GF5017 - Geodesia Aplicada a la Tectónica Activa

Repaso de Álgebra Lineal y Estimación de Parámetros con el Método de Mínimos Cuadrados

Versión Agosto 2019

Preparado por
Francisco Hernán Ortega Culaciati (ortega.francisco@uchile.cl)
Departamento de Geofísica
Facultad de Ciencias Físicas y Matemáticas
Universidad de Chile

1. Álgebra Lineal

En esta sección se plantean definiciones básicas del álgebra lineal, de operaciones y propiedades vectoriales y matriciales. A modo de ejemplo, estas definiciones permiten plantear y resolver sistemas de ecuaciones lineales.

1.1. Definición de vector

Se llama vector $\mathbf{V}$ de dimensión n a una tupla de n números reales, llamados componentes del vector. El conjunto de todos los vectores posibles de dimensión n se denomina $\mathbb{R}^n$. Luego un vector $\mathbf{V} \in \mathbb{R}^n$ se representa como

$$\mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} \tag{1}$$

1.2. Definición de espacio vectorial lineal

Sea $F \in \mathbb{R}$ un campo escalar y $\mathbf{V} \in \mathbb{R}^n$ un conjunto de n elementos $V_i \in \mathbb{R}$. Un espacio vectorial lineal sobre F es un conjunto de elementos $\mathbf{V}$ junto a una función llamada "adición" que transforma valores en el dominio $\mathbf{V} \times \mathbf{V}$ a valores en el rango $\mathbf{V}$; y una función llamada "multiplicación escalar" que transforma valores en el dominio $F \times \mathbf{V}$ a valores en el rango $\mathbf{V}$, que satisfacen las siguientes propiedades para todo $x, y, z \in \mathbf{V}$ y para todo $\alpha, \beta \in F$

1. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
2. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
3. Existe $\mathbf{0} \in \mathbf{V}$ tal que $\mathbf{x} + \mathbf{0} = \mathbf{x}$, $\forall \mathbf{x} \in \mathbf{V}$
4. Para cada $\mathbf{x} \in \mathbf{V}$ existe un elemento $-\mathbf{x} \in \mathbf{V}$ tal que $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$
6. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$
7. $\alpha(\beta\mathbf{x}) = \alpha\beta\mathbf{x}$
8. $1\mathbf{x} = \mathbf{x}$

1.3. Definición de matriz

Se denomina una matriz $\mathbf{A}$ de dimensiones $n \times m$ (n por m), denotada por $\mathbf{A}_{n \times m}$, a un arreglo bidimensional de números reales A_{ij}, llamadas componentes de la matriz. La representación de la matriz es de la forma

$$\mathbf{A} = [A_{ij}] = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \tag{2}$$

En A_{ij}, el primer índice (i) indica la fila y el segundo índice (j) indica la columna de la matriz $\mathbf{A}$. Notar que los índices de la matriz son números naturales, es decir, $i, j \in I\!N$.

Las matrices tienen múltiples aplicaciones, entre los cuales se puede destacar, por ejemplo, para representar los coeficientes de un sistema de ecuaciones lineales o para definir transformaciones lineales de un sistema de coordenadas a otro.

En particular, un vector de dimensión n se puede entender como una matriz de dimensiones $n \times 1$, con la salvedad que en el caso de matrices se puede hablar de un **vector columna** $\mathbf{a}$ o un **vector fila** $\mathbf{b}$ de dimensión n como se muestra a continuación,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} \leftarrow \text{vector columna;} \qquad \mathbf{b} = [b_1, b_2, b_3, \ldots, b_n] \quad \leftarrow \text{vector fila} \tag{3}$$

1.4. Operaciones básicas

1.4.1. Producto interno entre dos vectores

Sean los vectores $\mathbf{u}$, $\mathbf{v} \in I\!R^n$. Se define el producto interno entre los vectores $\mathbf{u}$ y $\mathbf{v}$ como la función que toma dos vectores de $\mathbf{V} \times \mathbf{V}$y entrega un valor en la recta real $I\!R$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{k=1}^{n} u_k v_k \tag{4}$$

Este producto entre dos vectores se conoce también como "producto escalar", o "producto punto". Notar que el producto interno es conmutativo, es decir, que $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ y que el producto interno entre dos vectores se puede interpretar también como el tamaño de la proyección de un vector en otro.

1.4.2. Norma de un vector

Se define la norma euclidiana de un vector $\mathbf{a} \in I\!R^n$ como:

$$\|\mathbf{a}\| = \sqrt{\sum_{k=1}^{n} a_k^2} \tag{5}$$

La norma euclidiana de un vector define el "tamaño" o longitud de un vector.

1.4.3. Ángulo entre dos vectores

El ángulo θ entre dos vectores $\mathbf{a}, \mathbf{b} \in I\!R^n$, se define como:

$$cos(\theta) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\|\|\mathbf{b}\|} \qquad \text{ó} \qquad \theta = \text{acos}\left(\frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\|\|\mathbf{b}\|}\right) \tag{6}$$

donde cos() es la función trigonométrica coseno, y acos() es la función arcoseno (la función inversa de la función coseno).

1.4.4. Producto cruz o producto vectorial entre dos vectores

Sean los vectores $\mathbf{u}$, $\mathbf{v} \in I\!R^3$. Se define el producto vectorial o producto cruz entre los vectores $\mathbf{u}$ y $\mathbf{v}$ como la función $\mathbf{w}$ que toma dos vectores de $\mathbf{V} \times \mathbf{V}$ y entrega un vector en $\mathbf{V}$, tal que:

$$\mathbf{w} = \mathbf{u} \wedge \mathbf{v} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix} \tag{7}$$

Se tiene que el producto vectorial no es conmutativo. De hecho, es válida la siguiente identidad:

$$(\mathbf{u} \wedge \mathbf{v}) = -(\mathbf{v} \wedge \mathbf{u}) \tag{8}$$

1.4.5. Identidades que involucran el producto interno y el producto vectorial

Sean los vectores $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in I\!R^3$, se tiene las siguientes identidades y/o propiedades:

1. Si $\mathbf{a}$ y $\mathbf{b}$ son dos vectores ortogonales (i.e., el ángulo entre los dos vectores es de 90 grados sexagesimales), se tiene que $< \mathbf{a}, \mathbf{b} >= 0$
2. $(\mathbf{u} \wedge \mathbf{v}) = -(\mathbf{v} \wedge \mathbf{u})$
3. $\langle \mathbf{a}, (\mathbf{a} \wedge \mathbf{b}) \rangle = 0$, es decir, el producto vectorial entre dos vectores es ortogonal a ambos vectores
4. Si $\mathbf{a} \wedge \mathbf{b} = \mathbf{0}$ con $\mathbf{a} \neq \mathbf{0}$ y $\mathbf{b} \neq \mathbf{0}$, esto es equivalente a que los vectores $\mathbf{a}$ y $\mathbf{b}$ son paralelos entre sí (i.e., el ángulo entre ellos es cero)
5. $\mathbf{a} \wedge \mathbf{a} = \mathbf{0}$ $\forall \mathbf{a} \in I\!R^n$
6. El producto vectorial no es asociativo, es decir, $(\mathbf{a} \wedge \mathbf{b}) \wedge \mathbf{c} \neq \mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{c})$
7. El producto vectorial es asociativo con respecto a un factor escalar $\lambda \in I\!R$, es decir, se tiene la identidad $\lambda(\mathbf{a} \wedge \mathbf{b}) = (\lambda\mathbf{a}) \wedge \mathbf{b} = \mathbf{a} \wedge (\lambda\mathbf{b})$
8. $\|\mathbf{a} \wedge \mathbf{b}\| = \sqrt{\|\mathbf{a}\|^2\|\mathbf{b}\|^2 - (\langle \mathbf{a}, \mathbf{b} \rangle)^2}$
9. $\|\mathbf{a} \wedge \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\| sin\theta$ donde θ es el ángulo menor entre los vectores $\mathbf{a}$ y $\mathbf{b}$

1.4.6. Suma de matrices

La suma de matrices se define para dos matrices de igual dimensiones $n \times m$. Sean $\mathbf{A}_{n\times m}$ y $\mathbf{B}_{n\times m}$ dos matrices de igual dimensiones, la matriz $\mathbf{C}$ que resulta de la suma de $\mathbf{A}$ y $\mathbf{B}$ tiene dimensiones $n \times m$ y se define como la suma de las componentes respectivas de las matrices, es decir, $\mathbf{C} = \mathbf{A} + \mathbf{B}$ es tal que:

$$C_{ij} = A_{ij} + B_{ij} \tag{9}$$

donde $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, m$.

La operación se define de manera similar para la suma de dos vectores, en que el vector resultante es la suma de las componentes respectivas de los vectores sumados.

1.4.7. Multiplicación de matrices

Sean las matrices $\mathbf{A}_{n\times m}$ y $\mathbf{B}_{m\times q}$, se define la multiplicación de las matrices $\mathbf{A}_{n\times m}$ y $\mathbf{B}_{m\times q}$ como la matriz de dimensiones:

$$\mathbf{C}_{n\times q} = \mathbf{A}_{n\times m}\mathbf{B}_{m\times q} \qquad \text{donde} \qquad c_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj} \tag{10}$$

donde el número de columnas de $\mathbf{A}$ es igual al número de filas de $\mathbf{B}$, ambos m, y $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, q$.

Notar que para el caso de la multiplicación de una matriz $\mathbf{A}_{n\times m}$ por un vector columna $\mathbf{b}_{m\times 1}$, se necesita que la cantidad de columnas de $\mathbf{A}$ sea igual a la dimensión del vector $\mathbf{b}$ para hacer la multiplicación $\mathbf{Ab}$.

Además, notar que la multiplicación de matrices no es conmutativa, es decir, si se tiene las matrices $A_{n\times m}$ y $B_{m\times n}$, en general $(\mathbf{AB})_{n\times n} \neq (\mathbf{BA})_{m\times m}$. La igualdad sólo se cumple en el caso particular donde las matrices $\mathbf{A}$ y $\mathbf{B}$ son cuadradas, de igual dimensión (i.e., $n = m$) y ambas son matrices simétricas (ver definición más adelante).

1.4.8. Matriz diagonal

Una matriz diagonal es una matriz cuadrada en que sólo los elementos de su diagonal son distintos de cero. Por ejemplo, una matriz diagonal de dimensión 4×4 es la siguiente:

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ 0 & 0 & A_{33} & 0 \\ 0 & 0 & 0 & A_{44} \end{bmatrix} \tag{11}$$

1.4.9. Matriz Identidad

Es una matriz diagonal, denotada por la letra $\mathbf{I}$, en que los elementos de la diagonal son todos iguales a la unidad (i.e., el número entero 1). Por ejemplo, una matriz identidad de 4×4 es la siguiente:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

La matriz identidad tiene la siguiente propiedad: para una matriz $\mathbf{A}_{n\times n}$, vector $\mathbf{x}_{n\times 1}$ y matriz identidad $\mathbf{I}_{n\times n}$ se cumple que:

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \tag{13}$$

$$\mathbf{xI} = \mathbf{Ix} = \mathbf{x} \tag{14}$$

1.4.10. Traspuesta de una matriz

La traspuesta de una matriz $\mathbf{A}_{n\times m}$, denotada por $\mathbf{A}^{\mathbf{T}}_{\mathbf{m\times n}}$, esta dada por:

$$(\mathbf{A^T})_{ij} = A_{ji} \qquad \text{donde } i = 1, \ldots, n; \quad j = 1, \ldots, m \tag{15}$$

Por ejemplo:

$$Si \quad \mathbf{A} = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \quad \Rightarrow \quad \mathbf{A^T} = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \tag{16}$$

Una propiedad de la trasposición de matrices: dadas las matrices $\mathbf{A}_{n\times m}$ y $\mathbf{B}_{m\times q}$, se tiene la siguiente propiedad

$$(\mathbf{AB})^{\mathbf{T}} = \mathbf{B^T A^T}$$

1.4.11. Matriz Simétrica

Una matriz $\mathbf{A}$ es simétrica si se tiene que $\mathbf{A^T} = \mathbf{A}$, es decir, que $A_{ji} = A_{ij}$.

1.4.12. Matriz Antimétrica

Una matriz $\mathbf{A}$ es antimétrica si se tiene que $\mathbf{A^T} = -\mathbf{A}$, es decir, que $A_{ji} = -A_{ij}$. Notar que los elementos de la diagonal de la matriz $\mathbf{A}$ son nulos si dicha matriz es antimétrica.

1.4.13. Matriz ortogonal

Una matriz $\mathbf{Q}_{n\times n}$ se dice matriz ortogonal, si

$$\mathbf{Q^T Q} = \mathbf{Q Q^T} = \mathbf{I}_{n\times n}$$

Se tiene la siguiente propiedad: dada una matriz ortogonal $\mathbf{Q}$ y un vector $\mathbf{x}$,

$$(\mathbf{Qx})^{\mathbf{T}}(\mathbf{Qx}) = \mathbf{x^T x}$$

Las matrices ortogonales son utilizadas para definir rotaciones del sistema de coordenadas.

1.4.14. Determinante de una matriz

El determinante de una matriz cuadrada $\mathbf{A}_{n\times n}$ es un número escalar denotado por $det(\mathbf{A})$ o $|A|$ en que para una matriz de 2×2 es:

$$det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc \tag{17}$$

Para una matriz $\mathbf{A}_{n\times n}$ se utiliza la fórmula de Laplace:

$$det\,(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} M_{ij} \qquad \text{para cualquier } i \in \{1, \ldots, n\} \tag{18}$$

donde M_{ij} es el determinante de la matriz que se obtiene al eliminar la fila i y la columna j de la matriz $\mathbf{A}$.

Por ejemplo, si se tiene una matriz $\mathbf{A}_{3\times 3}$

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \tag{19}$$

el determinante de dicha matriz se calcula como:

$$\det(\mathbf{A}) = \det\left(\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}\right) \tag{20}$$

$$= A_{11}\det\left(\begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}\right) - A_{12}\det\left(\begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix}\right) + A_{13}\det\left(\begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}\right) \tag{21}$$

$$= A_{11}(A_{22}A_{33} - A_{32}A_{23}) + A_{12}(A_{31}A_{23} - A_{21}A_{33}) + A_{13}(A_{21}A_{32} - A_{31}A_{22}) \tag{22}$$

1.4.15. Matriz Singular

Una matriz cuadrada es singular si su determinante es cero, es decir, $det(\mathbf{A}) = 0$.

1.4.16. Matriz definida positiva

Sea una matriz cuadrada $\mathbf{A}_{n\times n}$ y un vector columna $\mathbf{x}_{n\times 1}$. La matriz $\mathbf{A}$ es definida positiva si para todo $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x} \in I\!R^n$, se tiene que:

$$\mathbf{x^T A x} > \mathbf{0}$$

Las matrices definidas positivas no son matrices singulares.

1.4.17. Definición de Matriz Inversa

Sea la matriz cuadrada y no singular $\mathbf{A}_{n\times n}$. Se define la matriz inversa de $\mathbf{A}$, denotada por $\mathbf{A^{-1}}$, como la matriz que cumple la siguiente propiedad:

$$\mathbf{A A^{-1}} = \mathbf{A^{-1} A} = \mathbf{I_{n\times n}} \tag{23}$$

La matriz inversa se puede calcular usando la fórmula de los cofactores:

$$\mathbf{A^{-1}} = \frac{1}{det(\mathbf{A})}\mathbf{C^T} \tag{24}$$

donde $\mathbf{C}$ es la matriz de cofactores de $\mathbf{A}$, cuyos elementos son tales que:

$$C_{ij} = (-1)^{i+j} M_{ij}$$

y M_{ij} es el determinante de la matriz que se obtiene al eliminar la fila i y la columna j de la matriz $\mathbf{A}$.

Por ejemplo, la matriz inversa de una matriz $\mathbf{A}_{2\times 2}$ se calcula como sigue:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \Rightarrow \mathbf{A^{-1}} = \frac{1}{det(\mathbf{A})}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{25}$$

1.5. Resolución de sistemas de ecuaciones lineales

Sea el sistema lineal de ecuaciones algebraicas de n ecuaciones y n incógnitas representado por:

$$A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n = b_1 \quad (26)$$
$$A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n = b_2 \quad (27)$$
$$\ldots\ldots\ldots\ldots \quad (28)$$
$$A_{n1}x_1 + A_{n2}x_2 + \ldots + A_{nn}x_n = b_n \quad (29)$$

donde $x_1, x_2, \ldots, x_n$ son sus incógnitas.

El sistema de ecuaciones anterior se puede escribir de manera matricial de la siguiente forma:

$$\mathbf{Ax} = \mathbf{b}$$

donde se tiene la matriz:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \ldots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \ldots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{n1} & \mathbf{A}_{n2} & \ldots & \mathbf{A}_{nn} \end{bmatrix} \quad (\text{ una matriz cuadrada de } n \times n\ ). \quad (30)$$

y los vectores columnas de dimension n:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad ; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (31)$$

Si la matriz $\mathbf{A}$ es no singular, la solución del sistema de ecuaciones esta dada por:

$$\mathbf{x} = \mathbf{A^{-1}b}$$

Luego, la mayor dificultad para resolver el sistema de ecuaciones lineales está en encontrar los coeficientes de $\mathbf{A^{-1}}$, es decir, calcular la matriz inversa de $\mathbf{A}$. Para ello existen numerosas técnicas, tales como:

- Fórmula de los Cofactores
- Métodos de eliminación de Gauss
- Métodos iterativos para la estimación de $\mathbf{A^{-1}}$

Asimismo, existen técnicas que permiten resolver el sistema de ecuaciones lineales $\mathbf{Ax} = \mathbf{b}$ sin encontrar de manera explícita la matriz inversa de $\mathbf{A}$. El detalle de las técnicas para encontrar la solución al sistema de ecuaciones escapa el alcance de este tutorial y, por lo tanto, no serán explicadas (para mayor información ver material complementario NO obligatorio).

2. Estimación de Parámetros usando el Método de Mínimos Cuadrados

En las sección anterior se vieron definiciones de álgebra lineal y en particular el cómo resolver un sistema de ecuaciones lineales en que se tiene igual número de incógnitas y ecuaciones.

En esta sección se ejemplificará y discutirá dos metodologías para resolver sistemas de ecuaciones lineales en donde el número de ecuaciones es mayor que el número de incógnitas del problema a resolver, es decir, un sistema de ecuaciones sobredeterminado.

2.1. Problemas de estimación lineal

Los problemas de estimación lineal son aquellos en que se quiere determinar ciertos parámetros de un modelo y que se podrán expresar de manera matricial de la siguiente forma:

$$\mathbf{Gm} = \mathbf{d} \tag{32}$$

donde $\mathbf{G}$ es la matriz de diseño del problema, $\mathbf{m}$ es el vector de parámetros que se quiere estimar y $\mathbf{d}$ son las observaciones. En este caso se considerará que el sistema de ecuaciones tiene N_m incógnitas y $N_d > N_m$ ecuaciones, por lo que la matriz de diseño $\mathbf{G}$ no es cuadrada, sino que tiene N_d filas y N_m columnas.

A modo de ilustración, se analizará el problema de encontrar los valores de el coeficiente de intersección y de la pendiente de una recta mediante un proceso de ajuste del modelo matemático de la recta a puntos en el plano (X,Y) que provienen del resultado de un experimento físico.

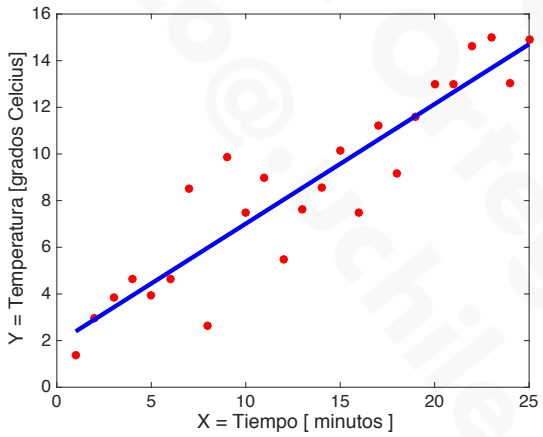

Figura 1: Ejemplo de ajuste de una recta a datos. Los puntos rojos son mediciones de temperatura de un objeto en tiempos determinados. La linea azul corresponde a la recta ajustada a dichas observaciones.

Suponer que se realiza un experimento en que se quiere estudiar la evolución de la temperatura de un cuerpo. Durante el procedimiento experimental se realizan mediciones de temperatura (denotada por y) a determinados instantes de tiempo (que denotamos x). Si se realizan N_d mediciones en total, se obtiene un conjunto de N_d pares

de valores $\{(x_i, y_i)\}_{i=1}^{N_d}$:

$$\begin{array}{rcl} (x_1 & , & y_1) \\ (x_2 & , & y_2) \\ (x_3 & , & y_3) \\ & \vdots & \\ (x_{N_d} & , & y_{N_d}) \end{array}$$

que corresponden a los puntos rojos en la Figura 1.

Basado en alguna razón física, se pretende ajustar una linea recta que modele la evolución de la temperatura del cuerpo en el tiempo. Para ello se plantea el modelo matemático de una recta:

$$y = a + bx \tag{33}$$

donde x representa el tiempo de la medición, y la temperatura medida en el experimento (los datos del experimento), y a, b representan las incógnitas , el coeficiente de intersección y la pendiente de la recta respectivamente.

A continuación, se explica como formar el sistema lineal de ecuaciones (32). Aquí se considera que cada medición debe obedecer a la ecuación de la recta (33), propuesta como modelo teórico o modelo físico para la evolución de la temperatura del cuerpo del experimento y que se quiere ajustar a los datos. Con esto se obtiene el siguiente sistema de ecuaciones:

$$\begin{array}{rcl} a + bx_1 & = & y_1 \\ a + bx_2 & = & y_2 \\ a + bx_3 & = & y_3 \\ & \vdots & \\ a + bx_{N_d} & = & y_{N_d} \end{array} \tag{34}$$

en donde cada ecuación se corresponde con una observación experimental (x_i, y_i). El sistema de ecuaciones anterior se puede describir de manera matricial de la forma $\mathbf{Gm} = \mathbf{d}$, cuyos elementos son los siguientes:

$$\mathbf{G} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{N_d} \end{bmatrix}; \qquad \mathbf{m} = \begin{bmatrix} a \\ b \end{bmatrix}; \qquad \mathbf{d} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{N_d} \end{bmatrix} \tag{35}$$

Se adelantará en esta sección que, si la matriz inversa de $\mathbf{G^T G}$ existe, la solución al sistema de ecuaciones lineales $\mathbf{Gm} = \mathbf{d}$ está dada por:

$$\mathbf{m} = \begin{bmatrix} a \\ b \end{bmatrix} = (\mathbf{G^T G})^{-1} \mathbf{G^T d}. \tag{36}$$

En al Figura 1 la línea azul es la recta en que se usa el coeficiente de intersección y la pendiente ajustada usando la ecuación (36), que corresponde a la solución por mínimos cuadrados del sistema de ecuaciones (32). Los detalles del método de mínimos cuadrados se discuten en la siguiente sección.

2.2. Método de Mínimos Cuadrados

En esta sección se muestra como obtener la solución del problema $\mathbf{Gm} = \mathbf{d}$ usando el método de mínimos cuadrados.

Se tiene el sistema de ecuaciones lineales $\mathbf{Gm} = \mathbf{d}$ en donde se quiere encontrar el valor de la incógnita $\mathbf{m}$. Por lo general se contará con más ecuaciones que incógnitas (como se ilustra en el ejemplo de la sección anterior), por lo que la matriz $\mathbf{G}$ no será una matriz cuadrada y no se puede resolver el problema simplemente calculando la matriz inversa de $\mathbf{G}$, ya que ésta no esta definida.

Aquí se plantea el problema de la siguiente manera. Primero, se define lo que corresponde a la predicción de los datos $\mathbf{d^{pred}}(\mathbf{m}) = \mathbf{Gm}$ para un vector $\mathbf{m}$ de parámetros del modelo teórico (e.g., la predicción de los valores de y en los tiempos x, para valores dados de coeficiente de intersección y pendiente de la recta en el ejemplo anterior).

Luego, usando el modelo teórico (i.e., la ecuación de la recta) se tiene una forma funcional para hacer una predicción de las observaciones dado un vector de parámetros. Luego, la idea es buscar el vector de parámetros $\mathbf{m}$ tal que la predicción de los datos sea lo más parecido posible al vector de datos observado a través de un procedimiento experimental ($\mathbf{d}$ - valores de temperatura medidos en el ejemplo de la sección anterior). Para ello se define el vector de error de ajuste de la predicción como el vector resultante de la diferencia entre la predicción de las observaciones para un vector de párametros dado y las observaciones:

$$\mathbf{e}(\mathbf{m}) = \mathbf{d} - \mathbf{d^{pred}}(\mathbf{m}) = \mathbf{d} - \mathbf{Gm} \tag{37}$$

El método de mínimos cuadrados busca como solución al problema de ecuaciones el vector $\mathbf{m}$ que minimiza la suma de las diferencias al cuadrado entre cada observación y la correspondiente predicción obtenida a través de $\mathbf{m}$, es decir, se busca resolver el problema de optimización:

$$\min_{\mathbf{m}} E(\mathbf{m}) \tag{38}$$

donde

$$\mathbf{E}(\mathbf{m}) = \sum_i e_i^2(\mathbf{m}) = \|\mathbf{e}(\mathbf{m})\|^2 = (\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm}) = \sum_i (\mathbf{d} - \mathbf{Gm})_i^2 \tag{39}$$

A $\mathbf{E}(\mathbf{m})$ se le denomina el error de ajuste. Aquí, el modelo estimado o vector de parámetros óptimo $\mathbf{m^{est}}$, es tal que $E(\mathbf{m^{est}}) \leq E(\mathbf{m})$ para cualquier vector de parámetros $\mathbf{m}$ posible.

La condición optimal para encontrar el modelo estimado $\mathbf{m}^{est}$ es:

$$\frac{\partial E(\mathbf{m})}{\partial m_q} = 0 \tag{40}$$

que en este caso asegura que el resultado obtenido sea un mínimo y no un máximo de $E(\mathbf{m})$ ya que la función que define el error de ajuste es una función convexa por construcción.

Se puede demostrar que imponiendo la condición optimal $\frac{\partial E(\mathbf{m})}{\partial m_q} = 0$ para encontrar el vector de parámetros $\mathbf{m}^{est}$ que minimiza el error de ajuste , se obtiene las *ecuaciones normales* del método de mínimos cuadrados:

$$\mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d} \tag{41}$$

Luego, resolver el sistema lineal $\mathbf{Gm} = \mathbf{d}$ por el método de mínimos cuadrados simples es equivalente a resolver las ecuaciones normales $\mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d}$. Notar que la matriz $\mathbf{G^TG}$ es una matriz cuadrada. Luego, cuando la matriz inversa de $\mathbf{G}^T\mathbf{G}$ existe, la solución al problema (41) se puede escribir de la siguiente manera:

$$\mathbf{m}^{est} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \tag{42}$$

Notar que el método de mínimos cuadrados falla cuando la matriz $\mathbf{G}^T\mathbf{G}$ no es invertible, lo que, por ejemplo, sucede cuando hay dos o más columnas de la matriz de diseño $\mathbf{G}$ que son linealmente dependientes, o cuando el número de ecuaciones es menor al número de incógnitas (i.e.,: el sistema de ecuaciones es subdeterminado). Cuando ocurre esta situación, la no existencia de $(\mathbf{G}^T\mathbf{G})^{-1}$ no quiere decir necesariamente que el problema inverso no tenga solución, y sólo implica que la metodología es incapaz de resolver el problema inverso planteado, posiblemente por la existencia de múltiples soluciones de éste.

Un factor igualmente importante al encontrar una solución del problema inverso, es poder calcular la incertidumbre asociada a ésta. De manera de saber que tan confiable o que tan acotados son los valores obtenidos (el rango de error o intervalos de confianza), se utilizará la fórmula de propagación de error para una relación lineal (esto se verá con más detalles en el repaso de probabilidades), la que se expresa como sigue:

$$\text{si} \quad \mathbf{m} = \mathbf{M}\mathbf{d} \quad \Rightarrow \quad \mathbf{C_m} = \mathbf{M}\mathbf{C_d}\mathbf{M^T} \tag{43}$$

donde $\mathbf{C_d}$ es la matriz de covarianza de las observaciones que representa los errores de las mediciones experimentales, $\mathbf{C_m}$ es la matriz de covarianza del vector de parámetros estimado, que representa los errores en la estimación de $\mathbf{m}$, y $\mathbf{M} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ según la fórmula (42).

Es posible demostrar que en la derivación del método de mínimos cuadrados planteado en este apunte, se asume que los errores de las observaciones siguen una distribución de probabilidades Normal (o Gaussiana) y que son independientes e idénticamente distribuidos. En el ejemplo de la sección anterior, esto se traduce en que cada medición (observación) de temperatura tiene asociada una varianza igual a σ_d^2, y por lo tanto, la matriz de covarianza de las observaciones es $\mathbf{C_d} = \sigma_d^2\mathbf{I}$ (donde $\mathbf{I}$ es la matriz identidad). Escrito de manera matricial, los errores de las mediciones quedan definidos a través de la siguiente matriz de covarianza:

$$\mathbf{C_d} = \begin{bmatrix} \sigma_d^2 & 0 & \dots & 0 \\ 0 & \sigma_d^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_d^2 \end{bmatrix} \leftarrow \text{matriz de covarianza diagonal} \tag{44}$$

Luego, usando la fórmula (43) se obtiene:

$$\begin{aligned} \mathbf{C_m} &= \mathbf{M}\mathbf{C_d}\mathbf{M^T} \\ &= \underbrace{(\mathbf{G^T}\mathbf{G})^{-1}\mathbf{G^T}}_{\mathbf{M}}\underbrace{\sigma_d^2\mathbf{I}}_{\mathbf{C_d}}\underbrace{\mathbf{G}(\mathbf{G^T}\mathbf{G})^{-1}}_{\mathbf{M^T}} \\ &= \sigma_d^2(\mathbf{G^T}\mathbf{G})^{-1}\underbrace{(\mathbf{G^T}\mathbf{G})(\mathbf{G^T}\mathbf{G})^{-1}}_{\mathbf{I}} \\ &= \sigma_d^2(\mathbf{G^T}\mathbf{G})^{-1} \end{aligned} \tag{45}$$

Notar que la matriz de covarianza de los parámetros estimados, si bien es una matriz cuadrada, no es necesariamente una matriz diagonal. El número de filas y columnas de la matriz es igual al número de parámetros que se estima (el número de incógnitas). En el caso particular del ejemplo de un ajuste de una recta a observaciones de medición de temperatura, se definió que el vector de parámetros (o incógnitas) es:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \text{coeficiente de intersección} \\ \text{pendiente} \end{bmatrix}. \tag{46}$$

En este caso, la matriz de covarianza de los valores de los parámetros estimados tendrá la siguiente forma:

$$\mathbf{C_m} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} \sigma_a^2 & C_{ab} \\ C_{ab} & \sigma_b^2 \end{bmatrix} \tag{47}$$

donde los valores de la diagonal son las varianzas σ_a^2, σ_b^2 del coeficiente de intersección (a) y la pendiente (b) de la ecuación de la recta (33).

Aquí se puede interpretar a σ_a y σ_b como los errores asociados a los valores estimados de a y b, con lo que se puede entender que, por ejemplo, si bien se ha encontrado un valor dado para a, debido a que éste tiene un error σ_a, es muy probable que la incógnita a pueda tomar valores entre $a - \sigma_a$ y $a + \sigma_a$, definiendo lo que se llama un intervalo de confianza de la incógnita a. En este contexto, es muy probable que b pueda tomar valores entre $b - \sigma_b$ y $b + \sigma_b$.

Los elementos fuera de la diagonal de la matriz $\mathbf{C_m}$ son las covarianzas de los parámetros estimados. Estos indican que tan fuerte es la relación o dependencia lineal entre los parámetros que se estiman. En el ejemplo de la recta, si $C_{ab} \neq 0$ indica que los errores del coeficiente de intersección a y de la pendiente de la recta b están relacionados linealmente, es decir, que no son independientes. La relación entre dos parámetros (o entre los errores asociados a la estimación de éstos) se puede cuantificar de mejor manera usando el coeficiente de correlación de Pearson

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \frac{C_{ij}}{\sigma_i \sigma_j} \qquad \text{para } i \neq j \tag{48}$$

el que tiene la propiedad

$$-1 \leq \rho_{ij} \leq 1 \tag{49}$$

Luego si dos parámetros a y b son tales que $\rho_{ab} > 0$ se dice que estos parámetros están positivamente correlacionados. En el caso de que $\rho_{ab} < 0$ se dice que estos parámetros están negativamente correlacionados, y en el caso de que $\rho_{ab} = 0$ se dice que los parámetros no están linealmente relacionados.[1]

2.3. Método de mínimos cuadrados con pesos

En la sección anterior se vió que el problema de mínimos cuadrados asume implícitamente que los errores de todas las observaciones son iguales, lo cual no es necesariamente correcto. Por ejemplo, las condiciones experimentales pueden cambiar entre una medición y otra, cambiando el error con el que se obtiene el valor de la medición.

Esto nos motiva a cambiar la forma en que se define el error de ajuste (39), reformulándolo como una "medida" de que tan bien la predicción de un vector de parámetros ajusta las observaciones, pero esta vez considerando los errores de las observaciones. Para ello se propone la siguiente definición del error de ajuste, en que la diferencia entre la predicción de un vector de parámetros y la medición u obervación es dividida por el error de dicha observación:

$$\mathbf{E}(\mathbf{m}) = \sum_i \frac{(\mathbf{d} - \mathbf{G}\mathbf{m})_i^2}{\sigma_i^2}. \tag{50}$$

donde i indica la observación (o ecuación) i-ésima y σ_i es el error con el que se obtiene dicha observación.

Como se vió en el ejemplo de ajustar una recta, se tiene que cada observación, medición o dato experimental va a generar una ecuación del sistema lineal de ecuaciones $\mathbf{Gm} = \mathbf{d}$ (ver ecuaciones (34) y (35)). Luego, el considerar la ecuación (50) como el error de ajuste, se traduce en que uno divide cada ecuación del sistema de ecuaciones $\mathbf{Gm} = \mathbf{d}$ por el error asociado a la observación que define la ecuación respectiva, es decir, uno obtiene el siguiente

[1] La interpretación del coeficiente de correlación de Pearson se verá con más de detalle en el repaso de probabilidades de la siguiente semana.

sistema de ecuaciones:

$$\begin{aligned} \frac{a+bx_1}{\sigma_1} &= \frac{y_1}{\sigma_1} \\ \frac{a+bx_2}{\sigma_2} &= \frac{y_2}{\sigma_2} \\ &\vdots \\ \frac{a+bx_N}{\sigma_N} &= \frac{y_N}{\sigma_N} \end{aligned} \tag{51}$$

en que cada ecuación se corresponde con una observación experimental (x_i, y_i) con el error σ_i asociado a la medición de y_i. El sistema de ecuaciones anterior se puede describir de manera matricial de la forma:

$$\mathbf{W}_d\mathbf{G}\mathbf{m} = \mathbf{W}_d\mathbf{d} \tag{52}$$

, donde:

$$\mathbf{G} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}; \quad \mathbf{m} = \begin{bmatrix} a \\ b \end{bmatrix}; \quad \mathbf{d} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix} \tag{53}$$

son las mismas matrices y vectores del sistema original de ecuaciones $\mathbf{Gm} = \mathbf{d}$ (generado usando las ecuaciones (34) y (35)), el cual se ha premultiplicado por una matriz de pesos $\mathbf{W}_d$ que es diagonal y donde los elementos de la diagonal son el recíproco de los errores de las mediciones,

$$\mathbf{W_d} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_N} \end{bmatrix} \leftarrow \text{matriz de pesos diagonal} \tag{54}$$

Notar que dada la definición de la matriz de pesos se tiene que $\mathbf{W^T}_d\mathbf{W}_d = \mathbf{C}_d^{-1}$, donde

$$\mathbf{C_d} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix} \leftarrow \text{matriz de covarianza diagonal} \tag{55}$$

es la matriz de covarianza de las observaciones o mediciones $\mathbf{d}$.

Luego, si hacemos el siguiente cambio de variables,

$$\tilde{\mathbf{G}} = \mathbf{W}_d\mathbf{G} \tag{56}$$
$$\tilde{\mathbf{d}} = \mathbf{W}_d\mathbf{d} \tag{57}$$

obtenemos otro sistema de ecuaciones:

$$\tilde{\mathbf{G}}\mathbf{m} = \tilde{\mathbf{d}} \tag{58}$$

que se resuelve usando el método de mínimos cuadrados simples a través las ecuaciones normales (41) en las nuevas variables, resultando el vector de parámetros estimados:

$$\mathbf{m}^{est} = \left(\tilde{\mathbf{G}}^T\tilde{\mathbf{G}}\right)^{-1}\tilde{\mathbf{G}}^T\tilde{\mathbf{d}} \tag{59}$$

Luego si se recuerda que $\mathbf{W^T}_d\mathbf{W}_d = \mathbf{C}_d^{-1}$, y además se reemplaza $\tilde{\mathbf{G}} = \mathbf{W}_d\mathbf{G}$, $\tilde{\mathbf{d}} = \mathbf{W}_d\mathbf{d}$, se puede escribir el vector de parámetros estimados en función de las variables originales como:

$$\begin{aligned}\mathbf{m}^{est} &= (\mathbf{G}^T \underbrace{\mathbf{W}_d^T\mathbf{W}_d}_{\mathbf{C}_d^{-1}} \mathbf{G})^{-1}\mathbf{G}^T \underbrace{\mathbf{W}_d^T\mathbf{W}_d}_{\mathbf{C}_d^{-1}}\mathbf{d} \\ &= \underbrace{(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}}_{\mathbf{M}}\mathbf{d} \\ &= \mathbf{M}\mathbf{d}\end{aligned} \tag{60}$$

que es la solución de $\mathbf{Gm} = \mathbf{d}$ usando el método de mínimos cuadrados con pesos.

Si se recuerda la fórmula (45) para calcular la matriz de covarianza de los parámetros estimados del modelo, $\mathbf{C_m} = \mathbf{M}\mathbf{C_d}\mathbf{M^T}$, la matriz de covarianza de los parametros estimados usando el método de mínimos cuadrados con pesos (ecuación (60)) queda como sigue:

$$\begin{aligned}\mathbf{C_m} &= \mathbf{M}\mathbf{C_d}\mathbf{M^T} \\ &= \underbrace{(\mathbf{G^T}\mathbf{C_d^{-1}}\mathbf{G})^{-1}\mathbf{G^T}\mathbf{C_d^{-1}}}_{\mathbf{M}}\mathbf{C_d}\underbrace{\mathbf{C_d^{-1}}\mathbf{G}(\mathbf{G^T}\mathbf{C_d^{-1}}\mathbf{G})^{-1}}_{\mathbf{M^T}} \\ &= (\mathbf{G^T}\mathbf{C_d^{-1}}\mathbf{G})^{-1}\end{aligned} \tag{61}$$

2.4. Ejemplo de Ajuste de una línea recta: la importancia de considerar los errores de las mediciones en el proceso de estimación de los parámetros

A modo de ilustrar la importancia de considerar los errores de las observaciones o mediciones experimentales en el proceso de estimación de parámetros se retomará el ejemplo del ajuste de la línea recta a las mediciones de temperatura. Aquí se compararán dos procesos de estimación de la pendiente y el coeficiente de intersección de la recta, uno en que se usa el método de mínimos cuadrados, en que se asume que todas las observaciones poseen el mismo error, y el método de mínimos cuadrados con pesos, en que se toma en cuenta los errores de cada medición.

Consideremos que se han realizado las mediciones de temperatura graficadas en la Figura 2. En esta figura se observa que hay dos mediciones anómalas, en el sentido de que "se salen de la tendencia del resto de las observaciones" debido a un problema en el instrumento con el que se mide la temperatura del cuerpo. El problema que causa las mediciones anómalas se ve reflejado en el error asociado a dichas observaciones.

A modo de comparación, realizamos el ajuste de la recta usando el método mínimos cuadrados simples, es decir, ecuaciones (34), (41) y (42) y usando el método de mínimos cuadrados con pesos, es decir, ecuaciones (51), (59) y (60). En la Figura 3 se presenta la recta ajustada a las observaciones usando el método de mínimos cuadrados simples (linea roja) y usando el método de mínimos cuadrados con pesos (linea verde).

Al analizar las rectas ajustadas usando ambos métodos ilustradas en Figura 3, se observa que la solución usando el método de mínimos cuadrados con pesos (linea verde) no se ve influenciada por las mediciones anómalas ya que el método toma en cuenta que éstas tienen un error mayor, representando bien las "buenas mediciones". Sin embargo, la recta ajustada con el método de mínimos cuadrados simples, al no considerar los errores de las mediciones trata de ajustar todas las observaciones, independientemente de que si son "buenas" (con error bajo) o "malas" (con error alto), produciendo un sesgo en la solución, lo que se traduce en que la recta ajustada no representa bien las mediciones experimentales. Luego es importante considerar no sólo el valor de una observación medida en un experimento sino que también el error o incerteza asociado a cada una de las mediciones.

Referencias

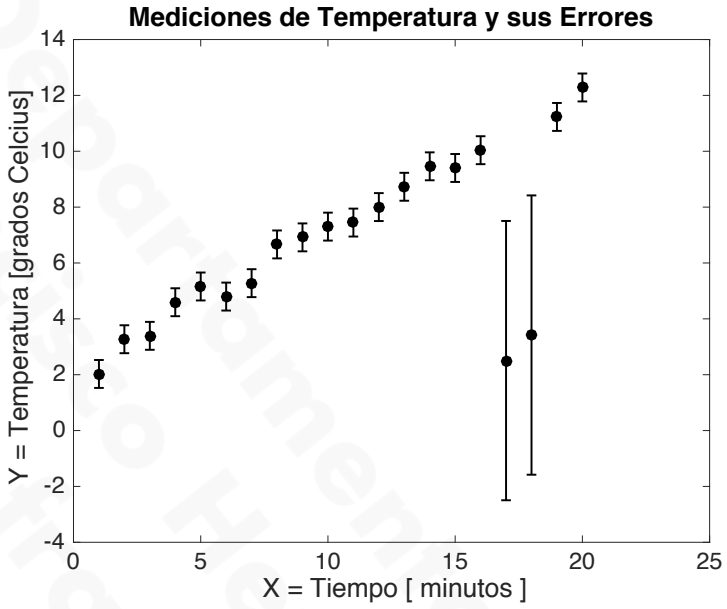

Figura 2: Ejemplo de ajuste mediciones experimentales de temperatura de un cuerpo en el tiempo. Las barras verticales indican el error asociado a cada medición. Notar que hay dos mediciones anómalas debido a problemas con el instrumento utilizado para obtenerlas, lo que se ve reflejado en los errores de dichas observaciones (barras de error más grandes).

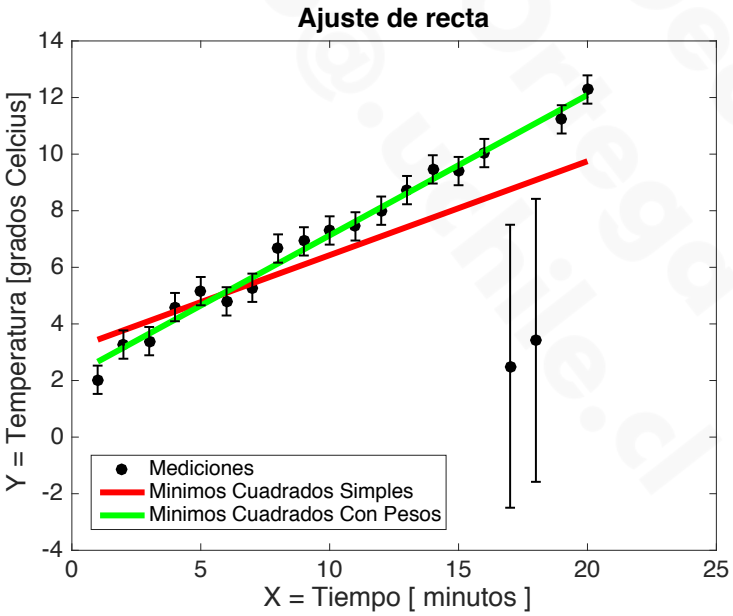

Figura 3: Gráfico de la recta ajustada a las mediciones experimentales. La línea roja corresponde a la recta ajustada usando el método de mínimos cuadrados simples y la línea verde corresponde a la recta ajustada usando el método de mínimos cuadrados con pesos. Notar como el método de mínimos cuadrados simples produce un resultado que es sesgado por las observaciones anómalas al no considerar los errores de cada medición por separado.