
Aproximación de funciones y regresión

PID_00266172

Álvaro Leitao Rodríguez

Tiempo mínimo de dedicación recomendado: 3 horas



Álvaro Leitao Rodríguez

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por la profesora: Teresa Sancho (2019)

Primera edición: septiembre 2019
© Álvaro Leitao Rodríguez
Todos los derechos reservados
© de esta edición, FUOC, 2019
Av. Tibidabo, 39-43, 08035 Barcelona
Realización editorial: FUOC

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Índice

Introducción	5
1. Aproximación de funciones por el método de mínimos cuadrados	10
1.1. Consideraciones previas	10
1.1.1. Normas y seminormas	10
1.1.2. Producto escalar	12
1.1.3. Sistemas ortogonales	13
1.2. Formulación del problema	13
1.3. Solución al problema de aproximación	14
1.4. Aproximación por polinomios ortogonales	17
1.4.1. Familia triangular de polinomios	17
1.4.2. Polinomios ortogonales	19
1.4.3. Polinomios de Chebyshev	20
1.4.4. Polinomios de Legendre	22
1.5. Métodos de Fourier	24
1.5.1. Conceptos básicos en análisis de Fourier	26
1.5.2. Series de Fourier	29
1.5.3. Teorema integral de Fourier	30
2. Regresión	33
2.1. Regresión lineal	33
2.1.1. Cuantificación del error de la regresión lineal	35
2.2. Linealización de relaciones no lineales	37
2.3. Regresión lineal polinómica	39
2.4. Regresión lineal múltiple	40
Bibliografía	41

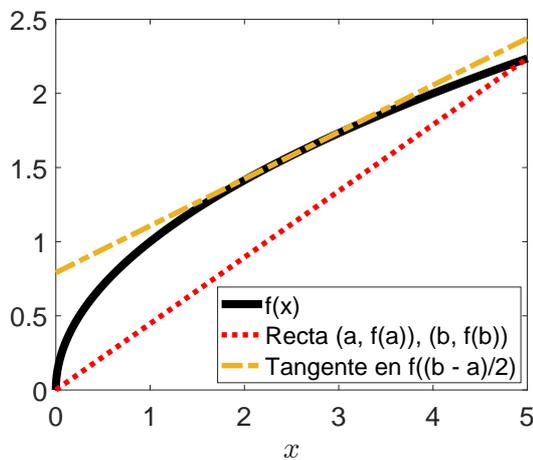
Introducción

Tal y como se ha comentado en la guía correspondiente a la actividad 2, existen muchas situaciones en las que se necesita aproximar una función, por ejemplo si esta es desconocida, conocida solo en algunos puntos o cuya definición es intrínsecamente complicada. El objetivo entonces es aproximar una función f por una función alternativa \hat{f} , miembro de una clase de funciones con la que es fácil trabajar matemáticamente (polinomios, funciones racionales, polinomios trigonométricos), donde cada función en dicha clase está definida por los valores numéricos de ciertos parámetros. Aquí nos restringiremos al problema de aproximar funciones de *una variable* en un intervalo cerrado.

Función definida en un intervalo $[a, b]$

En este ejemplo se presenta la curva de la función $f(x) = \sqrt{x}$ en el intervalo $[0,5]$. Además, incluimos dos posibles aproximaciones de f , dos de las más triviales, la recta que une los puntos de origen y fin de la función (en rojo) y la tangente en un punto ($x = \frac{b-a}{2} = 2,5$, en amarillo).

Figura 1

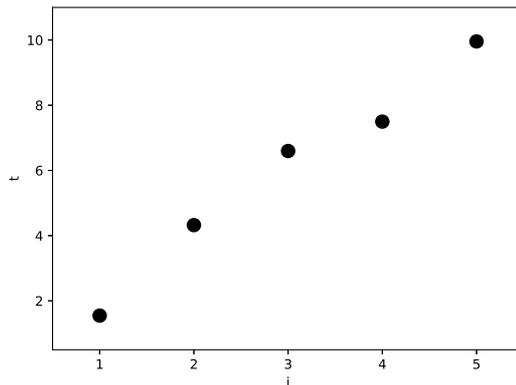


En tanto que hacemos una aproximación, hay diferencias entre la función original y la aproximada. Existen dos tipos de discrepancias que se deben tener en cuenta: discrepancias en los datos de entrada y discrepancias en el modelo (clase de funciones, forma, etc.) con el que queremos ajustar los datos de entrada. Denominaremos a estas discrepancias *error de medida* y *error en el modelo*, respectivamente.

Medición experimental

Los puntos en la figura 2 muestran los cinco tiempos de paso por el punto de equilibrio de un péndulo. Podemos llamar a cada uno de los tiempos como el tiempo i -ésimo de paso para $i = 1, 2, 3, 4, 5$. Las condiciones del experimento son tales que una relación de la forma $t = t_0 + iT$, con t_0 el tiempo inicial y T el tiempo que tarda el péndulo en cada ciclo, se considera válida, proporcionando alta precisión. En realidad, pueden aparecer algunos errores aleatorios en el procedimiento de medición y son los causantes de las desviaciones producidas con respecto a la linealidad (que sería lo esperado para un péndulo balanceándose). Estas desviaciones explican que los valores de los parámetros t_0 y T sean inciertos. Además tenemos cinco puntos y solo dos parámetros que determinar. Entonces el problema se llama **sobredeterminado**.

Figura 2. Tiempos de paso de un péndulo



Función definida como integral

Queremos evaluar, de la forma más eficiente posible, la función f definida como

$$f(x) = \int_0^x \exp(-t^2) dt,$$

para un $x \in [-1, 1]$, con un error relativo menor a 10^{-4} .

Esta integral no puede resolverse con papel y lápiz, mediante las técnicas clásicas de cálculo de primitivas, por lo que la función es complicada de evaluar en los puntos del intervalo dado. En este apartado veremos distintas formas de construir una aproximación que trate de minimizar el error del modelo y permita su evaluación sin tener que resolver la integral para cada evaluación.

Los ejemplos anteriores presentan dos casos distintos. En un caso tenemos un conjunto de puntos y queremos conocer valores intermedios, en el otro queremos evaluar una función definida mediante una expresión complicada. En la práctica, tanto un conjunto de datos como la utilización de funciones sencillas (típicamente polinomios) que sirven para “modelar” la función original son, en general, insuficientes, en el sentido de que un modelo es siempre una simplificación de la realidad y los datos pueden estar distorsionados (contener *ruido*). La aproximación de funciones puede ser vista como un caso especial de un problema más general e importante: ajustar un modelo matemático a unos datos dados y otros hechos conocidos.

Nos centraremos principalmente en el problema denominado **aproximación lineal**, esto es, una función desconocida f será aproximada usando una función \hat{f} que puede ser expresada como combinación lineal de otras.

Combinación lineal

Podemos ver un polinomio de grado 1 como una combinación lineal de la función constante 1 (que no aparece explícitamente) y la función x :

$$\hat{f}(x) = c_0 + c_1x,$$

Esta expresión aparece en muchas situaciones, como en el ejemplo del péndulo visto anteriormente o en la ecuación de la recta, que veremos en otro ejemplo.

La combinación lineal del ejemplo anterior se puede generalizar en dos direcciones: el número de sumandos o coeficientes y la forma de las correspondientes funciones. Así, la función \hat{f} (que aproxima linealmente a f), tendrá la forma genérica

$$\hat{f}(x) = c_0\phi_0(x) + c_1\phi_1(x) + \dots + c_n\phi_n(x),$$

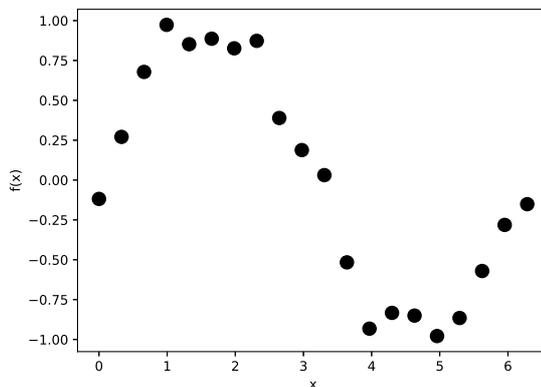
con $n + 1$ funciones $\phi_0, \phi_1, \dots, \phi_n$, que son elegidas previamente. Los coeficientes c_0, c_1, \dots, c_n son parámetros cuyo valor necesitar ser determinado para aproximar f . Por ejemplo, si $\phi_0(x) = 1, \phi_1(x) = x, \phi_2(x) = x^2, \dots, \phi_n(x) = x^n$ o, equivalentemente $\phi_i(x) = x^i$, la clase de posibles \hat{f} es la clase de polinomios de grado n . Entonces, el conjunto formado por $\{1, x, x^2, \dots, x^n\}$ se denomina **base** del conjunto de todos los polinomios de grado n .

Como ya hemos dicho, la función f puede venir dada de diferentes maneras. Una situación bastante común es que f es solo conocida a través de un conjunto de evaluaciones $f(x_0), f(x_1), \dots, f(x_m)$ de valores en una malla $x_0, x_1, x_2, \dots, x_m$. Asumiremos siempre que los $m + 1$ valores de la malla son distintos.

Función observada

En este ejemplo se presenta una función conocida solo por su valor en ciertos puntos, también llamadas observaciones.

Figura 3



Si por el contrario, la función f es originariamente expresada como una curva o a través de una fórmula complicada, la malla de puntos puede ser construida.

Recta que pasa por dos puntos

La ecuación de la recta viene dada por $y = \alpha + \beta x$. Como es conocido, el problema de hallar la expresión de una recta dados dos puntos tiene una única solución exacta, que además es fácilmente calculable. Este problema encaja en la formulación general de aproximación de funciones, simplemente tomando $\hat{f}(x) = y$, $n = 1$, $\phi_0 = 1$, $\phi_1 = x$, $c_0 = \alpha$ y $c_1 = \beta$. Asumiendo que los dos puntos dados tienen coordenadas $(x_0 = 1, y_0 = 1)$ y $(x_1 = 2, y_1 = 3)$, podemos escribir el sistema de ecuaciones,

$$\alpha + \beta = 1$$

$$\alpha + 2\beta = 3,$$

o, equivalentemente,

$$c_0 + c_1 = 1$$

$$c_0 + 2c_1 = 3,$$

de donde podemos deducir que, como ya sabíamos, la **pendiente** $\beta = \frac{y_1 - y_0}{x_1 - x_0}$ y α (respectivamente $c_1 = \frac{\hat{f}(x_1) - \hat{f}(x_0)}{x_1 - x_0}$ y c_0) se obtienen inmediatamente resolviendo el sistema: $\beta = c_1 = 2$ y $\alpha = c_0 = -1$.

Supongamos que queremos determinar los $n + 1$ coeficientes c_0, c_1, \dots, c_n de tal modo que $\hat{f}(x_i) = f(x_i)$ o que está lo más cerca posible, para todos los $m + 1$ puntos de la malla. Esta situación, considerando la aproximación lineal, nos llevaría a un sistema de ecuaciones lineal con $n + 1$ incógnitas y $m + 1$ ecuaciones,

$$\phi_0(x_0)c_0 + \phi_1(x_0)c_1 + \dots + \phi_n(x_0)c_n = f(x_0)$$

$$\phi_0(x_1)c_0 + \phi_1(x_1)c_1 + \dots + \phi_n(x_1)c_n = f(x_1)$$

⋮

$$\phi_0(x_m)c_0 + \phi_1(x_m)c_1 + \dots + \phi_n(x_m)c_n = f(x_m),$$

que, escrito en forma matriz-vector, es

$$(\phi_0(\mathbf{x}) | \phi_1(\mathbf{x}) | \dots | \phi_n(\mathbf{x})) \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} = f(\mathbf{x}), \tag{1}$$

donde $\mathbf{x} = (x_0, x_1, \dots, x_m)^T$, $\phi_j(\mathbf{x}) = (\phi_j(x_0), \phi_j(x_1), \dots, \phi_j(x_m))^T$ y $f(\mathbf{x}) = (f(x_0), f(x_1), \dots, f(x_m))^T$ son vectores columna.

Si $m = n$, entonces el sistema de ecuaciones tendrá, en condiciones normales, una única solución. En ese caso, esta forma de determinar \hat{f} se denomina **interpolación**, y se explica en la guía correspondiente a la actividad 2. La condición que se debe satisfacer para que la solución sea única es que los vectores $\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x})$ deben ser linealmente independientes. Esta condición también se puede expresar diciendo que las funciones ϕ_j deben ser linealmente independientes de la malla.

Si $m > n$, entonces solo en casos excepcionales se puede conseguir que $\hat{f}(x_i) = f(x_i)$ para todos los valores de la malla, x_i . El sistema tiene más ecuaciones que incógnitas, esto es, el sistema es **sobredeterminado**. En este caso, el sistema de ecuaciones se puede satisfacer solo parcialmente. Tal y como se menciona en la guía de la actividad 1, un método apropiado para resolver sistemas de ecuaciones sobredeterminados es, por ejemplo, la **descomposición o factorización QR**.

La **sobredeterminación** se usa para alcanzar dos tipos diferentes de suavizado en la aproximación:

- Para reducir el efecto de errores aleatorios en los valores de la función (filtrado).
- Para dar a la curva una forma más suave entre los valores de la malla (suavizado).

Una de las técnicas más utilizadas para resolver sistemas de ecuaciones lineales sobredeterminados es el **método de mínimos cuadrados**. Su aplicación a la aproximación de funciones implica una computación relativamente simple y, en muchas situaciones, puede estar motivada por argumentos estadísticos. El método será descrito en el apartado 1.

1. Aproximación de funciones por el método de mínimos cuadrados

Una forma de aproximar una función por otra es mediante los denominados “mínimos cuadrados”. En este caso se trata de encontrar la función \hat{f} más parecida a f de acuerdo con una medida concreta, precisamente la de “mínimos cuadrados”. Para definir esta medida, necesitaremos algunos resultados previos, que recordaremos brevemente, para formular después el problema de aproximación de una manera matemáticamente correcta.

A partir de este momento, diferenciaremos entre los casos donde la función que queremos aproximar sea continua y conocida por una expresión complicada (caso *continuo*) o expresada en términos de unos valores dados en una malla (caso *discreto*).

1.1. Consideraciones previas

1.1.1. Normas y seminormas

Una manera muy útil para estudiar el análisis de funciones es ver una función como un vector. Veremos que algunos tipos de funciones, como los polinomios, se pueden definir mediante “coordenadas”, lo que facilita su manejo.

Si consideramos el conjunto de todos los polinomios de grado n -ésimo, un polinomio concreto puede escribirse de la forma

$$p(x) = c_0 + c_1x + \dots + c_nx^n,$$

y está completamente determinado por todos sus coeficientes c_0, c_1, \dots, c_n . De esta manera, el polinomio $p(x)$ se puede escribir como $(c_0, c_1, \dots, c_n)^T$. Este vector tiene $n+1$ componentes y, por lo tanto, está en un espacio de dimensión $n+1$.

Parábola

Si consideramos la parábola

$$y = x^2,$$

se corresponde con un polinomio de grado $n = 2$ con los coeficientes $c_0 = 0, c_1 = 0$ y $c_2 = 1$. Entonces, el “vector” asociado es $(0,0,1)$ en un espacio de dimensión 3.

Fuente del apartado

Apartado basado en la siguiente obra: **Germund Dahlquist; Ake Bjorck** (2003). *Numerical Methods*. Dover Publications, Inc.

Medida

Definimos *medida* como una función que asigna un valor positivo (magnitud) a una realidad observable modelizada como objeto matemático. Por ejemplo, la longitud, el área y el volumen son tipos de medidas que empleamos regularmente en nuestra vida diaria.

En la guía correspondiente a la actividad 1 hemos visto cómo se definen y calculan las distancias entre vectores, particularmente utilizando las **normas** vectoriales. El concepto de norma también se puede utilizar cuando empleamos funciones en vez de vectores.

Como se ha especificado en la guía de la actividad 1, todas las normas tienen que cumplir una serie de propiedades que, adaptadas al contexto de las funciones, se concretan a continuación.

Un función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ se denomina **norma** si está definida en todo el espacio y cumple las condiciones siguientes:

- 1) $\|f\| \geq 0$ para todas las funciones f en el espacio.
- 2) $\|\alpha f\| = |\alpha| \cdot \|f\|$ para todo escalar real α .
- 3) $\|f+g\| \leq \|f\|+\|g\|$ para todo f y g en el espacio. Conocido como **desigualdad triangular**.
- 4) $\|f\| = 0 \Leftrightarrow f = 0$. En el caso particular de espacio de funciones continuas en el intervalo $[a,b]$, esta condición significa que $f(x) = 0$ en todo el intervalo $[a,b]$.

Las normas más comúnmente usadas para una función f continua en un intervalo cerrado $[a,b]$ son:

- **Norma infinito:** $\|f\|_\infty = \max_{x \in [a,b]} |f(x)|$.
- **Norma euclídea (L_2) o norma-2:** $\|f\|_2 = \sqrt{\int_a^b |f(x)|^2 dx}$.
- **Norma euclídea ponderada:** $\|f\|_{2,w} = \sqrt{\int_a^b |f(x)|^2 w(x) dx}$, donde w es una función peso continua y estrictamente positiva en $[a,b]$.

Las normas euclídea e infinito son casos especiales, $p = 2$ y $p \rightarrow \infty$ respectivamente, de una familia de normas denominadas **normas- p** , definidas

$$\|f\|_p = \sqrt[p]{\int_a^b |f(x)|^p dx}, \quad p \geq 1. \quad (2)$$

Cuando estamos en el caso discreto, la función desconocida viene dada por una malla de puntos, x_1, x_2, \dots, x_m , y las integrales se transforman en sumatorios. Por ejemplo, el equivalente discreto a la norma euclídea de $f(\mathbf{x})$ se expresaría como

$$\|f(\mathbf{x})\| = \sqrt{\sum_{j=0}^m |f(x)|^2}.$$

Notación $x \rightarrow \infty$

Cuando utilizamos la expresión $x \rightarrow \infty$, nos referimos a que la variable x se hace muy grande y se dice que x tiende a infinito.

Aunque se podría pensar en esta fórmula como una norma para la función f , estrictamente hablando, no lo es, ya que la condición 4 no se cumple. Esta expresión se conoce como **seminorma** de f con respecto al intervalo $[x_0, x_n]$. Usaremos la misma notación $\| \cdot \|$ para normas y seminormas, y cuando sea necesario usaremos subíndices para distinguirlas, por ejemplo $\| \cdot \|_2$ y $\| \cdot \|_{2,w}$, respectivamente.

Muchos de los métodos de aproximación están basados en el principio de minimizar alguna norma o seminorma del error $\hat{f} - f$, donde \hat{f} aproxima f y tiene una forma preestablecida. Es bastante importante que la norma sea elegida con respecto a la \hat{f} para la que se va a usar.

1.1.2. Producto escalar

Introducimos ahora un formalismo relacionado con las ideas geométricas mencionadas anteriormente, que será útil para estudiar la aproximación de mínimos cuadrados.

El **producto escalar** de dos funciones, f y g , reales y continuas, denotado por $\langle f, g \rangle$, se define por la relación

$$\langle f, g \rangle = \begin{cases} \int_a^b f(x)g(x)w(x)dx, & \text{en caso continuo,} \\ \sum_{i=0}^m f(x_i)g(x_i)w_i, & \text{en caso discreto.} \end{cases}$$

Producto escalar de vectores

Recordemos que el producto escalar de vectores se corresponde con la suma de los productos coordenada a coordenada de cada vector. Por ejemplo, en dos dimensiones $u \cdot v = (u_1, u_2) \cdot (v_1, v_2) = u_1 v_1 + u_2 v_2$.

En el caso discreto, si todos los pesos $w_j = 1$, entonces $\langle f, g \rangle$ se corresponde con el producto escalar de vectores de $f(\mathbf{x})$ y $g(\mathbf{x})$ (con la notación vectorial presentada en la introducción).

Al igual que en el caso de los productos escalares que aparecen en álgebra lineal, la definición genérica de $\langle f, g \rangle$ nos permite derivar reglas fundamentales asociadas al producto escalar de funciones. Sean f, g y ϕ funciones, y α y β números reales, tenemos las siguientes propiedades:

- **conmutatividad:** $\langle f, g \rangle = \langle g, f \rangle$
- **linealidad:** $\langle \alpha f + \beta g, \phi \rangle = \alpha \langle f, \phi \rangle + \beta \langle g, \phi \rangle$
- **positividad:** $\langle f, f \rangle \geq 0$

Es fácil deducir que $\langle f, f \rangle = \|f\|^2$, donde $\| \cdot \|$ denota la norma o seminorma euclídea ponderada en los casos continuo o discreto, respectivamente.

1.1.3. Sistemas ortogonales

Dos funciones f y g se dicen **ortogonales** si su producto escalar es cero, esto es $\langle f, g \rangle = 0$. Una secuencia finita o infinita de funciones $\phi_0, \phi_1, \dots, \phi_n$ forma un **sistema ortogonal** si $\langle \phi_i, \phi_j \rangle = 0$ para $i \neq j$ y $\|\phi_i\| \neq 0, \forall i$. Si además $\|\phi_i\| = 1$, para cualquier valor de i , entonces la secuencia se denomina **sistema ortonormal**.

Si $\langle f, g \rangle = 0$, esto es, f y g son ortogonales, entonces se puede formular el llamado **teorema de Pitágoras de las funciones**,

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2.$$

1.2. Formulación del problema

Vamos ahora a presentar más concretamente el problema que pretendemos solucionar. Como hemos dicho en la introducción, pretendemos aproximar una función f desconocida en el intervalo $[a, b]$ por otra \hat{f} . Además, la función \hat{f} tendrá la forma de combinación lineal escrita como

$$\hat{f}(x) = c_0\phi_0(x) + c_1\phi_1(x) + \dots + c_n\phi_n(x),$$

o, equivalentemente,

$$\hat{f}(x) = \sum_{j=0}^n c_j\phi_j(x),$$

donde tenemos $n + 1$ funciones $\phi_0(x), \phi_1(x), \dots, \phi_n(x)$ dadas. Los coeficientes c_0, c_1, \dots, c_n se tienen que obtener de tal forma que una norma (o seminorma) euclídea ponderada de la función de error $\hat{f} - f$ sea lo más pequeña posible, esto es, calcular los coeficientes de tal modo que

$$\|\hat{f} - f\|_{2,w}^2 = \int_a^b |\hat{f}(x) - f(x)|^2 w(x) dx, \quad \text{en el caso continuo,}$$

$$\|\hat{f} - f\|_{2,w}^2 = \sum_{i=0}^m |\hat{f}(x_i) - f(x_i)|^2 w_i, \quad \text{en el caso discreto,}$$

sean tan pequeñas como sea posible. Esta formulación se conoce como **problema de mínimos cuadrados**.

1.3. Solución al problema de aproximación

Se puede demostrar que, cuando $\phi_0, \phi_1, \dots, \phi_n$ son linealmente independientes, el problema de aproximación mediante mínimos cuadrados tiene solución única,

$$\hat{f} = \sum_{j=0}^n \hat{c}_j \phi_j, \quad (3)$$

donde los coeficientes \hat{c}_j son los que hacen mínima la norma euclídea.

La solución está además caracterizada por la propiedad de ortogonalidad, que implica que $\hat{f} - f$ tiene que ser ortogonal a todas las funciones ϕ_j . Así, si se cumple esto, entonces \hat{f} es una solución del problema de aproximación. Aplicando la definición de ortogonalidad, obtendríamos que las $n + 1$ igualdades

$$\left\langle \sum_{j=0}^n \hat{c}_j \phi_j - f, \phi_k \right\rangle = 0, \quad k = 0, 1, 2, \dots, n,$$

se deben cumplir. Además, de la propiedad de linealidad del producto escalar se deduce que

$$\left\langle \sum_{j=0}^n \hat{c}_j \phi_j, \phi_k \right\rangle = \sum_{j=0}^n \hat{c}_j \langle \phi_j, \phi_k \rangle.$$

Entonces, las condiciones anteriores son equivalentes a resolver el sistema de ecuaciones

$$\begin{aligned} \langle \phi_0, \phi_0 \rangle \hat{c}_0 + \langle \phi_0, \phi_1 \rangle \hat{c}_1 + \dots + \langle \phi_0, \phi_n \rangle \hat{c}_n &= \langle \phi_0, f \rangle \\ \langle \phi_1, \phi_0 \rangle \hat{c}_0 + \langle \phi_1, \phi_1 \rangle \hat{c}_1 + \dots + \langle \phi_1, \phi_n \rangle \hat{c}_n &= \langle \phi_1, f \rangle \\ &\vdots \\ \langle \phi_n, \phi_0 \rangle \hat{c}_0 + \langle \phi_n, \phi_1 \rangle \hat{c}_1 + \dots + \langle \phi_n, \phi_n \rangle \hat{c}_n &= \langle \phi_n, f \rangle, \end{aligned}$$

que, formulado de forma compacta,

$$\sum_{j=0}^n \langle \phi_j, \phi_k \rangle \hat{c}_j = \langle f, \phi_k \rangle, \quad k = 0, 1, 2, \dots, n. \quad (4)$$

Este sistema de ecuaciones se denomina **ecuaciones normales**. Resolviendo el sistema se determinan los coeficientes \hat{c}_j . En la guía de la actividad 1 se han presentado distintos métodos de resolución de sistemas.

Si nos encontramos en el caso discreto y la norma utilizada es la norma euclídea, esto es, los pesos son todos la unidad, $w_j = 1$, las ecuaciones normales se puede obtener de una forma directa, sin calcular uno por uno todos los productos escalares necesarios. Recordemos el sistema de ecuaciones en (1) visto en la introducción y denotamos

$$\Phi := (\phi_0(\mathbf{x}) | \phi_1(\mathbf{x}) | \dots | \phi_n(\mathbf{x})) \quad \text{y} \quad \mathbf{c} := (c_0, c_1, \dots, c_n)^T.$$

Entonces lo que queremos es resolver el sistema sobredeterminado $\Phi \mathbf{c} = f(\mathbf{x})$. Es fácil demostrar que podemos construir las ecuaciones normales correspondientes a este sistema con respecto a la norma euclídea simplemente haciendo la siguiente transformación:

$$\Phi^T \Phi \mathbf{c} = \Phi^T f(\mathbf{x}). \quad (5)$$

Un caso particular muy importante es cuando $\phi_0, \phi_1, \dots, \phi_n$ forman un sistema ortogonal. Entonces las ecuaciones normales se pueden resolver inmediatamente, ya que en cada ecuación los términos donde $j \neq k$ son cero. Teniendo esto en cuenta, cada ecuación normal se reduce a

$$\langle \phi_j, \phi_j \rangle \hat{c}_j = \langle f, \phi_j \rangle, \quad j = 0, 1, 2, \dots, n.$$

Despejando, los coeficientes son calculados de una manera muy sencilla, mediante la expresión

$$\hat{c}_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}, \quad j = 0, 1, 2, \dots, n. \quad (6)$$

Los coeficientes \hat{c}_j se denominan **coeficientes ortogonales** (ocasionalmente, coeficientes de Fourier).

Ecuaciones normales

Supongamos que queremos usar el método de mínimos cuadrados para ajustar una función de la forma $\hat{f}(x) = c_0 + c_1 x$ a unos datos medidos en un experimento y presentados en la tabla 1. Asumimos que todos los pesos son la unidad.

Tabla 1

x	$f(x)$
1	-2,1
3	-0,9
4	-0,6
6	0,6
7	0,9

Datos experimentales

Usando la notación anterior, tendríamos que $\phi_0(x) = 1$ y $\phi_1(x) = x$, con $m = 5$. Las ecuaciones normales entonces vienen dadas por

$$\langle \phi_0, \phi_0 \rangle \hat{c}_0 + \langle \phi_1, \phi_0 \rangle \hat{c}_1 = \langle f, \phi_0 \rangle$$

$$\langle \phi_0, \phi_1 \rangle \hat{c}_0 + \langle \phi_1, \phi_1 \rangle \hat{c}_1 = \langle f, \phi_1 \rangle.$$

Se necesitan los productos escalares de los vectores

$$\langle \phi_0, \phi_0 \rangle = \sum_{i=1}^5 \phi_0(x_i) \phi_0(x_i) = \sum_{i=1}^5 1 = 5,$$

$$\langle \phi_0, \phi_1 \rangle = \sum_{i=1}^5 \phi_0(x_i) \phi_1(x_i) = \sum_{i=1}^5 x_i = 21,$$

$$\langle \phi_1, \phi_0 \rangle = \sum_{i=1}^5 \phi_1(x_i) \phi_0(x_i) = \sum_{i=1}^5 x_i = 21,$$

$$\langle \phi_1, \phi_1 \rangle = \sum_{i=1}^5 \phi_1(x_i) \phi_1(x_i) = \sum_{i=1}^5 x_i^2 = 111,$$

$$\langle f, \phi_0 \rangle = \sum_{i=1}^5 f(x_i) \phi_0(x_i) = \sum_{i=1}^5 f(x_i) = -2,1,$$

$$\langle f, \phi_1 \rangle = \sum_{i=1}^5 f(x_i) \phi_1(x_i) = \sum_{i=1}^5 f(x_i) x_i = 2,7,$$

Entonces se obtiene el sistema

$$5\hat{c}_0 + 21\hat{c}_1 = -2,1$$

$$21\hat{c}_0 + 111\hat{c}_1 = 2,7,$$

cuya solución es $\hat{c}_0 = -2,542$ y $\hat{c}_1 = 0,5053$. El sistema resultante se podría haber obtenido directamente mediante la transformación (5).

Media como aproximación de funciones

Consideremos el caso $n = 0$, $\phi_0(x) = 1$. Recordemos que

$$\langle f, g \rangle = \sum_{i=0}^m w_i f(x_i) g(x_i).$$

Las “ecuaciones normales” en la ecuación (4) se reducen a una sola ecuación

$$\langle \phi_0, \phi_0 \rangle c_0 = \langle f, \phi_0 \rangle,$$

donde

$$c_0 = \frac{\sum_{i=0}^m w_i f(x_i)}{\sum_{i=0}^m w_i}.$$

El coeficiente c_0 es la *media ponderada* de los valores de la función. Si todos los pesos son iguales, entonces obtenemos

$$c_0 = \frac{\sum_{i=0}^m f(x_i)}{m+1},$$

con lo que vemos que el cálculo de la media es un caso especial del método de mínimos cuadrados.

Es importante destacar que, cuando $\phi_1, \phi_2, \dots, \phi_n$ forman un sistema ortogonal, el coeficiente de Fourier \hat{c}_j es independiente de n (ver ecuación (6)). La ventaja de esto es que se puede incrementar el número total de parámetros sin recalcularse ninguno de los previamente obtenidos. El uso de sistemas ortogonales es muy ventajoso, no sólo porque simplifican los cálculos sino también porque permiten evitar problemas numéricos que pueden aparecer cuando se resuelven sistemas de ecuaciones normales para un conjunto de funciones base no ortogonales.

1.4. Aproximación por polinomios ortogonales

Por una familia de polinomios ortogonales nos referimos a una familia triangular (que definiremos a continuación) de polinomios que es un sistema ortogonal con respecto a una función peso dada. Las **aproximaciones de funciones en términos de polinomios ortogonales** son muy útiles: son **fáciles de manipular**, tienen **buenas propiedades de convergencia** y proporcionan una **representación bien condicionada** de una función. La teoría de los polinomios ortogonales también constituye la base de métodos numéricos para muchos problemas que, en principio, puede parecer que tienen poco que ver con la aproximación de mínimos cuadrados (integración numérica, fracciones continuas o el problema algebraico de autovalores, por ejemplo).

Número de condición

El número de condición de una función mide cuánto afecta al valor de salida un cambio importante en los valores de entrada. Si el valor del número de condición es próximo a uno, se dice que la función está **bien condicionada**. Si, por el contrario, el número de condición es mucho más grande que uno, se dice que está **mal condicionada**.

1.4.1. Familia triangular de polinomios

Un polinomio de grado n es una función de la forma

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0.$$

El coeficiente c_n se llama *primer coeficiente*. Si $c_n \neq 0$, entonces el polinomio se denomina *polinomio genuino de grado n* . La clase de polinomio de grado n contiene todos los polinomios de grado menor como casos especiales. Una constante es un polinomio de grado cero.

Existen muchas formas de especificar un polinomio. Una de ellas es escribir los términos en forma de sumatorio, también llamado expansión, con lo que obtenemos

$$\sum_{j=0}^n c_j x^j.$$

Aunque la forma de este sumatorio puede parecer la más intuitiva, no es la única alternativa. Por ejemplo, se puede reescribir lo anterior de la forma

$$\sum_{j=0}^n d_j (x-h)^j,$$

para algún $h \neq 0$.

Matemáticamente, las expresiones anteriores son equivalentes pero, computacionalmente, trabajar con *valores redondeados* de los coeficientes puede suponer una gran diferencia. Si se quisiera usar la segunda forma para una aproximación polinómica en el intervalo $[a, b]$, entonces se debería seleccionar $h = \frac{1}{2}(a+b)$, esto es, el punto medio del intervalo.

Las dos formas presentadas son las más naturales, en las que primero podríamos pensar, pero existen muchas otras representaciones que son incluso más ventajosas. Pensemos en la secuencia de polinomios $\phi_0, \phi_1, \phi_2, \dots$ (finita o infinita)

$$\begin{aligned} \phi_0(x) &= c_{00} \\ \phi_1(x) &= c_{10} + c_{11}x \\ \phi_2(x) &= c_{20} + c_{21}x + c_{22}x^2 \\ &\vdots \\ \phi_n(x) &= c_{n0} + c_{n1}x + c_{n2}x^2 + \dots + c_{nn}x^n \\ &\vdots \end{aligned}$$

donde todos los primeros coeficientes son distintos de cero, esto es $c_{ii} \neq 0$. Esta representación se conoce como **familia triangular** de polinomios. Como veremos, un polinomio escrito como familia triangular facilita su tratamiento y evaluación.

Utilizando la forma triangular, las potencias de x pueden ser expresadas recursiva y únicamente como combinación lineal de $\phi_0, \phi_1, \phi_2, \dots$,

$$x^0 = \frac{\phi_0(x)}{c_{00}}, \quad x^1 = \frac{\phi_1(x) - c_{10}}{c_{11}} = \frac{\phi_1(x) - c_{10} \frac{\phi_0(x)}{c_{00}}}{c_{11}}, \dots$$

1.4.2. Polinomios ortogonales

En este apartado definiremos formalmente los polinomios ortogonales. En los siguientes apartados introduciremos algunos ejemplos representativos de polinomios ortogonales, como los **polinomios de Chebyshev** y los **polinomios de Legendre**.

Cualquiera que sea la distribución de pesos para el producto escalar, existe un sistema ortogonal asociado $\phi_0, \phi_1, \phi_2, \dots$, que es una familia triangular de polinomios. La familia está determinada de manera única, independientemente de que los primeros coeficientes $A_0 = c_{00}, A_1 = c_{11}, A_2 = c_{22}, \dots$ de los polinomios de la familia triangular puedan venir dados por valores arbitrarios distintos de cero.

Dada una distribución de pesos en una malla con $m + 1$ puntos, la familia acaba con $\phi_m(x)$. La función $\phi_{m+1}(x)$ es cero en cada punto de la malla. En el caso continuo, la familia tiene infinitos miembros.

Para $n \geq 0$, los polinomios ortogonales satisfacen un fórmula recursiva de tres términos,

$$\phi_{n+1}(x) = \alpha_n(x - \beta_n)\phi_n(x) - \gamma_n\phi_{n-1}(x), \quad \phi_{-1}(x) = 0, \quad \phi_0(x) = A_0, \quad (7)$$

donde $\alpha_n = \frac{A_{n+1}}{A_n}$, y β_n y γ_n vienen dadas por las expresiones

$$\beta_n = \frac{\langle x\phi_n, \phi_n \rangle}{\|\phi_n\|^2}, \quad \gamma_n = \frac{\alpha_n \|\phi_n\|^2}{\alpha_{n-1} \|\phi_{n-1}\|^2}.$$

En el cálculo de los coeficientes en una expansión de la forma

$$p(x) = c_0\phi_0(x) + c_1\phi_1(x) + \dots + c_n\phi_n(x), \quad (8)$$

usando la fórmula del coeficiente ortogonal $c_j = \langle p, \phi_j \rangle / \|\phi_j\|^2$, se puede usar la fórmula recursiva de la ecuación 7. En el caso discreto, es aconsejable calcular recursivamente los vectores $\phi_j(\mathbf{x}), j = 0, 1, 2, \dots, n$.

El modo más fácil de evaluar una función definida mediante una expansión ortogonal es usar la **fórmula recursiva de Clenshaw**. Utilizando la notación de la ecuación 7, tenemos, para $k = n, n - 1, \dots, 1, 0$,

$$y_k = \alpha_k(x - \beta_k)y_{k+1} - \gamma_{k+1}y_{k+2} + c_k,$$

donde $y_{n+2} = y_{n+1} = 0$. Al final obtenemos que $p(x) = A_0y_0$.

1.4.3. Polinomios de Chebyshev

Los polinomios de Chebyshev constituyen una de las familias triangulares de polinomios ortogonales más utilizadas. Sus propiedades pueden ser deducidas con bastante facilidad, incluso sin la teoría enunciada en el apartado anterior.

Consideremos la fórmula

$$\cos((n+1)\phi) + \cos(n-1)\phi = 2 \cos(\phi) \cos(n\phi), \quad n \geq 1.$$

Esta fórmula nos permite expresar $\cos(n\phi)$ como un polinomio en términos de $\cos(\phi)$,

$$\cos(2\phi) = 2 \cos^2(\phi) - 1$$

$$\cos(3\phi) = 2 \cos(\phi) \cos(2\phi) - \cos(\phi) = 4 \cos^3(\phi) - 3 \cos(\phi)$$

$$\cos(4\phi) = 2 \cos(\phi) \cos(3\phi) - \cos(2\phi) = 8 \cos^4(\phi) - 8 \cos^2(\phi) + 1$$

⋮

Si asumimos que $x = \cos(\phi)$, siendo $\phi = \arccos(x)$, entonces obtenemos una familia triangular de polinomios, los polinomios de Chebyshev, para $-1 \leq x \leq 1$, $n = 0, 1, 2, \dots$ definidos por la fórmula

$$T_n(x) = \cos(n \arccos(x)).$$

Así, utilizando las expresiones anteriores para $\cos(n\phi)$, obtenemos

$$T_0(x) = 1,$$

$$T_1(x) = x,$$

$$T_2(x) = 2x^2 - 1,$$

$$T_3(x) = 4x^3 - 3x,$$

$$T_4(x) = 8x^4 - 8x^2 + 1$$

⋮

Propiedades de los polinomios de Chebyshev

1) Fórmula recursiva:

$$T_0(x) = 1,$$

$$T_1(x) = x,$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n \geq 1.$$

2) **Primer coeficiente:** Es 2^{n-1} para $n \geq 1$ y 1 para $n = 0$.

3) Simetría:

$$T_n(-x) = (-1)^n T_n(x).$$

4) **Ceros y puntos críticos:** $T_n(x)$ tiene n ceros en el intervalo $[-1,1]$, que vienen dados por

$$x_k = \cos\left(\frac{2k+1}{n} \frac{\pi}{2}\right), \quad k = 0, 1, 2, \dots, n-1,$$

conocidos como **abscisas de Chebyshev**, y $n+1$ puntos críticos,

$$x'_k = \cos\left(\frac{k\pi}{n}\right), \quad T_n(x'_k) = (-1)^k, \quad k = 0, 1, 2, \dots, n$$

Estos resultados se deducen directamente por el hecho de que $\cos(n\phi) = 0$ para

$$\phi = \frac{2k-1}{n} \frac{\pi}{2},$$

y que $|\cos(n\phi)|$ tiene máximos en $\phi = k\pi/n$.

5) **Ortogonalidad, caso continuo:** los polinomios de Chebyshev son ortogonales con respecto al producto escalar

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x)(1-x^2)^{-\frac{1}{2}} dx.$$

Entonces,

$$\langle T_i, T_j \rangle = \begin{cases} 0 & \text{si } i \neq j \\ \frac{1}{2}\pi & \text{si } i = j \neq 0 \\ \pi & \text{si } i = j = 0 \end{cases}$$

6) Propiedad de ortogonalidad, caso discreto: sea

$$\langle f, g \rangle = \sum_{k=0}^m f(x_k)g(x_k),$$

donde $\{x_k\}$ son los ceros de $T_{m+1}(x)$. Entonces, para $0 \leq i \leq m$ y $0 \leq j \leq m$, tenemos

$$\langle T_i, T_j \rangle = \begin{cases} 0 & \text{si } i \neq j \\ \frac{1}{2}(m+1) & \text{si } i = j \neq 0 \\ m+1 & \text{si } i = j = 0 \end{cases}$$

7) Propiedad minimax: de todos los polinomios de grado n con primer coeficiente 1, el polinomio $2^{1-n}T_n$ tiene la norma del máximo más pequeña en $[-1, 1]$. El valor de su norma del máximo es 2^{1-n} .

Las expansiones en términos de polinomios de Chebyshev son una ayuda importante en el estudio de funciones en el intervalo $[-1, 1]$. Si queremos trabajar en términos de una variable t que vive en el intervalo $[a, b]$, entonces se debe hacer el cambio de variables (sustitución),

$$t = \frac{1}{2}(a + b) + \frac{1}{2}(b - a)x, \quad t \in [a, b] \leftrightarrow x \in [-1, 1].$$

1.4.4. Polinomios de Legendre

Los polinomios de Legendre se definen mediante la fórmula

$$\begin{aligned} L_0(x) &= 1, \\ L_n(x) &= \frac{1}{2^n n!} \frac{d^n}{dx^n} \left((x^2 - 1)^n \right). \end{aligned} \tag{9}$$

Nota

La derivada n -ésima de una función, que describimos $f(x)d^n/dx^n(f(x))$ es la derivada de la derivada de la derivada... n veces.

Propiedades de los polinomios de Legendre

1) Fórmula recursiva:

$$\begin{aligned}
 L_0(x) &= 1, \\
 L_1(x) &= x, \\
 L_2(x) &= \frac{1}{2}(3x^2 - 1), \\
 L_{n+1}(x) &= \frac{2n+1}{n+1}xL_n(x) - \frac{n}{n+1}L_{n-1}(x)
 \end{aligned}
 \tag{10}$$

2) Primer coeficiente: como $(x^2-1)^n$ es un polinomio de grado $2n$, $L_n(x)$ es un polinomio de grado n . El primer coeficiente A_n es mismo que el del polinomio,

$$\frac{1}{2^n n!} \frac{d^n}{dx^n} (x^{2n}) = \frac{1}{2^n n!} 2n(2n-1)(2n-2) \dots (n+1)x^n$$

Entonces

$$\alpha_n = \frac{A_{n+1}}{A_n} = \frac{2n+1}{n+1}$$

3) Simetría:

$$L_n(-x) = (-1)^n L_n(x).$$

4) Ortogonalidad: los polinomios de Legendre son ortogonales con respecto al producto escalar,

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x)dx,$$

por lo que,

$$\langle L_n, L_j \rangle = \begin{cases} 0 & \text{si } n \neq j \\ \frac{2}{2n+1} & \text{si } n = j \end{cases}$$

5) Cota:

$$|L_n(x)| \leq 1, \quad x \in [-1, 1].$$

Observando la propiedad de ortogonalidad, los polinomios de Legendre son bastante útiles para aproximaciones con norma euclídea y distribución de peso $w(x) = 1$.

1.5. Métodos de Fourier

En este apartado vamos a estudiar una aplicación particular de la teoría de aproximación de funciones vista hasta ahora. Se trata de un marco teórico altamente desarrollado, el *análisis de Fourier*, en el que se enmarcan los llamados polinomios trigonométricos. Estos son una herramienta con mucho potencial en la aproximación de funciones, especialmente *funciones periódicas* (aunque no solo estas). El análisis de Fourier utiliza muchos de los resultados generales de los polinomios ortogonales.

Muchos fenómenos naturales (acústicos, ópticos, etc.) presentan un comportamiento periódico. Es bien sabido que un sonido está compuesto por oscilaciones regulares: el tono fundamental (o puro), con una frecuencia n , y los tonos armónicos, con frecuencias $2n, 3n, 4n, \dots$. El ratio entre el tono fundamental y los tonos armónicos determina nuestra percepción del sonido. Pueden también existir sonidos sin tonos armónicos, es decir, basados solo en el tono fundamental, por ejemplo en la música electrónica, donde se utilizan tonos sinusoidales puros.

En un oscilador electrónico, una corriente eléctrica es generada con una fuerza que depende del tiempo t de acuerdo con la fórmula

$$r \sin(\omega t + \nu),$$

donde r es la llamada *amplitud* de oscilación, ω es la frecuencia angular (2π veces la frecuencia) y ν es una constante que define el estado del sonido a tiempo $t = 0$.

En un altavoz, las variaciones de corriente eléctrica son convertidas en variaciones de presión del aire, que, bajo condiciones ideales, se describen con la misma función. Sin embargo, en la práctica siempre se producen ciertas distorsiones que provocan que aparezcan los tonos armónicos. Así, armónicos de orden $k - 1$ contribuyen al sonido en la forma $r_k \sin(k\omega t + \nu_k)$. Las variaciones en la presión del aire que llegan a nuestros oídos se pueden describir, desde este punto de vista, como la suma

$$\sum_{k=0}^{\infty} r_k \sin(k\omega t + \nu_k). \quad (11)$$

La separación de un fenómeno periódico en un tono fundamental y varios tonos armónicos aparece no solo en acústica sino también en muchas otras áreas, y está relacionado con un teorema puramente matemático muy importante desarrollado por primera vez por Fourier. De acuerdo con este teorema, toda función $f(t)$ con periodo $2\pi/\omega$, bajo ciertas condiciones generales, se pue-

de expandir en una serie de la forma del sumatorio anterior (ecuación (11)). Más adelante daremos una formulación más precisa de este resultado.

Una función periódica es la que cumple que

$$f(t + p) = f(t)$$

Una expansión de la forma de la ecuación (11) se puede expresar de muchas maneras equivalentes. Si tomamos $a_k = r_k \sin(v_k)$ y $b_k = r_k \cos(v_k)$, por las propiedades del seno de una suma podemos escribir

$$f(t) = \sum_{k=0}^{\infty} (a_k \cos(k\omega t) + b_k \sin(k\omega t)), \quad (12)$$

donde a_k y b_k son constantes reales. En la práctica, el sumatorio se debe *truncar*, es decir, utilizar un número finito de sumandos, $n + 1$,

$$f(t) = \sum_{k=0}^n (a_k \cos(k\omega t) + b_k \sin(k\omega t)),$$

Otra forma que también puede resultar conveniente se puede obtener mediante la fórmula de Euler,

$$\cos(x) = \frac{\exp(ix) + \exp(-ix)}{2}, \quad \sin(x) = \frac{\exp(ix) - \exp(-ix)}{2i},$$

donde $i = \sqrt{-1}$ es la unidad imaginaria. Si, para $k = 1, 2, 3, \dots$, tenemos

$$\begin{aligned} c_k &= \frac{1}{2} (a_k - ib_k), \\ c_{-k} &= \frac{1}{2} (a_k + ib_k), \end{aligned} \quad (13)$$

entonces se obtiene

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \exp(ik\omega t), \quad (14)$$

o, utilizando el sumatorio truncado,

$$f(t) = \sum_{k=-n}^n c_k \exp(ik\omega t).$$

En lo que resta, utilizaremos el término **serie de Fourier** a una expansión de la forma de la ecuación (12) o la ecuación (14). Las formas truncadas de estas series se conocen como **polinomios trigonométricos**. Las series de Fourier son muy importantes en el estudio de fenómenos periódicos en el tiempo (vibraciones, sonido, luz, corriente alterna, etc.) o en el espacio (ondas, estructura cristalina, etc.).

1.5.1. Conceptos básicos en análisis de Fourier

Nos restringiremos ahora a funciones de una variable y periodo 2π . Si una función de t tiene periodo p , entonces el cambio de variables $x = 2\pi t/p$ transforma la función de t inicial en una función de x con periodo 2π . Asumimos que la función puede tener valores complejos, ya que la exponencial compleja resulta muy conveniente para ciertas manipulaciones algebraicas. El producto escalar de dos funciones que toman valores complejos, f y g de periodo 2π , se define, en el caso continuo, como

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)\bar{g}(x)dx, \quad (15)$$

y en el caso discreto,

$$\langle f, g \rangle = \sum_{\alpha=0}^M f(x_{\alpha})\bar{g}(x_{\alpha}), \quad x_{\alpha} = \frac{2\pi\alpha}{M+1}, \quad (16)$$

donde la barra horizontal encima de g indica conjugación compleja.

La norma de la función f se define como

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

Los resultados para el producto escalar vistos en los apartados anteriores aplican igualmente con esta nueva definición, con algunas modificaciones obvias. Nótese especialmente que $\langle g, f \rangle = \overline{\langle f, g \rangle}$. En el caso continuo no importa en qué intervalo nos encontremos; siempre y cuando tenga longitud 2π , el valor del producto escalar no cambia.

Sean las funciones (que aparecen en la ecuación (14)),

$$\phi_j(x) = \exp(ijx), \quad j = \dots, -2, -1, 0, 1, 2, \dots,$$

se tienen las siguientes relaciones de ortogonalidad,

$$\langle \phi_j, \phi_k \rangle = \begin{cases} 2\pi & \text{si } j = k, \\ 0 & \text{si } j \neq k, \end{cases}$$

para el caso continuo y,

$$\langle \phi_j, \phi_k \rangle = \begin{cases} M+1 & \text{si } \frac{j-k}{M+1} \text{ es un entero} \\ 0 & \text{en otro caso,} \end{cases}$$

para el caso discreto.

Si se sabe que f se puede expandir de la forma

$$f = \sum_{j=a}^b c_j \phi_j,$$

donde $a = -\infty$ y $b = \infty$ en el caso continuo, y $b-a = M$ en el discreto, entonces, formalmente tenemos, para $a \leq k \leq b$, que

$$\langle f, \phi_k \rangle = \sum_{j=a}^b c_j \langle \phi_j, \phi_k \rangle,$$

ya que $\langle \phi_j, \phi_k \rangle = 0$ para $j \neq k$. Así, cambiando k por j , tenemos

$$c_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle} = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-ijx) dx & \text{en el caso continuo} \\ \frac{1}{M+1} \sum_{\alpha=0}^M f(x_\alpha) \exp(-ijx_\alpha) & \text{en el caso discreto.} \end{cases} \quad (17)$$

Estos coeficiente se denominan **coeficientes de Fourier** y son un caso particular de los obtenidos en el apartado 1.3.

En consonancia con la ecuación (13),

$$a_j = c_j + c_{-j}, \quad b_j = i(c_j - c_{-j}).$$

Cuando $j = 0$, el coeficiente a_0 se calcularía sumando dos veces el mismo valor c_0 , lo que sería incorrecto. Por lo tanto, el coeficiente a_0 aparece siempre dividido por 2 para corregir esta discrepancia. Como se puede deducir fácilmente, $b_0 = 0$.

Por ejemplo, para el caso continuo, tendríamos que, para $j > 0$

$$\begin{aligned}
 a_j &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(jx) dx, \\
 b_j &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(jx) dx.
 \end{aligned}
 \tag{18}$$

Datos periódicos

En la tabla 2 presentamos las temperaturas registradas en Washington D.C. el 1 de enero de 2001. Elegimos un modelo que permita capturar las oscilaciones que observamos en los datos.

Tabla 2

Hora del día	Fracción de día, (t)	Temperatura
00:00	0	-2,2
03:00	$\frac{1}{8}$	-2,8
06:00	$\frac{1}{4}$	-6,1
09:00	$\frac{3}{8}$	-3,9
12:00	$\frac{1}{2}$	0,0
15:00	$\frac{5}{8}$	1,1
18:00	$\frac{3}{4}$	-0,6
21:00	$\frac{7}{8}$	-1,1

De acuerdo con lo visto en este apartado, el más sencillo en el que podemos pensar es $f(t) = \frac{a_0}{2} + a_1 \cos(2\pi t) + b_1 \sin(2\pi t)$, que se corresponde con la expresión (12) truncada a $n = 1$. Utilizamos este modelo porque sabemos que la temperatura es más o menos periódica con un período de 24 horas. Utilizando la notación del apartado tendríamos, $M = 7$ (8 datos) y $\alpha = 0,1, \dots, 7$ (ver segunda columna). Por lo tanto, $x_\alpha = 2\pi t$, y $f(x_\alpha)$ son las correspondientes temperaturas (tercera columna). Si aplicamos ahora las expresiones de los coeficientes, c_j , para el caso discreto obtenemos

$$c_0 = -1,95, \quad c_{-1} = -0,37 - 1,28i, \quad c_1 = -0,37 + 1,28i.$$

Ahora podemos calcular los coeficientes a_0 , a_1 y b_1 como

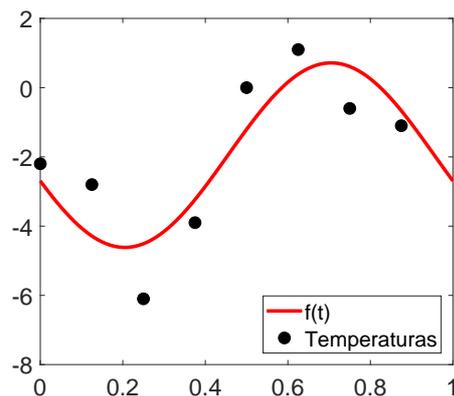
$$a_0 = c_0 + c_0 = -3,9, \quad \frac{a_0}{2} = -1,95,$$

$$a_1 = c_1 + c_{-1} = -0,74,$$

$$b_1 = i(c_1 - c_{-1}) = -2,56.$$

En la figura 4 presentamos la aproximación obtenida.

Figura 4



1.5.2. Series de Fourier

Toda función f continua a trozos con periodo 2π puede ser asociada con una serie de Fourier de las siguientes dos formas vistas anteriormente,

$$\frac{1}{2}a_0 + \sum_{j=1}^{\infty} (a_j \cos(jx) + b_j \sin(jx)), \quad \text{o} \quad \sum_{j=-\infty}^{\infty} c_j \exp(ijx).$$

Los coeficientes c_j y a_j y b_j , se pueden calcular mediante las ecuaciones (17) y (18), respectivamente. Si f y su primera derivada, f' , son continuas en todo el dominio, entonces la serie converge en todo el dominio a $f(x)$. Se permite incluso que f y f' tengan un número finito de discontinuidades en cada periodo. En tales discontinuidades, la serie da la media de los valores límite por la derecha e izquierda en el punto estudiado. La suma parcial o truncada de la serie es la mejor aproximación posible a $f(x)$, en el sentido de los mínimos cuadrados, mediante el uso de polinomios trigonométricos.

Teniendo en cuenta la ecuación (18), es fácil deducir las siguientes afirmaciones:

- Si f es una función par, esto es $f(x) = f(-x)$, para cualquier x , entonces $b_j = 0$, para todo j . La serie de Fourier es entonces una serie de cosenos.
- Si f es una función impar, esto es $f(x) = -f(-x)$, para cualquier x , entonces $a_j = 0$, para todo j . La serie de Fourier es entonces una serie de senos.

Expansión de Fourier de la onda cuadrada

Consideremos la siguiente función de onda cuadrada (ver figura 5)

$$f(x) = \begin{cases} -1 & \text{si } -\pi < x < 0, \\ 1 & \text{si } 0 < x < \pi. \end{cases}$$

Imaginemos que podemos hacer una continuación periódica de $f(x)$ fuera del intervalo $(-\pi, \pi)$. Entonces f es una función impar, con lo que ya sabemos que todos los coeficientes de "coseno" son cero, $a_j = 0$. Los coeficientes de "seno", b_j , se calculan

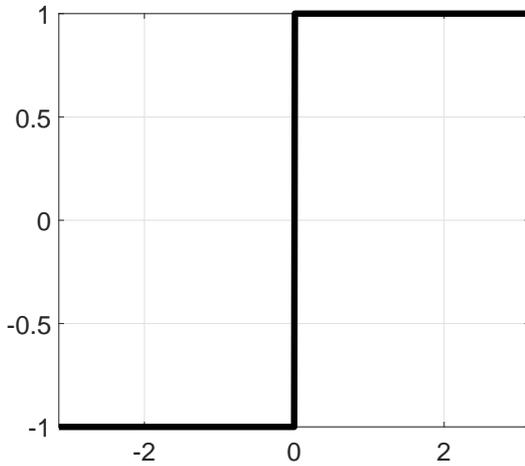
$$\begin{aligned} b_j &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(jx) dx = -\frac{1}{\pi} \int_{-\pi}^0 \sin(jx) dx + \frac{1}{\pi} \int_0^{\pi} \sin(jx) dx \\ &= \frac{2}{\pi} \int_0^{\pi} \sin(jx) dx = \frac{2}{\pi} \frac{1 - \cos(j\pi)}{j} = \begin{cases} 0 & \text{si } j \text{ es par,} \\ \frac{4}{j\pi} & \text{si } j \text{ es impar.} \end{cases} \end{aligned}$$

Entonces

$$f(x) = \frac{4}{\pi} \left(\sin(x) + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \dots \right).$$

La suma de la serie es cero en los puntos donde f presenta una discontinuidad. Esto concuerda con el hecho de que la suma debe ser igual a la media de los valores límite por la derecha y la izquierda en la discontinuidad. En la figura 6 presentamos la aproximación obtenida al truncar la serie infinita a n términos, considerando diferentes valores de n . Observamos que, a mayor n la aproximación claramente mejora, esto es, converge a f .

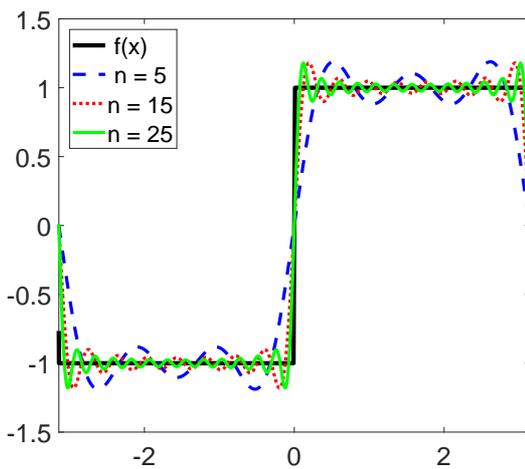
Figura 5



1.5.3. Teorema integral de Fourier

Supongamos que la función $\psi(\xi)$ definida en todo el dominio real y que satisface las propiedades de regularidad requeridas para que las series de Fourier sean convergentes,

Figura 6



Sean

$$x = \frac{2\pi\xi}{L}, \quad \psi(\xi) = f(x).$$

Si tenemos

$$c_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \exp(-ijx) dx = \frac{1}{L} \int_{-\frac{1}{2}L}^{\frac{1}{2}L} \psi(\xi) \exp(-2\pi i \xi j/L) d\xi,$$

entonces, considerando las series de Fourier para $x \in (-\pi, \pi)$ o, equivalentemente, $\xi \in (-\frac{1}{2}L, \frac{1}{2}L)$, resulta

$$f(x) = \sum_{j=-\infty}^{\infty} c_j \exp(ijx), \quad \psi(\xi) = \sum_{j=-\infty}^{\infty} c_j \exp(2\pi i \xi j/L).$$

Si ahora tenemos que

$$g_L(t) = \frac{1}{L} \int_{-\frac{1}{2}L}^{\frac{1}{2}L} \psi(\xi) \exp(-2\pi i \xi t) d\xi, \quad (19)$$

entonces obtenemos

$$\psi(\xi) = \frac{1}{L} \sum_{j=-\infty}^{\infty} g_L\left(\frac{j}{L}\right) \exp(2\pi i \xi j/L), \quad \xi \in \left(-\frac{1}{2}L, \frac{1}{2}L\right). \quad (20)$$

Pasando al límite, esto es $L \rightarrow \infty$, se evita hacer una continuación periódica artificial de la función fuera del intervalo finito. El sumatorio en la ecuación (20) es una “suma de rectángulos”, parecida a la que aparece en la definición de integral definida. Sin embargo, aquí el argumento varía entre $-\infty$ y ∞ , y la función $g_L(t)$ depende de L . Mediante un paso al límite no muy ortodoxo, el par de fórmulas en las ecuaciones (19) y (20) se convierten en

$$g(t) = \int_{-\infty}^{\infty} \psi(\xi) \exp(-2\pi i \xi t) d\xi \iff \psi(t) = \int_{-\infty}^{\infty} g(\xi) \exp(2\pi i \xi t) dt. \quad (21)$$

De hecho, se puede demostrar, siguiendo un análisis matemático bastante complicado, que el resultado anterior es correcto. La demostración requiere, además de las condiciones de regularidad ya mencionadas (de carácter local), unas condiciones de regularidad “global”,

$$\int_{-\infty}^{\infty} |\psi(\xi)| d\xi, \quad \text{es convergente.}$$

La relación (casi simétrica) presentada en la ecuación (21) se denomina **teorema integral de Fourier**. Este teorema (y otras versiones de él) es de gran

importancia tanto en la matemática pura como en la aplicada. La función g se conoce como la **transformada de Fourier** de ψ .

Transformada de Fourier

La función $\psi(\xi) = \exp(-|\xi|)$ tiene la transformada de Fourier,

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty} \exp(-|\xi|) \exp(-2\pi i \xi t) d\xi = \int_{-\infty}^{\infty} (\exp(-(1+2\pi i t)\xi) + \exp(-(1-2\pi i t)\xi)) d\xi \\ &= \frac{1}{1+2\pi i t} + \frac{1}{1-2\pi i t} = \frac{1}{1+4\pi^2 t^2}. \end{aligned}$$

Utilizando ahora la ecuación (21), obtenemos

$$\int_{-\infty}^{\infty} \frac{1}{1+4\pi^2 t^2} \exp(2\pi i \xi t) dt = \exp(-|\xi|),$$

o, si consideramos $x = 2\pi t$,

$$\int_0^{\infty} \frac{1}{1+x^2} \cos(\xi x) dx = \frac{\pi}{2} \exp(-|\xi|), \quad \xi \in \mathbb{R}.$$

2. Regresión

La *regresión* es una aplicación (o un caso particular) de aproximación de funciones. De hecho, esta técnica es también conocida como *regresión por mínimos cuadrados*, en alusión a que proporciona una solución óptima sobre la base del marco teórico presentado en el apartado anterior.

Aquí nos centraremos en la llamada regresión lineal, donde el concepto “lineal” se refiere a la relación entre lo que serán las incógnitas del problema: los coeficientes de la expansión. Veremos cómo varios métodos diferentes encajan en esta definición.

La regresión es una metodología que ayuda a tratar conjuntos de datos experimentales, especialmente cuando estos no se presentan de una forma conveniente en cantidad y/o calidad, y extraer de ellos algún tipo de medida o desarrollar relaciones predictivas. Pretende una evaluación descriptiva de los datos que permita extraer conclusiones útiles para su estudio. En ese sentido, cuando usamos regresión estamos en el caso discreto de aproximación de funciones.

Existen dos problemas típicos que nos solemos encontrar cuando manejamos datos experimentales: análisis de tendencia y contraste de hipótesis.

El **análisis de tendencia** se basa en usar un patrón observado en los datos para sacar conclusiones generales del experimento analizado o hacer predicciones futuras de su evolución.

En el **contraste de hipótesis** se compara un modelo matemático existente con los datos observados. Si los parámetros del modelo son desconocidos, puede ser necesario determinarlos para tener el mejor ajuste posible con los datos medidos. Por otra parte, si los coeficientes ya están disponibles, puede ser conveniente comparar los valores predichos por el modelo y los valores reales observados, para evaluar la adecuación del modelo. Habitualmente varios modelos son comparados para conseguir el mejor ajuste a un conjunto de observaciones dado.

2.1. Regresión lineal

Cuando tenemos datos que son de naturaleza variable, inestable o pseudoaleatoria, o que pueden presentar errores de medida (“ruido”), la mejor estra-

Fuente del apartado

Apartado basado en la siguiente obra: **Steven C. Chapra** (2018) *Applied Numerical Methods with MATLAB for Engineers and Scientists, Fourth Edition*. Nueva York, NY: McGraw-Hill Education.

tegia para describir las observaciones suele ser una línea (función) que pueda reproducir la forma o tendencia general de los datos, más que ajustar cada punto individualmente, lo que se correspondería con un problema de interpolación.

Intuitivamente, la regresión sería el equivalente objetivo y matemáticamente sustentado a una metodología más visual consistente en, dados los datos representados en una gráfica, dibujar la “mejor” línea sobre los puntos que los describe. El criterio que, aplicando el sentido común, se utilizaría es de dibujar una línea que minimice de algún modo la discrepancia entre los puntos y la curva. Entonces, aprovechando el marco teórico de la aproximación de funciones, podemos concluir que la forma de medir esa distancia de manera que su minimización nos proporcione la mejor solución es la de los **mínimos cuadrados**. Recordemos que, entre muchas otras ventajas, la solución que se obtiene es única, algo que ya resulta intuitivo para pensar que es la solución óptima.

Pensemos ahora en el caso más sencillo: describir unos datos dados por una línea recta. Supongamos que tenemos un conjunto de datos por pares, $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$. La expresión matemática de una recta es

$$y = c_0 + c_1x,$$

donde los coeficientes c_0 y c_1 son la *ordenada en origen* y la *pendiente*, respectivamente. Asumiendo que la recta no aproxima todas las observaciones perfectamente, para cada par (y_i, x_i) cometemos un error ϵ_i , esto es

$$\epsilon_i = y_i - c_0 - c_1x_i,$$

conocido también como *residuo*. Al considerar mínimos cuadrados, particularmente el caso discreto, obtenemos

$$\sum_{i=1}^m \epsilon_i^2 = \sum_{i=1}^m (y_i - c_0 - c_1x_i)^2.$$

Recordemos en este momento la solución que se presenta en el apartado 1, que tenemos que adaptar al contexto de la regresión mediante una recta, tomando $n = 1$, $\phi_0(x) = 1$, $\phi_1(x) = x$, $f(x_i) = y_i$ y $w_i = \frac{1}{m}$. De esta manera, la obtención de los coeficientes se simplifica mucho, siendo

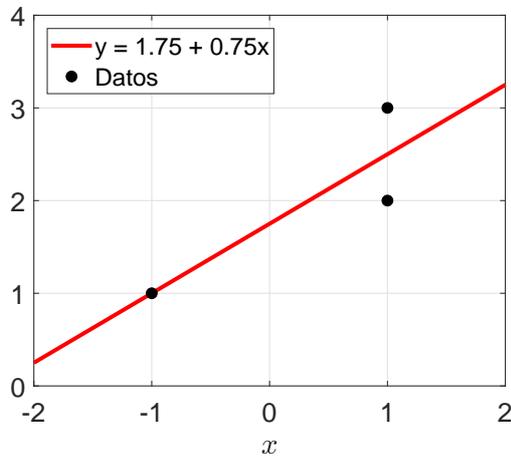
$$c_1 = \frac{m \sum_{i=1}^m x_i y_i - \sum_{i=1}^m x_i \sum_{i=1}^m y_i}{m \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2},$$

$$c_0 = \frac{1}{m} \sum_{i=1}^m y_i - c_1 \frac{1}{m} \sum_{i=1}^m x_i$$

Regresión lineal

Encontrad la recta que mejor se ajusta a los puntos (1,2),(-1,1),(1,3). Resolviendo el sistema de ecuaciones, obtenemos que $c_0 = \frac{7}{4}$ y $c_1 = \frac{3}{4}$. En la figura 7 podemos ver el resultado.

Figura 7



Podemos ahora calcular los residuos o errores que se cometen en la aproximación, presentados en la tabla 3.

Tabla 3

x_i	y_i	Recta, y	Residuos, ϵ_i
1	2	2,5	-0,5
-1	1	1,0	0,0
1	3	2,5	0,5

2.1.1. Cuantificación del error de la regresión lineal

Como ya hemos mostrado en el apartado 1, la solución basada en mínimos cuadrados es única y, por lo tanto, cualquier otra recta produciría una suma de residuos mayor. Una serie de propiedades adicionales se pueden deducir simplemente examinando un poco más de cerca la manera como los residuos son calculados. Si definimos su suma, S_r , como

$$S_r = \sum_{i=1}^m (y_i - c_0 - c_1 x_i)^2,$$

observamos un parecido a una expresión muy conocida, la de la desviación típica de y dadas muestras y_i ,

$$s_y = \sqrt{\frac{S_t}{m-1}}, \quad S_t = \sum_{i=1}^m (y_i - \bar{y})^2$$

donde \bar{y} representa la media. Sabemos que la desviación típica expresa la discrepancia de las muestras con respecto a un valor central, la media. S_r puede

ser visto como la discrepancia de los datos con respecto a otra medida central, en este caso la línea recta.

La analogía se puede llevar aún más lejos a casos donde, por una parte, los datos se separan de la línea con una magnitud similar en todo el dominio de estos y, por otra parte, la distribución de los puntos alrededor de la línea es normal. En esta situación se puede demostrar que la solución de mínimos cuadrados es la que mejor (incluso utilizando otras medidas) estima c_0 y c_1 . En estadística se conoce como el *principio de máxima verosimilitud*. Además, si las condiciones previas se cumplen, se puede calcular una “desviación típica” para la regresión lineal, definida

$$s = \sqrt{\frac{S_r}{m-2}},$$

que se conoce como *error estándar de la estimación*. Nótese que, a diferencia de la varianza estadística (o la desviación típica), en el denominador tenemos $m-2$, ya que estimamos dos valores c_0 y c_1 , por lo que hemos perdido dos grados de libertad.

Del mismo modo que la desviación típica estadística, el error estándar de la estimación cuantifica la dispersión de los datos, aunque en este caso alrededor de la recta.

Estos conceptos pueden ser usados para cuantificar la “bondad” del ajuste, lo que resulta muy útil a la hora de comparar diferentes tipos de regresiones. Un caso particular de especial importancia es cuando consideramos las medidas de dispersión S_r y S_t , que corresponden a la dispersión con respecto a la recta y a la media, respectivamente. La diferencia $S_t - S_r$ cuantifica la reducción del error debido a describir los datos en términos de la media o una recta. Como la magnitud de esta medida depende de la escala, lo normalizamos a S_t y obtenemos el llamado *coeficiente de determinación*

$$r^2 = \frac{S_t - S_r}{S_t}, \quad (22)$$

cuya raíz cuadrada, r , es el *coeficiente de correlación*. Si $r^2 = 1$ ($S_r = 0$), tenemos que la línea recta explica perfectamente el 100% de la variabilidad de los datos. Sin embargo, para $r^2 = 0$ ($S_r = S_t$), la línea recta no mejora a la media a la hora de describir los datos. Una formulación alternativa de r que resulta ser más eficiente para su implementación en el ordenador es

$$r = \frac{m \sum_{i=1}^m x_i y_i - \sum_{i=1}^m x_i \sum_{i=1}^m y_i}{\sqrt{m \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2} \sqrt{m \sum_{i=1}^m y_i^2 - (\sum_{i=1}^m y_i)^2}}.$$

Error estándar

Continuando con el ejemplo anterior, podemos calcular ahora las medidas de error que hemos introducido. La suma de residuos al cuadrado sería

$$S_r = (-0,5)^2 + 0^2 + 0,5^2 = 0,5.$$

Una vez calculado S_r , el error estándar se obtiene fácilmente, recordando que en este ejemplo $m = 3$,

$$s = s = \sqrt{\frac{S_r}{m-2}} = s = \sqrt{0,5} = 0,71.$$

Finalmente, el coeficiente de correlación r

$$r = \sqrt{\frac{S_t - S_r}{S_t}} = \sqrt{\frac{2 - 0,5}{2}} = 0,87.$$

2.2. Linealización de relaciones no lineales

La regresión lineal es una metodología muy potente para describir de la mejor manera un conjunto de datos con una línea. Sin embargo, su aplicabilidad y eficiencia dependen significativamente de que la relación entre las variables dependiente e independiente sea lineal (o cercana a lineal). Por supuesto, existen múltiples casos de disciplinas de la ciencia y la ingeniería en los que los datos no presentan un comportamiento lineal, como el estudio de poblaciones, radiación, etc. Como primer paso intuitivo en el análisis de los datos, basta con presentarlos en un gráfico y discernir si el modelo lineal es el más conveniente.

Cuando un comportamiento no lineal es observado, se puede hacer uso de técnicas de regresión no lineal. Sin embargo, en algunos casos, relaciones altamente no lineales se pueden transformar para expresar los datos en una forma que sea compatible con la regresión lineal. Así se simplifica mucho su tratamiento. A continuación presentamos algunos ejemplos:

- **Modelo exponencial**

$$y = \alpha_1 \exp \beta_1 x,$$

donde α_1 y β_1 son constantes. Este modelo se emplea en muchas disciplinas de ingeniería y ciencia para simular valores que crecen (β_1 positivo) o decrecen (β_1 negativo) a una velocidad directamente proporcional a su propia magnitud. La relación entre las variables x e y es, en principio, claramente no lineal. Sin embargo, aplicando una transformación bastante simple, tomando el logaritmo natural, obtenemos

$$\log(y) = \log(\alpha_1) + \beta_1 x,$$

que se ajusta al modelo lineal.

Automóviles en el mundo

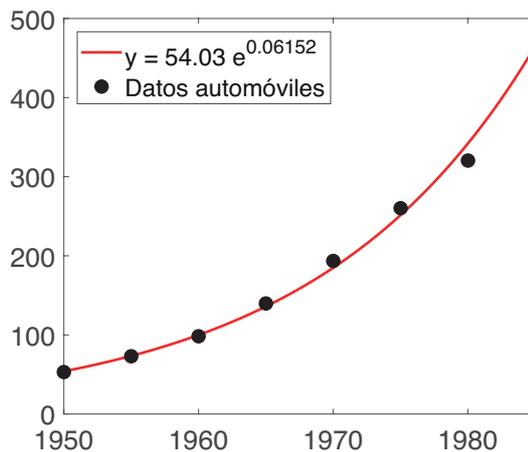
Los datos de la tabla 4 describen el número de automóviles operando en todo el mundo para diferentes años desde 1950. A simple vista, parece que la cantidad de coches ha seguido un crecimiento exponencial. Entonces vamos a ajustar un modelo exponencial para estos datos o, lo que es lo mismo, determinar α_1 y β_1 de la descripción anterior.

Tabla 4

Año	Automóviles ($\times 10^6$)
1950	53,05
1955	73,04
1960	98,31
1965	139,78
1970	193,48
1975	260,20
1980	320,39

Después de hacer la transformación logarítmica y resolver el sistema de las ecuaciones normales obtenemos: $c_0 = \log(\alpha_1) = 3,986$ y $c_1 = \beta_1 = 0,06152$. Deshaciendo la transformación calculamos $\alpha_1 = \exp(c_0) = 54,03$. En la figura 8 podemos ver los datos y la forma de la curva exponencial de regresión.

Figura 8



• **Ecuación potencia**

$$y = \alpha_2 x^{\beta_2},$$

donde α_1 y β_1 son constantes. Esta relación se utiliza mucho para ajustar datos experimentales cuando el modelo matemático es desconocido. Si $\beta_2 \neq 0$ la ecuación es no lineal. De forma similar al anterior, pero tomando logaritmos en base 10, tenemos

$$\log_{10}(y) = \log_{10}(\alpha_2) + \beta_2 \log_{10}(x),$$

que resulta en una recta con pendiente β_2 y ordenada en el origen $\log_{10}(\alpha_2)$. Nótese que se podría usar cualquier base para el logaritmo, aunque la base 10 es la más habitual.

- **Ecuación de la tasa de crecimiento de saturación**

$$y = \alpha_3 \frac{x}{\beta_3 + x},$$

donde α_1 y β_1 son constantes. La ecuación describe un crecimiento que se amortigua cuando x se hace grande, lo que representa un efecto de “saturación”. Para linealizar esta ecuación basta con invertirla,

$$\frac{1}{y} = \frac{1}{\alpha_3} + \frac{\beta_3}{\alpha_3} \frac{1}{x}.$$

Estas transformaciones permiten utilizar estos modelos en el marco de la regresión lineal, y calcular los coeficientes. Una vez calculados, se recupera el modelo original deshaciendo la transformación y pudiéndolo usar para propósitos de predicción.

2.3. Regresión lineal polinómica

Hasta ahora solo hemos hablado de cómo utilizar la recta para describir los datos. Sin embargo, hay situaciones donde, aunque la distribución de los puntos presenta un patrón claro, este no se puede ajustar mediante una línea recta. Para estos casos, una curva suele ser una mejor opción. Así, además de la linealización mediante transformaciones presentada anteriormente, se puede optar por ajustar un polinomio a los datos, utilizando la llamada *regresión polinómica*. Este tipo de regresión es, en cierto modo, una generalización del basado en la recta, aunque sigue siendo lineal y un caso particular de la teoría de aproximaciones de funciones por mínimos cuadrados.

Entonces, generalizaremos el caso de la recta a un polinomio de grado mayor. Supongamos que ajustamos un polinomio de segundo orden (cuadrático), de la forma

$$y = c_0 + c_1x + c_2x^2.$$

Como en el caso anterior, esta formulación es una especialización del problema de mínimos cuadrados presentado en el apartado 1. Simplemente debemos tomar $n = 2$, $\phi_0(x) = 1$, $\phi_1(x) = x$, $\phi_2(x) = x^2$ y $f(x) = y$. Ahora la suma del cuadrado de los residuos es

$$S_r = \sum_{i=0}^m \left(y_i - c_0 - c_1x_i - c_2x_i^2 \right)^2,$$

donde podemos observar que ahora tenemos un término más para tratar de reducir las diferencias individuales. Aplicando la solución ya estudiada basada en la ecuaciones normales, los coeficientes c_0 , c_1 y c_2 son fácilmente calculados. Cabe aclarar de nuevo que este tipo de ajuste sigue perteneciendo al contexto de regresión lineal, ya que las incógnitas del problema son los coeficientes, cuya relación es lineal.

Finalmente, la regresión polinómica se puede utilizar con polinomios de grado mayor, n . Así tendríamos,

$$y = c_0 + c_1x + c_2x^2 + \dots + c_nx^n.$$

De nuevo, mediante el sistema de ecuaciones normales ($n + 1$ ecuaciones lineales) se obtienen los coeficientes $c_0, c_1, c_2, \dots, c_n$. El error estándar se formula como

$$s = \sqrt{\frac{S_r}{m - (n + 1)}},$$

donde el denominador es $m - (n + 1)$ porque $n + 1$ coeficientes se utilizan para calcular S_r , con lo que hemos perdido $n + 1$ grados de libertad. El coeficiente de determinación r^2 se puede calcular utilizando la ecuación (22).

2.4. Regresión lineal múltiple

Otra extensión muy empleada que también se enmarca en la regresión lineal aparece cuando la variable y es una función lineal de dos o más variables. Por ejemplo, y puede ser una función de x_1 y x_2 de la forma

$$y = c_0 + c_1x_1 + c_2x_2.$$

Esta formulación es particularmente útil cuando se ajustan datos experimentales donde la variable estudiada depende de otras dos variables. Por ello, esta técnica también se suele llamar **variables explicativas**, ya que se trata de determinar la influencia de x_1 y x_2 en la variable y . En este caso concreto bidimensional, la "línea" se sustituye por un "plano".

Como en los casos anteriores, definimos la suma de residuos como

$$S_r = \sum_{i=0}^m (y_i - c_0 - c_1x_{1,i} - c_2x_{2,i})^2.$$

Los coeficientes c_0 , c_1 y c_2 se obtiene resolviendo el problema de mínimos cuadrados con $n = 2$, $\phi_0(x) = 1$, $\phi_1(x) = x_1$, $\phi_2(x) = x_2$ y $f(x) = y$.

De nuevo, el caso bidimensional se puede extender al caso de n dimensiones, teniendo

$$y = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n,$$

donde el error estándar se formula como

$$s = \sqrt{\frac{S_r}{m - (n + 1)}},$$

y el coeficiente de determinación se calcula con la ecuación (22).

Además de los casos donde una variable es linealmente dependiente de dos o más variables, la regresión lineal múltiple puede ser utilizada en la derivación de ecuaciones de potencias de la forma genérica

$$y = c_0x_1^{c_1}x_2^{c_2} \dots x_n^{c_n}.$$

Este tipo de ecuaciones son extremadamente útiles ajustando datos experimentales. Para usar regresión lineal múltiple en este contexto, basta con aplicar una transformación logarítmica para tener

$$\log(y) = \log(c_0) + c_1 \log(x_1) + c_2 \log(x_2) + \dots + c_n \log(x_n).$$

En esta guía hemos presentado varios modelos y técnicas numéricas para aproximar funciones. La selección del modelo adecuado dado un conjunto de datos no es una tarea trivial y, en muchas ocasiones, requiere un análisis exhaustivo tanto de las propiedades del modelo como de la calidad/cantidad de los datos. Aquí, particularmente en el apartado de regresión, hemos dado algunas medidas de la bondad en el ajuste, que pueden ser usadas para la comparación de distintos modelos, aunque muchos otros factores tienen que ser valorados: robustez, capacidad predictiva, facilidad de uso, etc. En la práctica, cualquier conocimiento previo o la intervención de expertos en la materia son también aspectos que se deben tener en cuenta.

Bibliografía

Chapra, Steven C. (2018). *Applied Numerical Methods with MATLAB for Engineers and Scientists, Fourth Edition*. Nueva York, NY: McGraw-Hill Education.

Dahlquist, Germund; Bjorck, Ake (2003). *Numerical Methods*. Dover Publications, Inc.