Tema 9. Aproximación

1. Introducción

La idea básica de la aproximación es, dada una serie de datos, encontrar una función tal que su gráfica pase "cerca" de los datos y que, además, represente bien la forma de la nube de puntos determinada por los datos.

En principio, el problema planteado es parecido al de interpolación, pero hay una diferencia importante que se refleja justamente en la primera exigencia realizada:

- Interpolación: la gráfica de la función interpoladora ha de pasar exactamente por los datos.
- Aproximación: la gráfica de la función que aproxima "sólo" ha de pasar cerca de los datos.

Por ejemplo, veamos el siguiente dibujo (realizado con Octave).

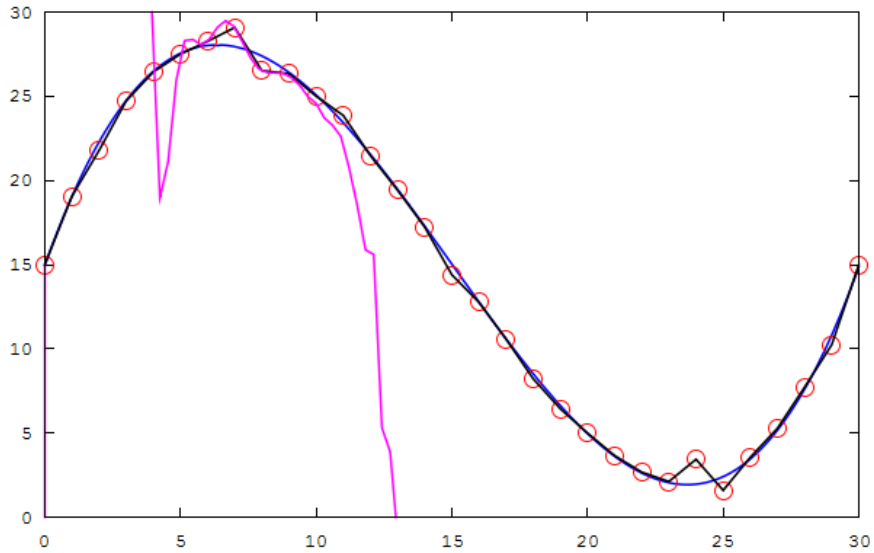

Figura 1: aproximación versus interpolación.

En la figura aparecen los siguientes elementos.

- 31 datos, que están marcados en rojo.
- En magenta tenemos parte de la gráfica del polinomio de grado 30 que interpola estos datos. Como se puede comprobar, el resultado que proporciona este polinomio de interpolación deja bastante que desear.
- La gráfica negra corresponde al spline lineal, que se ajusta mejor a los datos que el polinomio de interpolación pero que sólo es continua (existen "picos").
- Por último, la gráfica azul es la del polinomio de tercer grado que mejor aproxima, en el sentido de los mínimos cuadrados discreto, los datos. Aunque este polinomio no pasa exactamente por los datos, si representa mejor la nube de puntos determinada por ellos.

Por otra parte, como únicamente queremos "aproximar" datos, tenemos dos posibilidades a la hora de atacar el problema:

- Aproximación discreta: el conjunto de datos es finito.

- Aproximación continua: el dato es una función conocida en todo un intervalo pero complicada de manejar.

En ambos casos, nuestro objetivo será encontrar una función que sea fácil de calcular, fácil de evaluar, suficientemente derivable, etc.

La cuestión principal en este tema es cómo decidir cuándo una función es mejor que otra al aproximar datos. A tal fin, definiremos una función error de manera que, el problema planteado, resultará equivalente a medir cuánto nos equivocamos con cada posible función y, de esta forma, poder escoger la que da lugar al error "más pequeño posible".

En la siguiente tabla se recogen varias posibilidades para definir la función error. En la misma consideramos que los datos son el conjunto de pares $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ o una función $f : [a, b] \to \mathbb{R}$ y que, en ambos casos, la función que aproxima es $u(x)$.

Tipo	Discreto	Continuo
Error	$e_i = \lvert u(x-i) - y_i \rvert$	$e(x) = \lvert u(x) - f(x) \rvert$
Mínima suma	$\sum_{i=1}^{N} e_i$	$\int_a^b e(x)\mathrm{d}x$
Mínimos cuadrados	$\sum_{i=1}^{N} e_i^2$	$\int_a^b e^2(x)\mathrm{d}x$
Minimax	$\operatorname*{máx}_{0\le i\le n} e_i$	$\operatorname*{máx}_{[a,b]} e(x)$

Por cuestiones de simplicidad al realizar los cálculos, sólo vamos a estudiar los casos de mínimos cuadrados.

2. El método de aproximación por mínimos cuadrados discreto

Para medir el error que cometemos con cada función, consideramos el *error cuadrático*. Para entender más claramente cómo se define este error, veamos un ejemplo concreto.

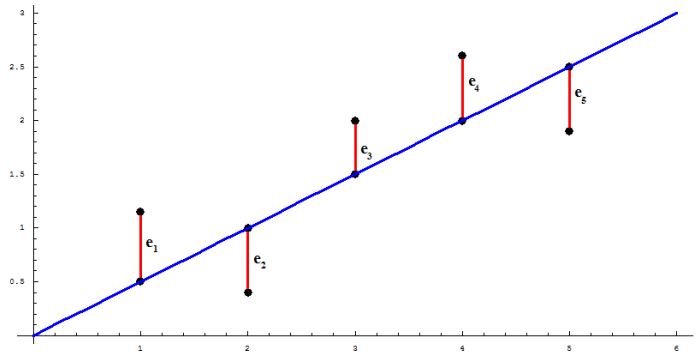

Figura 2: Ejemplo de aproximación por mínimos cuadrados discreto.

Tenemos el conjunto de datos $\{(1, 1.15), (2, 0.4), (3, 2), (4, 2.6), (5, 1.9)\}$. Si tomamos la recta $y = 0.5x$ como aproximación de estos datos, el error cuadrático cometido es

$$\begin{aligned} E_2 &= e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2 \\ &= (0.5 - 1.15)^2 + (1 - 0.4)^2 + (2.5 - 2)^2 + (2 - 2.6)^2 + (2.5 - 1.9)^2 \\ &= 0.4225 + 0.36 + 0.25 + 0.36 + 0.36 = 1.7525. \end{aligned}$$

Para cualquier otra recta $y = a + bx$ el error cometido será

$$\begin{aligned} E_2(a, b) &= e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2 \\ &= (a + b - 1.15)^2 + (a + 2b - 0.4)^2 + (a + 3b - 2)^2 + (a + 4b - 2.6)^2 \\ &\quad + (a + 5b - 1.9)^2. \end{aligned}$$

Tenemos, pues, que el error cuadrático depende de los coeficientes de la recta, es decir, de a y b. Así, el problema de encontrar la recta que ajusta con el menor error cuadrático se transforma en la búsqueda del mínimo de la función de dos variables $E_2(a, b)$.

Puesto que $E_2(a,b)$ es, básicamente, una función polinómica de segundo grado con coeficientes positivos en los términos a^2 y b^2 (es decir, una "parábola tridimensional hacia arriba"), podemos estar seguros de que, si somos capaces de calcular un extremo relativo de dicha función, entonces tendremos un mínimo absoluto.

Para calcular los posibles extremos recurrimos a las derivadas parciales de la función $E_2(a,b)$. De esta forma tenemos el sistema

$$\left.\begin{array}{l} \frac{\partial E_2}{\partial a}(a,b)=0 \\ \frac{\partial E_2}{\partial b}(a,b)=0 \end{array}\right\}.$$

En nuestro ejemplo el sistema es

$$\left.\begin{array}{r} 2(a+b-1.15)+2(a+2b-0.4)+2(a+3b-2)+2(a+4b-2.6)+2(a+5b-1.9)=0 \\ 2(a+b-1.15)+2(a+2b-0.4)2+2(a+3b-2)3+2(a+4b-2.6)4+2(a+5b-1.9)5=0 \end{array}\right\}.$$

Tras operar y simplificar, nos queda

$$\left.\begin{array}{r} 5a+15b=8.05 \\ 15a+55b=27.85 \end{array}\right\}.$$

Este es un sistema lineal, de dos ecuaciones con dos incógnitas, bastante fácil de resolver. En concreto, la solución es $a=0.5$, $b=0.37$. Así, la recta que mejor ajusta por mínimos cuadrados viene dada por la expresión $y=0.5+0.37x$. Para esta recta el error cuadrático cometido es

$$\begin{aligned} E_2(0.5,0.37) &= e_1^2+e_2^2+e_3^2+e_4^2+e_5^2 \\ &= (0.87-1.15)^2+(1.24-0.4)^2+(1.61-2)^2+(1.98-2.6)^2+(2.35-1.9)^2 \\ &= 0.0784+0.7056+0.1521+0.3844+0.2025=1.5230, \end{aligned}$$

que, por supuesto, es menor que el calculado antes para la recta $y=0.5x$.

3. Cómo obtener el sistema que debemos resolver

Cuando se busca la recta de mínimos cuadrados, hay un par de reglas que nos permiten obtener fácilmente el sistema que debemos resolver. Además, tales métodos se generalizan muy bien al caso de aproximación por mínimos cuadrados discreto para cualquier familia de funciones polinomiales.

3.1. Primera regla: elaborando tablas

Esta regla consiste en realizar una tabla de valores (cuyas columnas corresponden a los valores de x, y, x^2, xy) y calcular la suma de los elementos de cada columna. En nuestro ejemplo quedaría

	x	y	x^2	xy
	1	1.15	1	1.15
	2	0.4	4	0.8
	3	2	9	6
	4	2.6	16	10.4
	5	1.9	25	9.5
Suma:	15	8.05	55	27.85

Podemos ver que las cuatro sumas corresponden a los coeficientes (en realidad, faltaría uno) del sistema resuelto en la sección anterior. En general, el sistema a resolver es

$$\left.\begin{array}{r} \left(\sum 1\right)a+\left(\sum x\right)b=\sum y \\ \left(\sum x\right)a+\left(\sum x^2\right)b=\sum(xy) \end{array}\right\},$$

donde

- $\sum 1$ es igual al número de datos considerados (5 en el ejemplo);
- $\sum x$ es la suma de las primeras componentes de cada dato (15 en el ejemplo);
- $\sum x^2$ es la suma de los cuadrados de las primeras componentes (55 en el ejemplo);
- $\sum y$ es la suma de las segundas componentes de cada dato (8.05 en el ejemplo);

- $\sum xy$ es la suma de los productos de las componentes de cada dato (27.85 en el ejemplo).

Si queremos aproximar por una parábola (esto es, una función del tipo $y = a + bx + cx^2$), necesitaremos un sistema de tres ecuaciones para determinar biunívocamente los valores de a, b, c. Observando que, para la recta, la segunda ecuación se "puede obtener" a partir de la primera "sin más que multiplicar por x" cada una de las sumatorias, no es difícil imaginar que, para la parábola, el sistema necesario será

$$\left.\begin{array}{c} (\sum 1)\, a + (\sum x)\, b + (\sum x^2)\, c = \sum y \\ (\sum x)\, a + (\sum x^2)\, b + (\sum x^3)\, c = \sum (xy) \\ (\sum x^2)\, a + (\sum x^3)\, b + (\sum x^4)\, c = \sum (x^2 y) \end{array}\right\}.$$

Por consiguiente, tendremos que elaborar una tabla con siete columnas: $x, y, x^2, x^3, x^4, xy, x^2y$.

Ejercicio 1. A partir de lo visto para la recta y la parábola, determinar cuál es el sistema necesario (y, por tanto, la tabla necesaria) para calcular el polinomio de grado menor o igual que m que mejor aproxima por mínimos cuadrados un conjunto de n datos, siendo $m < n + 1$.

Nota 2. La necesidad de la condición $m < n + 1$ se justifica fácilmente sin más que observar que, por dos puntos cualesquiera, pasan infinitas parábolas.

3.2. Segunda regla: operando con matrices

Supongamos, por un momento, que los datos de nuestro ejemplo estuvieran alineados. En tal caso existiría una recta $y = a + bx$ para la que serían ciertas las siguientes igualdades

$$\left.\begin{array}{l} a + b \times 1 = 1.15 \\ a + b \times 2 = 0.4 \\ a + b \times 3 = 2 \\ a + b \times 4 = 2.6 \\ a + b \times 5 = 1.9 \end{array}\right\}.$$

Sin embargo, el sistema de ecuaciones resultante no tiene solución. Por ejemplo, $(a, b) = (1.9, 0.75)$ es el único par de números que cumple las dos primeras igualdades; sin embargo, no satisface las tres últimas.

A pesar de todo, vamos a calcular la mejor de todas las "no-soluciones". Para ello, empezamos expresando matricialmente el sistema anterior, esto es, escribimos

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1.15 \\ 0.4 \\ 2 \\ 2.6 \\ 1.9 \end{pmatrix}.$$

Observemos que, en la matriz de coeficientes, la primera columna está formada por 1's y la segunda por los valores de las primeras componentes de los datos. Además, el vector de términos independientes se corresponde con los valores de las segundas componentes de los datos.

Para tener un sistema con el mismo número de incógnitas y de ecuaciones, multiplicamos en ambos lados del sistema por la transpuesta de la matriz de coeficientes,

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1.15 \\ 0.4 \\ 2 \\ 2.6 \\ 1.9 \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8.05 \\ 27.85 \end{pmatrix},$$

y así btenemos, precisamente, el sistema que hemos resuelto en la primera regla.

Si queremos calcular la parábola que mejor ajusta, entonces el sistema a resolver será

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{pmatrix} \begin{pmatrix} 1.15 \\ 0.4 \\ 2 \\ 2.6 \\ 1.9 \end{pmatrix} \Rightarrow$$

$$\begin{pmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8.05 \\ 27.85 \\ 109.85 \end{pmatrix}.$$

En este caso, la matriz de coeficientes tiene tres columnas, estando formada la tercera de ellas por los cuadrados de las primeras componentes de los datos.

Ejercicio 3. Determinar cómo se construye el sistema necesario para calcular el polinomio de grado m que mejor aproxima por mínimos cuadrados, que aproxima n datos (con $m < n + 1$) usando esta regla.

4. Otras funciones para aproximar

En las Sección 2, al realizar las derivadas parciales de la función error cuadrático, el sistema resultante siempre es lineal. Esto ocurre al aproximar por polinomios. Sin embargo, si usamos otras familias de funciones (las exponenciales, por ejemplo) entonces el sistema resultante no es lineal y, en general, imposible de resolver de manera exacta. Ante este problema surgen dos posibles estrategias.

1. Resolución de manera aproximada del sistema.
2. Transformación de los datos originales de forma que baste con calcular una recta que aproxime (por mínimos cuadrados) los nuevos datos.

En el primer caso se trata de calcular una sucesión de soluciones de manera que, a ser posible, cada una sea mejor que la anterior. Para calcular dicha sucesión tenemos los métodos iterativos llamados de descenso (por ejemplo, los de máximo descenso o los de gradiente conjugado). Una explicación más detallada de estos métodos excede los límites de estas notas, por lo que no profundizaremos en su estudio.

La segunda vía se conoce como el *método de linealización* y es bastante más simple a la hora de comprender los cálculos a realizar aunque, en general, proporciona un peor resultado (esto es, un error cuadrático mayor) que el de la solución obtenida por un método de descenso. De todas formas, la aproximación obtenida es, en la práctica, suficientemente buena.

4.1. Exponenciales

Consideremos que $\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ es el conjunto de datos que queremos aproximar mediante la familia de funciones exponenciales dada de la forma $y = ae^{bx}$. Para "eliminar" la función exponencial, introducimos el logaritmo neperiano y operamos,

$$y = ae^{bx} \Rightarrow \ln(y) = \ln(ae^{bx}) = \ln(a) + \ln(e^{bx}) = \ln(a) + bx = A + bx.$$

Comprobamos así que hay una relación lineal (es decir, gráficamente están alineados sobre una recta) entre los valores de las primeras componentes (las x's) y los neperianos de las segundas componentes (las y's) de los datos originales. Ahora, calculando la recta por mínimos cuadrados para $\{(x_0, Y_0), (x_1, Y_1), \ldots, (x_n, Y_n)\} = \{(x_0, \ln(y_0)), (x_1, \ln(y_1)), \ldots, (x_n, \ln(y_n))\}$, obtenemos A y b. Por último, para determinar el valor de a en la exponencial, basta con hacer $a = e^A$.

4.2. Potenciales

Si queremos aproximar mediante la familia de funciones potenciales $y = ax^b$, consideramos los datos transformados $\{(X_i, Y_i) \mid 0 \leq i \leq n\} = \{(\ln(x_i), \ln(y_i)) \mid 0 \leq i \leq n\}$ y calculamos la recta $Y = A + BX$, donde $X = \ln(x)$, $Y = \ln(y)$, $A = \ln(a)$, $B = b$.

4.3. Hiperbólicas

Para aproximar con la familia de funciones hiperbólicas $y = a\frac{b}{x}$, se toman los datos transformados $\{(X_i, Y_i) \mid 0 \leq i \leq n\} = \left\{\left(\frac{1}{x_i}, y_i\right) \mid 0 \leq i \leq n\right\}$ y calculamos la recta $Y = A + BX$, donde $X = \frac{1}{x}$, $Y = y$, $A = a$, $B = b$.

4.4. Logísticas

Para aproximar con la familia de funciones logísticas $y = \frac{1}{1+ae^{bx}}$, tomamos los datos transformados $\{(X_i, Y_i) \mid 0 \leq i \leq n\} = \left\{\left(x_i, \frac{1}{y_i} - 1\right) \mid 0 \leq i \leq n\right\}$ y calculamos la recta $Y = A + BX$, donde $X = x$, $Y = \frac{1}{y} - 1$, $A = \ln(a)$, $B = b$.

5. El método de aproximación por mínimos cuadrados continuo

Supongamos que queremos aproximar una función $f : [a, b] \to \mathbb{R}$ mediante una combinación lineal de m funciones $\phi_1(x), \phi_2(x), \ldots, \phi_m(x)$ (de tipo polinomial, trigonométrico, exponencial, etc.) mediante mínimos cuadrados continuo. Entonces debemos minimizar la función

$$E(c_1, c_2, \ldots, c_n) = \int_a^b \big(c_1\phi(x) + c_2\phi(x) + \ldots + c_m\phi(m) - f(x)\big)^2 \,\mathrm{d}x.$$

Tomando derivadas parciales y simplificando, el sistema que nos permite calcular los extremos relativos de $E(c_1, c_2, \ldots, c_n)$ es

$$\left.\begin{array}{c}
c_1 \int_a^b \phi_1(x)\phi_1(x)\,\mathrm{d}x + c_2 \int_a^b \phi_1(x)\phi_2(x)\,\mathrm{d}x + \ldots + c_m \int_a^b \phi_1(x)\phi_m(x)\,\mathrm{d}x = \int_a^b \phi_1(x)f(x)\,\mathrm{d}x \\
c_1 \int_a^b \phi_2(x)\phi_1(x)\,\mathrm{d}x + c_2 \int_a^b \phi_2(x)\phi_2(x)\,\mathrm{d}x + \ldots + c_m \int_a^b \phi_2(x)\phi_m(x)\,\mathrm{d}x = \int_a^b \phi_2(x)f(x)\,\mathrm{d}x \\
\ldots \\
c_1 \int_a^b \phi_m(x)\phi_1(x)\,\mathrm{d}x + c_2 \int_a^b \phi_m(x)\phi_2(x)\,\mathrm{d}x + \ldots + c_m \int_a^b \phi_m(x)\phi_m(x)\,\mathrm{d}x = \int_a^b \phi_m(x)f(x)\,\mathrm{d}x
\end{array}\right\}.$$

5.1. Ejemplo

Queremos calcular el polinomio $p(x) = c_1 + c_2x + c_3x^2$ que mejor aproxima, mediante mínimos cuadrados continuo, la función $f(x) = |x|$ en el intervalo $[-1, 1]$.

El sistema que debemos resolver es

$$\left.\begin{array}{c}
c_1 \int_{-1}^1 1\,\mathrm{d}x + c_2 \int_{-1}^1 x\,\mathrm{d}x + c_3 \int_{-1}^1 x^2\,\mathrm{d}x = \int_{-1}^1 |x|\,\mathrm{d}x \\
c_1 \int_{-1}^1 x\,\mathrm{d}x + c_2 \int_{-1}^1 x^2\,\mathrm{d}x + c_3 \int_{-1}^1 x^3\,\mathrm{d}x = \int_{-1}^1 x|x|\,\mathrm{d}x \\
2c_1 + c_2 \int_{-1}^1 x^3\,\mathrm{d}x + c_3 \int_{-1}^1 x^4\,\mathrm{d}x = \int_{-1}^1 x^2|x|\,\mathrm{d}x
\end{array}\right\} \Rightarrow \left.\begin{array}{c}
2c_1 + 0c_2 + \frac{2}{3}c_3 = 1 \\
0c_1 + \frac{2}{3}c_2 + 0c_3 = 0 \\
\frac{2}{3}c_1 + 0c_2 + \frac{2}{5}c_3 = \frac{1}{2}
\end{array}\right\}.$$

En este ejemplo el único extremo relativo es justamente el mínimo absoluto de la función error y viene dado por $p(x) = \frac{3+15x^2}{16}$, $\forall x \in [-1, 1]$.

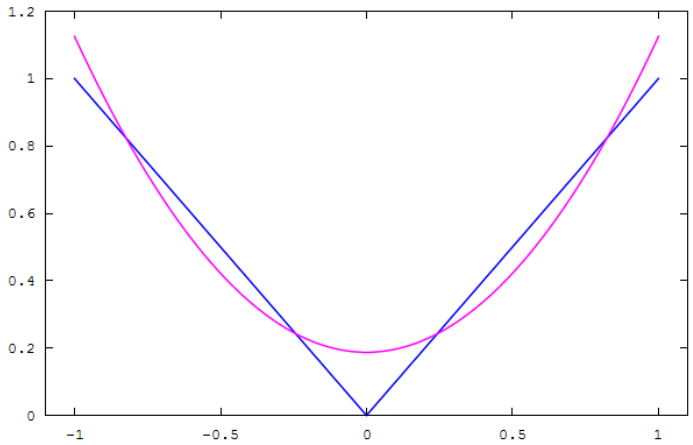

Figura 3: $p(x) = \frac{3+15x^2}{16}$ (en magenta) como aproximación por mínimos cuadrados continuo en $[-1, 1]$ de $f(x) = |x|$ (en azul).