Centro de Investigación en Matemáticas A.C.
Métodos Numéricos(2019)

Tarea 20 - Ajuste Lineal y Cuadrático

Marco Antonio Esquivel Basaldua (MCMESQ)

En la aproximación por mínimos cuadrados se aproximan los datos dados de acuerdo a una función estándar predefinida como por ejemplo una linea recta, un polinomio, una función exponencial, etc. En este trabajo se desarrolla y ejemplifica el ajuste de datos para una función lineal y una función cuadrática.

Métodos Numéricos (2019)

Tarea 20

Marco Antonio Esquivel Basaldua (MCMESQ)

Ajuste Lineal y cuadrático

Supóngase que se tiene que ajustar una función $g(x)$ a n puntos (x_i, y_i); $i = 1, 2, \ldots, n$. En cualquier punto x_i la función evaluada $g(x_i)$ es aproximada al valor exacto y_i. El error en este punto está dado por:

$$e_i = y_i - g(x_i)$$

Usando la suma de los cuadrados de los errores:

$$E = \sum_{i=1}^{n} e_i^2$$

se busca la función $g(x)$ tal que el error E sea el mínimo

→ Linea recta

Consideramos la linea recta $g(x) = a + bx$.

Se busca a y b tales que el error E es mínimo

$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - g(x_i))^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

Para encontrar los puntos extremos de la variable E se utilizan las derivadas parciales con respecto a a y b igualando a cero.

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} \frac{\partial}{\partial a}(y_i - (a+bx_i))^2 = 0$$

$$= \sum_{i=1}^{n} 2(y_i - a - bx_i)(0-1-0) = 0$$

$$= \sum_{i=1}^{n} -2(y_i - a - bx_i) = 0$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i = 0$$

$$\Rightarrow na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (1)$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} \frac{\partial}{\partial b}(y_i - (a+bx_i))^2 = 0$$

$$= \sum_{i=1}^{n} 2(y_i - a - bx_i)(0-0-x_i) = 0$$

$$= \sum_{i=1}^{n} -2x_i(y_i - a - bx_i) = 0$$

$$= \sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0$$

$$= \sum_{i=1}^{n} (x_i y_i - a x_i - b x_i^2) = 0$$

$$= \sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \quad a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (2)$$

Tomando (1) y (2) se obtiene el sistema de ecuaciones:

$$na + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

De forma matricial se obtiene

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

La matriz aumentada queda de la forma

$$\left[\begin{array}{cc|c} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \end{array}\right]$$

Resolviendo por el método de Gauss-Jordan

$$\left[\begin{array}{cc|c} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum x_i y_i \end{array}\right]$$

$$\left[\begin{array}{cc|c} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} y_i \\ 0 & \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 & \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \end{array}\right]$$

$$\left[\begin{array}{cc|c} 1 & \frac{1}{n}\sum_{i=1}^{n} x_i & \frac{1}{n}\sum_{i=1}^{n} y_i \\ 0 & 1 & \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \end{array}\right]$$

Se obtienen las fórmulas cerradas

$$b = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

→ Parábola

De forma similar que para la línea recta, se tiene ahora la función $g(x) = a + bx + cx^2$. El término de error está dado por

$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - g(x_i))^2 = \sum_{i=1}^{n} (y_i - (a + bx_i + cx_i^2))^2$$

Se deriva la variable E en función de a, b y c y se iguala a 0.

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} \frac{\partial}{\partial a} (y_i - (a + bx_i + cx_i^2))^2 = 0$$

$$= \sum_{i=1}^{n} 2(y_i - a - bx_i - cx_i^2)(0-1-0-0) = 0$$

$$= \sum_{i=1}^{n} -2(y_i - a - bx_i - cx_i^2) = 0$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2) = 0$$

$$= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i - \sum_{i=1}^{n} cx_i^2 = 0$$

$$\Rightarrow na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \qquad (3)$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} \frac{\partial}{\partial b}\left(y_i - (a + bx_i + cx_i^2)\right)^2 = 0$$

$$= \sum_{i=1}^{n} 2\left(y_i - a - bx_i - cx_i^2\right)\left(0 - 0 - x_i - 0\right) = 0$$

$$= \sum_{i=1}^{n} -2x_i\left(y_i - a - bx_i - cx_i^2\right) = 0$$

$$= \sum_{i=1}^{n} x_i\left(y_i - a - bx_i - cx_i^2\right) = 0$$

$$= \sum_{i=1}^{n} \left(x_i y_i - ax_i - bx_i^2 - cx_i^3\right) = 0$$

$$= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} ax_i - \sum_{i=1}^{n} bx_i^2 - \sum_{i=1}^{n} cx_i^3 = 0$$

$$\Rightarrow a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i \qquad (4)$$

$$\frac{\partial E}{\partial c} = \sum_{i=1}^{n} \frac{\partial}{\partial c}\left(y_i - (a + bx_i + cx_i^2)\right)^2 = 0$$

$$= \sum_{i=1}^{n} 2\left(y_i - a - bx_i - cx_i^2\right)\left(-x_i^2\right) = 0$$

$$= \sum_{i=1}^{n} -2x_i^2\left(y_i - a - bx_i - cx_i^2\right) = 0$$

$$= \sum_{i=1}^{n} x_i^2\left(y_i - a - bx_i - cx_i^2\right) = 0$$

$$= \sum_{i=1}^{n} \left(x_i^2 y_i - ax_i^2 - bx_i^3 - cx_i^4\right) = 0$$

$$\Rightarrow a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i \qquad (5)$$

Tomando (3), (4) y (5) se obtiene el sistema de ecuaciones:

$$na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

De forma matricial se obtiene:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix}$$

Al resolver el sistema de ecuaciones se obtiene

$$c = \frac{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]\left[\sum_{i=1}^{n} x_i^2 y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right)\right] - \left[\sum_{i=1}^{n} x_i^3 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} x_i\right)\right]\left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right]}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]\left[\sum_{i=1}^{n} x_i^4 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)^2\right] - \left[\sum_{i=1}^{n} x_i^3 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} x_i\right)\right]^2}$$

$$b = \frac{\left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right]\left[\sum_{i=1}^{n} x_i^4 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)^2\right] - \left[\sum_{i=1}^{n} x_i^2 y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum y_i\right)\right]\left[\sum_{i=1}^{n} x_i^3 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} x_i\right)\right]}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]\left[\sum_{i=1}^{n} x_i^4 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)^2\right] - \left[\sum_{i=1}^{n} x_i^3 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} x_i\right)\right]^2}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i - c\sum_{i=1}^{n} x_i^2}{n}$$

Implementación y Resultados

Para el caso del ajuste lineal se crearon 201 datos sintéticos a partir de la función $f(x) = 3x + 2$ en el dominio $[0, 200]$ a la cual se le agregó un término de ruido con la ayuda del software *Matlab*. Estos datos generados se encuentran en el archivo *3x+2.txt* dentro de la carpeta *Lineal* mismos que son cargados como parámetro de entrada del archivo *linear_regression.c*. Al generar el ajuste lineal se obtiene la función $g(x) = 7.494367 + 2.96541x$ con un error total acumulado $E = 111375$. Los puntos originales y el ajuste lineal pueden se pueden apreciar en la figura 1.

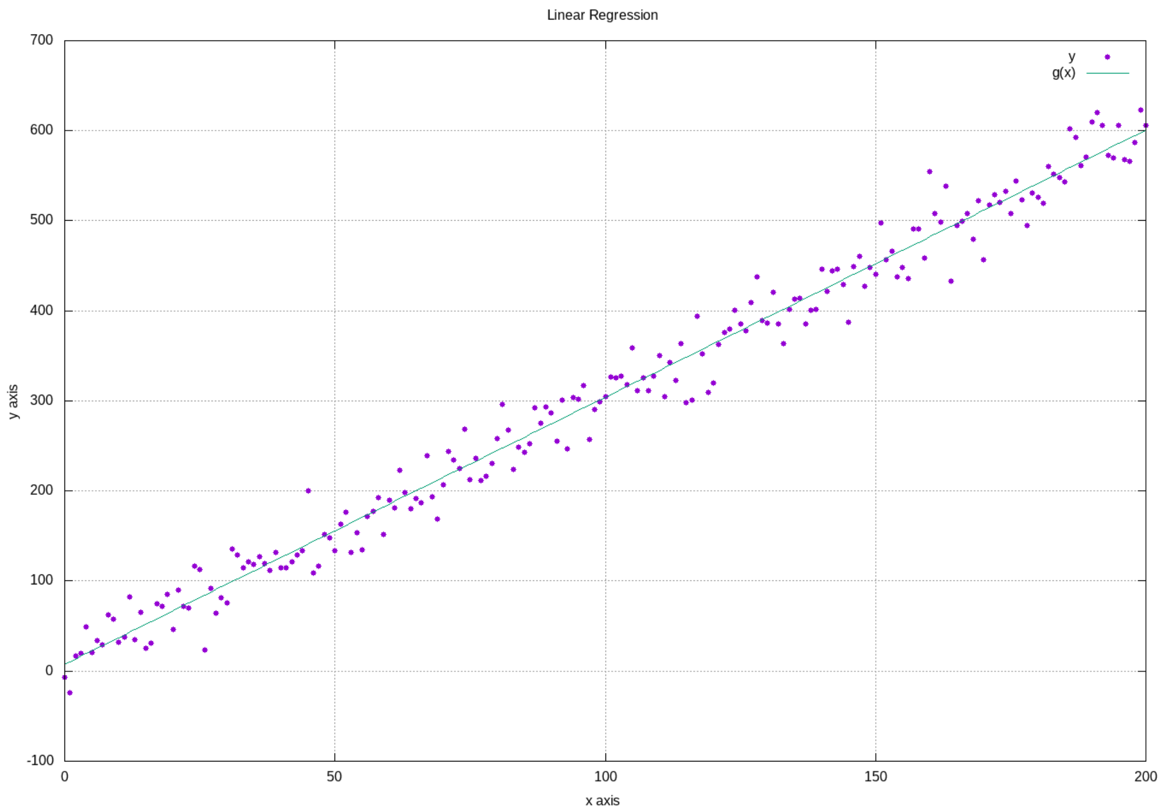

Figura 1. Aproximación Lineal

Para el caso del ajuste cuadràtico se crearon 101 datos sintéticos a partir de la función $f(x) = -2x^2 + 1$ en el dominio $[0, 10]$ a la cual se le agregó un término de ruido con la ayuda del software *Matlab*. Estos datos generados se encuentran en el archivo *2xx+1.txt* dentro de la carpeta *Cuadratica* mismos que son cargados como parámetro de entrada del archivo *quadratic_regression.c*. Al generar el ajuste lineal se obtiene la función $g(x) = -2.914567 + 2.302633x - 2.199210x^2$ con un error total acumulado $E = 30571.7$. Los puntos originales y el ajuste lineal pueden se pueden apreciar en la figura 2.

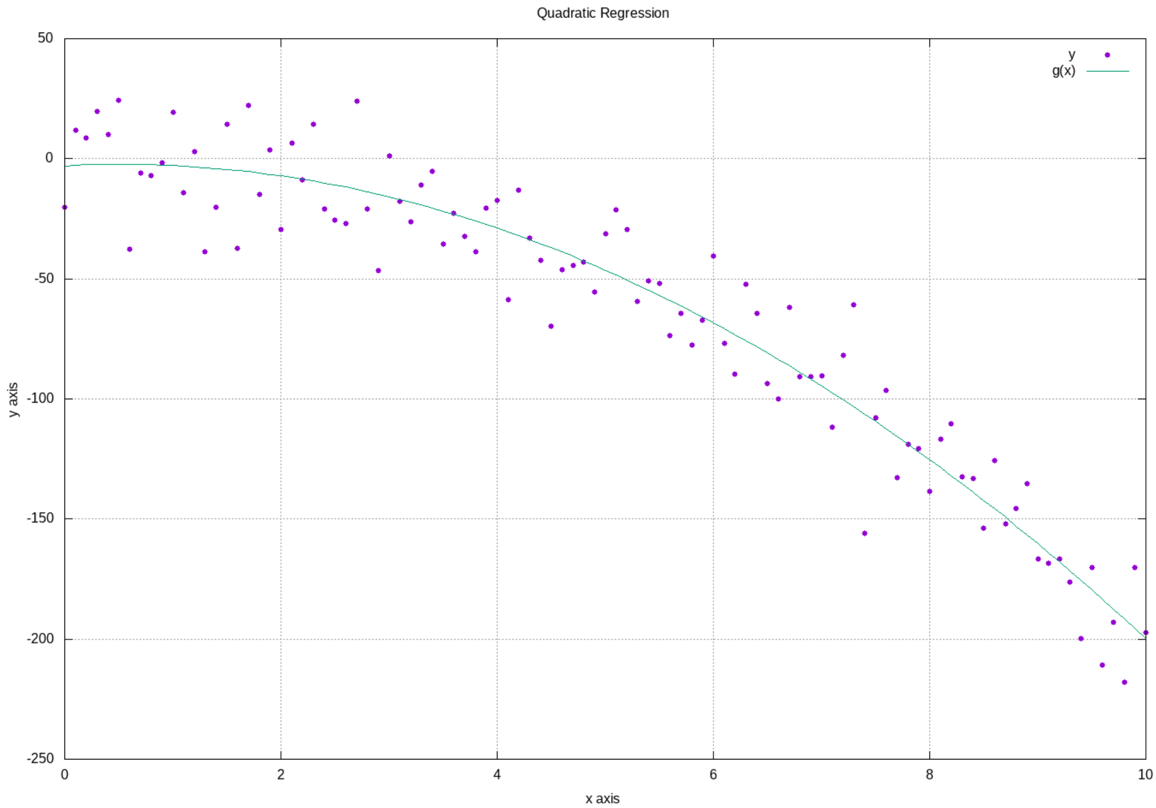

Figura 2. Ajuste Cuadrático.