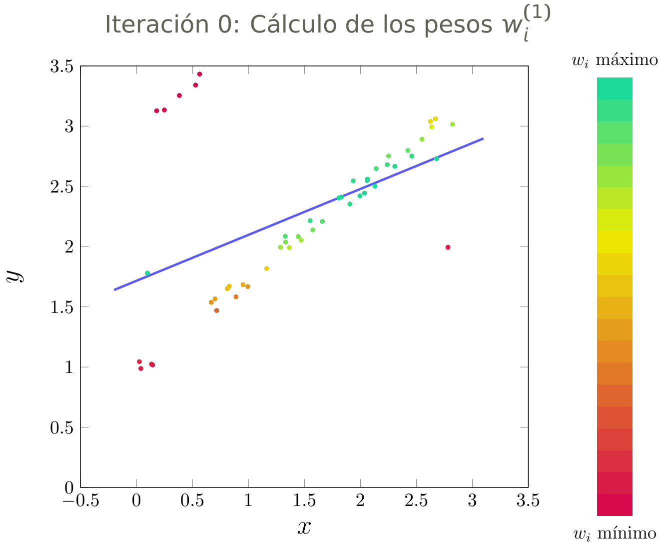
En las siguientes gráficas se muestra el resultado de aplicar el método de mínimos cuadrados pesados.
Los puntos son coloreados de acuerdo al valor del peso w_i que tienen asociado.

Ejemplo de mínimos cuadrados pesados (II)

Iteración 0: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(0)} \boldsymbol{A}\boldsymbol{c}^{(0)} = \boldsymbol{A}^\top \boldsymbol{W}^{(0)} \boldsymbol{y}$ con $w_i^{(0)} = 1$

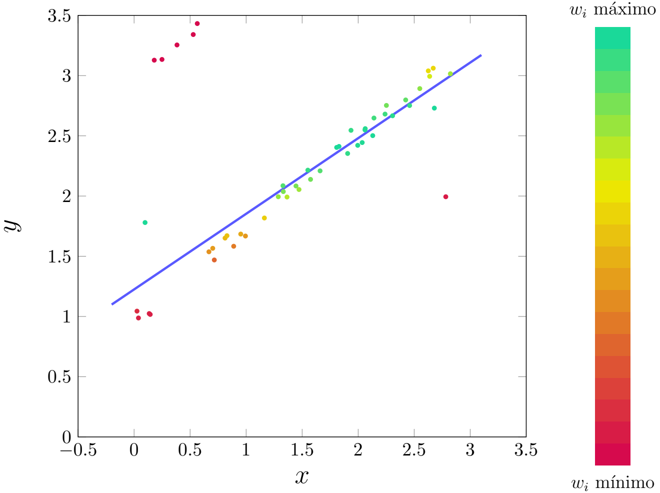


Ejemplo de mínimos cuadrados pesados (III)

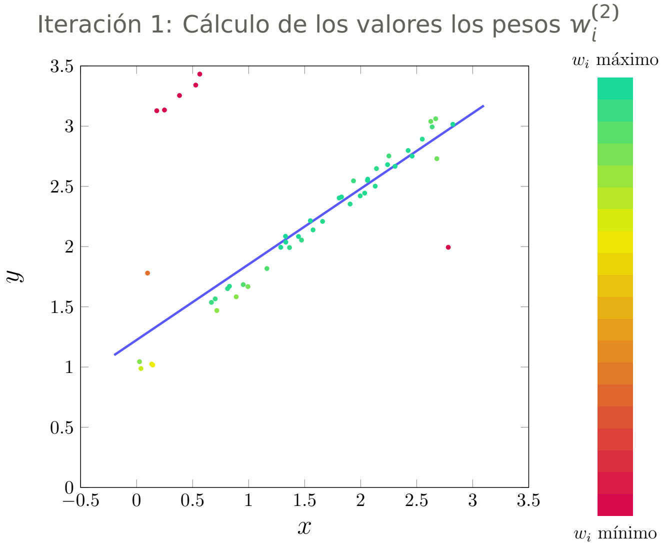


Ejemplo de mínimos cuadrados pesados (IV)

Iteración 1: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(1)} \boldsymbol{A} \boldsymbol{c}^{(1)} = \boldsymbol{A}^\top \boldsymbol{W}^{(1)} \boldsymbol{y}$

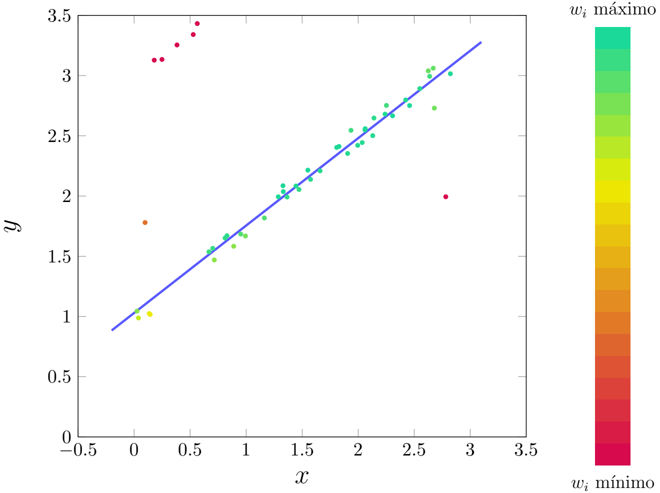


Ejemplo de mínimos cuadrados pesados (V)



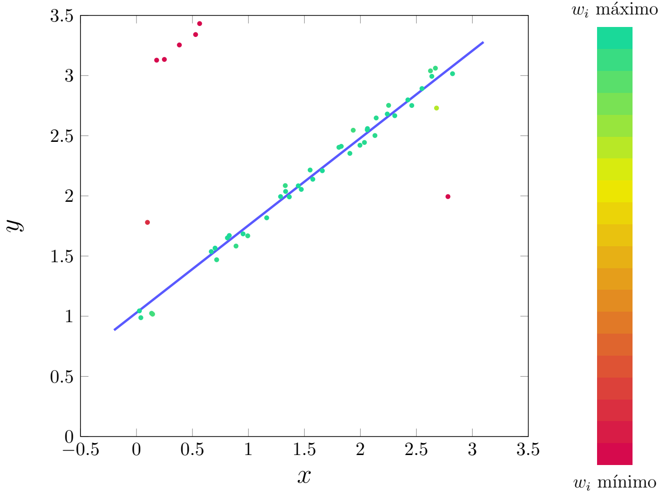
Ejemplo de mínimos cuadrados pesados (VI)

Iteración 2: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(2)} \boldsymbol{A} \boldsymbol{c}^{(2)} = \boldsymbol{A}^\top \boldsymbol{W}^{(2)} \boldsymbol{y}$



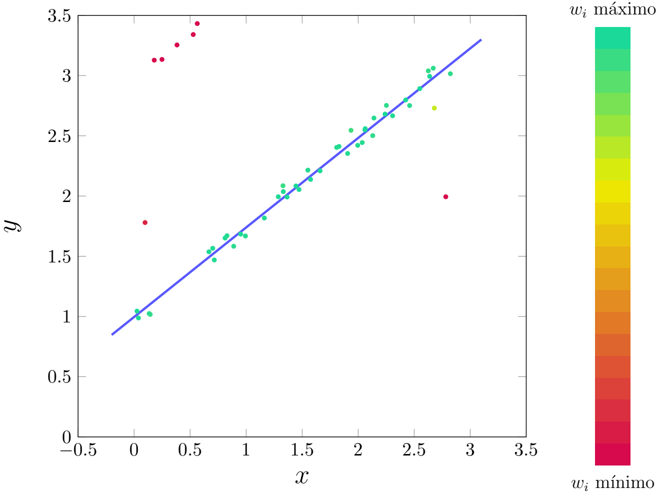
Ejemplo de mínimos cuadrados pesados (VII)

Iteración 2: Actualización de los valores los pesos $w_i^{(3)}$



Ejemplo de mínimos cuadrados pesados (VIII)

Iteración 10: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(10)} \boldsymbol{A} \boldsymbol{c}^{(10)} = \boldsymbol{A}^\top \boldsymbol{W}^{(10)} \boldsymbol{y}$



Ejemplo de mínimos cuadrados pesados (IX)

- El método de mínimos cuadrados pesados no siempre funciona.
- El método depende de la solución en la primera iteración y de la manera en la que se asignan los pesos.

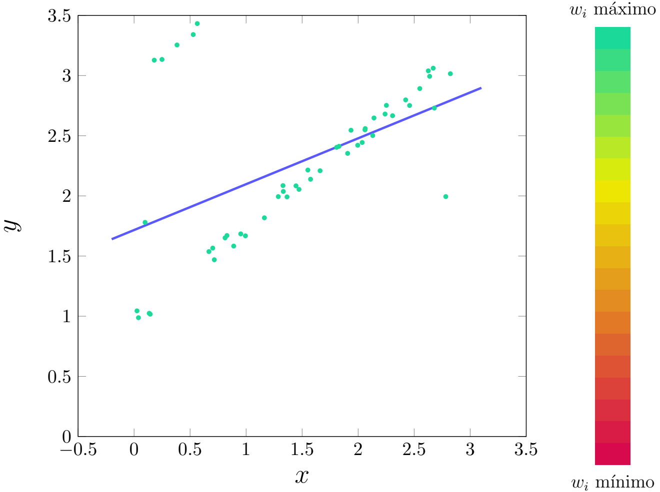
Por ejemplo, si se cambia la manera de calcular los pesos de la siguiente manera:

$$w_i^{(k+1)} = \exp(-150\epsilon_i^2), \qquad \text{con } \epsilon_i = \left(\boldsymbol{A}\boldsymbol{c}^{(k)}\right)_i - y_i,$$

se obtiene los siguientes resultados.

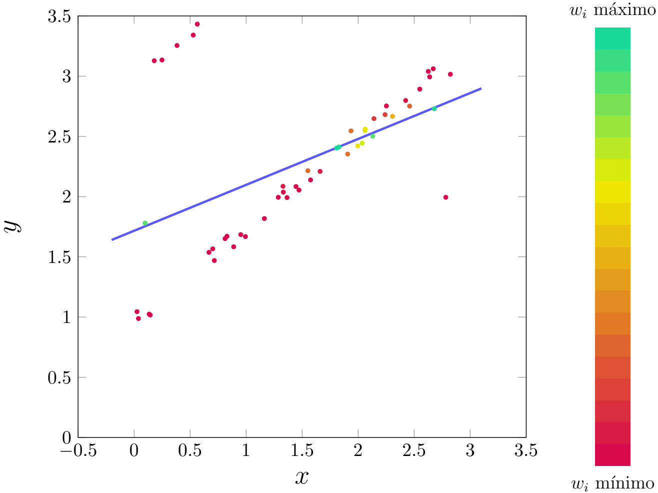
Ejemplo de mínimos cuadrados pesados (X)

Iteración 0: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(0)} \boldsymbol{A} \boldsymbol{c}^{(0)} = \boldsymbol{A}^\top \boldsymbol{W}^{(0)} \boldsymbol{y}$ con $w_i^{(0)} = 1$



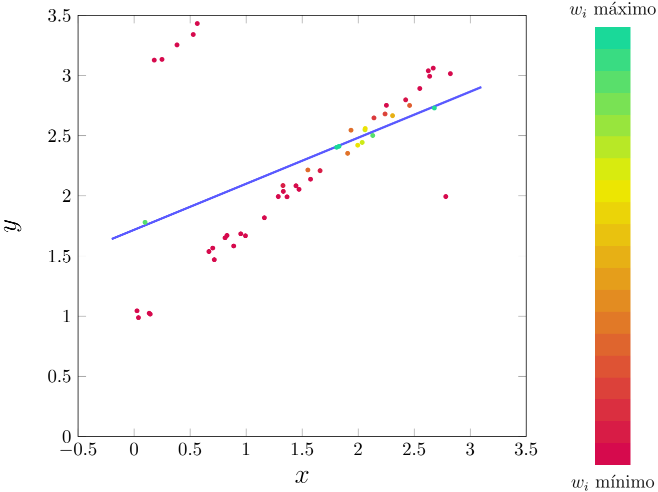
Ejemplo de mínimos cuadrados pesados **(XI)**

Iteración 0: Actualización de los valores de pesos $w_i^{(1)}$



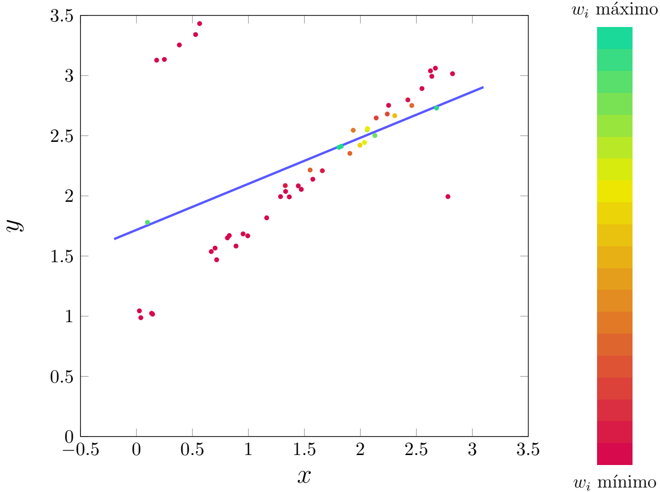
Ejemplo de mínimos cuadrados pesados (XII)

Iteración 1: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(1)} \boldsymbol{A} \boldsymbol{c}^{(1)} = \boldsymbol{A}^\top \boldsymbol{W}^{(1)} \boldsymbol{y}$



Ejemplo de mínimos cuadrados pesados (XIII)

Iteración 1: Actualización de los valores los pesos $w_i^{(2)}$



Ejemplo de mínimos cuadrados pesados (XIV)

Iteración 118: Cálculo de la solución $\boldsymbol{A}^\top \boldsymbol{W}^{(118)} \boldsymbol{A} \boldsymbol{c}^{(118)} = \boldsymbol{A}^\top \boldsymbol{W}^{(118)} \boldsymbol{y}$