Ajuste de curvas e interpolación para predecir el comportamiento de datos

Neil Otniel Moreno Rivera, 373687
Universidad de Guanajuato, no.morenorivera@ugto.mx

***Resumen*—Este documento se explicarán las actividades que se realizaron para el ajuste de curvas y para la interpolación en el programa MATLAB, como también explicar en que se basaron estas actividades para obtener un resultado correcto y convincente.**

***Abstract*-- This document will explain the activities that were carried out for the adjustment of curves and for the interpolation in the MATLAB program, as well as explain what these activities were based on to obtain a correct and convincing result.**

I. Introducción

El ajuste de curvas y la interpolación son técnicas utilizadas en el análisis de datos para modelar y predecir el comportamiento de variables dependientes a partir de variables independientes. Estas técnicas se basan en encontrar una función matemática que se ajuste de manera óptima a un conjunto de puntos conocidos, permitiendo estimaciones y predicciones.

El ajuste de curvas busca una función continua que se ajuste a los datos, minimizando la discrepancia entre los valores observados y los estimados. La interpolación, por otro lado, estima valores desconocidos dentro del rango de los datos conocidos.

En este documento, exploraremos el ajuste de curvas e interpolación, analizando técnicas comunes como la regresión polinómica y la interpolación de lineal. Estas técnicas son fundamentales en diversos campos científicos y de ingeniería, permitiendo obtener información valiosa y realizar proyecciones confiables con datos limitados.

II. Ajuste de curvas

El ajuste de curvas es un proceso mediante el cual, dado un conjunto de N pares de puntos {xi, yi} (siendo x la variable independiente e y la dependiente), se determina una función matemática f(x) de tal manera que la suma de los cuadrados de la diferencia entre la imagen real y la correspondiente obtenida mediante la función ajustada en cada punto sea mínima. [1]

Utilizamos varios métodos, entre ellos la regresión lineal, La regresión lineal es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos. Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente y (también conocida como la respuesta) como una función de una o varias variables independientes Xi (denominadas predictores). La ecuación general correspondiente a un modelo de regresión lineal es:

$$Y = \beta_0 + \sum \beta_K X_K + \epsilon_i$$

Donde β, representa las estimaciones de parámetros lineales que se deben calcular y $\in$ representa los términos del error. [2]

Si nos damos cuenta, esto tiene forma de la ecuación de una recta, gracias a esto podemos conocer la relación que se tiene entre los dos ejes con una línea recta.
Si desglosamos la ecuación derivando la ecuación, podemos obtener los siguientes despejes:

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \quad b = \frac{\sum y - a\sum x}{n}$$

Tomando en cuenta la fórmula de la ecuación de la recta.

$$y = ax + b$$

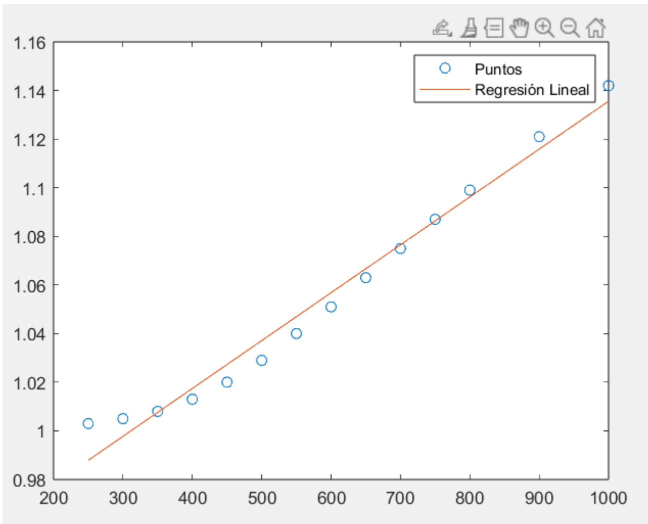

Fig. 1. Grafica del ajuste lineal con los datos dados del problema.

Otro método del que usamos fue el ajuste polinómico, Básicamente el ajuste de curvas se utiliza cuando se tiene una serie de datos calculados y se desea conocer valores intermedios no conocidos o también en aquellos casos que se desee una versión simplificada de una función que se ajuste a un número de valores concretos y posteriormente usar la función simplificada para derivar nuevos valores. [3]

* Universidad de Guanajuato – Neil Otniel Moreno Rivera

Para este caso, usamos el siguiente planteamiento, usamos una resolución que se basa en lo mismo, el planteamiento original, es un sistema de ecuaciones, similar a este

$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} = y1$$
$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} = y2$$
$$\ldots$$
$$a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} = yn$$

Luego resolver este sistema de ecuaciones para encontrar los coeficientes, para saber una aproximación polinómica.
Para esto, es necesario que n, sea igual al número de ecuaciones.

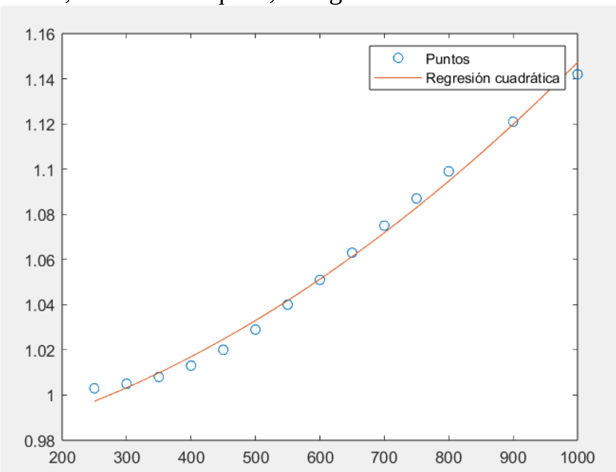

Fig. 2. Grafica del ajuste polinomial con los datos dados del problema.

Con eso terminamos el ajuste de curvas lineal y polinómico, por lo que pasamos a la interpolación.

III. Interpolación

La interpolación es un proceso para estimar valores que quedan entre puntos de datos conocidos.
La interpolación implica construir una función f cuyo valor sea el de unos valores de datos proporcionados yi, en unos sitios de datos proporcionados xi, de forma que f(xi) = yi, para todo i.
La interpolación f se construye normalmente como una función única de la forma. En este caso se uso la interpolación lineal y la interpolación de grado n.
Para la interpolación lineal, es un método utilizado para estimar valores desconocidos dentro de un rango dado utilizando una línea recta que conecta dos puntos conocidos. Este método asume una relación lineal entre los puntos conocidos y se utiliza ampliamente en diversos campos, como la matemática, la física, la ingeniería y la programación.
El uso de la interpolación lineal, es similar al del ajuste de lineal, por lo que se usan las mismas formulas, o sea la formula de una recta con dos puntos, pero con la gran diferencia de que la recta solamente existe entre los dos puntos, por lo que se toma dos puntos para calcular pequeñas rectas e ir creando algo que se ajuste mejor a la relación entre X y Y.

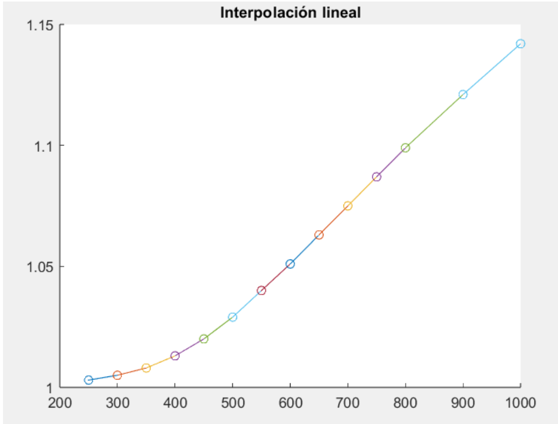

Fig. 3. Grafica de la interpolación lineal con los datos dados.

Ahora para la interpolación de grado n, o de cualquier grado, puede ser resuelta de muchas formas, en este documento solo se explicarán dos, una auto ajustable y otra que es más mecánica.
Primero comenzamos con la que es mecánica, con mecánica hacemos referencia a que se tiene que hacer lo mismo, muchas veces, en este caso se resolvió de la siguiente manera.
Se pensó como si fuera un ajuste polinomial, explicado en la parte superior, pero tomando solo los puntos de grado n, por lo que el resolver un sistema de más de 10 ecuaciones comienza a ser cansado y algo innecesario.

Por lo que queda la otra opción, que siempre es mejor, la cual es usar la matriz de Vandermonde, esta opción fue obtenida de un ensayo titulado "Matriz Vandermonde y ajuste polinómico, como herramienta para estimar la duración del área foliar en el maíz", la matriz de Vandermonde es una matriz cuadrada que se utiliza en la interpolación, esta matriz se construye a partir de un conjunto de puntos conocidos, esta matriz es adaptativa, puesto que podemos hacer que varie con el grado que tomamos y no cambiamos sus propiedades, la cual es que de todos los puntos al resolver la matriz por determinantes tenemos una diferencia de cada uno de los valores introducidos, esto nos ayuda a resolver un sistema de ecuaciones de una forma sencilla y casi sin darnos cuenta encontrar los exponentes, para nuestra ecuación de grado n, que pasa por todos los puntos tomados.

$$|V| = (\alpha_2 - \alpha_1)(\alpha_3 - \alpha_1)\ldots(\alpha_n - \alpha_1)\begin{vmatrix} 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{n-2} \\ 1 & \alpha_3 & \alpha_3^2 & \ldots & \alpha_3^{n-2} \\ 1 & \alpha_4 & \alpha_4^2 & \ldots & \alpha_4^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \ldots & \alpha_n^{n-2} \end{vmatrix}$$

Fig. 4. Matriz de Vandermonde.

IV. Desarrollo

Para el desarrollo de esto, se uso el programa MATLAB, en este se estableció un arreglo con los datos proporcionados y su respectivo eje, en este caso el contexto era una serie de datos que tenían un eje en la temperatura y otro en la energía a presión constante.

Temperatura, K	c_p kJ/kg · K	c_v kJ/kg · K	k
		Aire	
250	1.003	0.716	1.401
300	1.005	0.718	1.400
350	1.008	0.721	1.398
$\boxed{400}$	1.013	0.726	1.395
450	1.020	0.733	1.391
500	1.029	0.742	1.387
550	1.040	0.753	1.381
600	1.051	0.764	1.376
650	1.063	0.776	1.370
700	1.075	0.788	1.364
750	1.087	0.800	1.359
800	1.099	0.812	1.354
900	1.121	0.834	1.344
1000	1.142	0.855	1.336

Imagen 1. Datos del problema.

Con estos datos fueron cargador, en particular con las dos primeras columnas, tomándose como Y y X respectivamente, en este caso la resolución se llevo acabo por medio de funciones, con esto hago referencia a un programa que contenga los datos y otro que tenga los procedimientos o métodos a aplicar, en este caso se desarrollaron 2 códigos para ajuste lineal y para ajuste cuadrático, obteniendo resultados aceptables. Siendo las figuras 1 y 2 de este documento.

Luego se opto por hacer un par de códigos para la interpolación por definición. Que fue la interpolación lineal, ya vista en la figura 3 y para la interpolación de grado 13, siento n-1, el número máximo de grado al que se puede llegar en cualquier conjunto de puntos. Obteniendo esta gráfica:

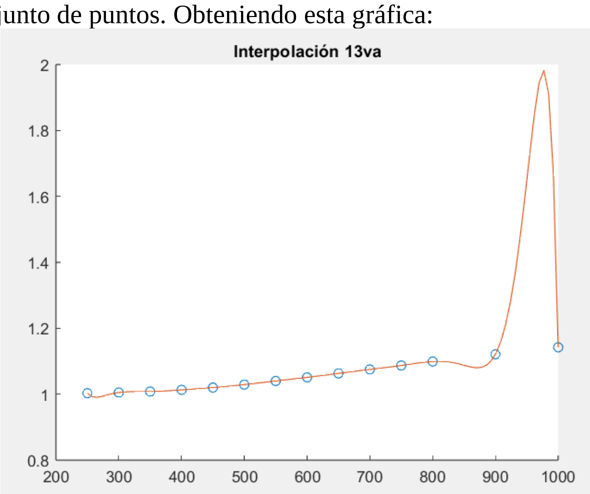

Fig. 5 Grafica de la interpolación de grado 13.

Y ya por ultimo se busco hacer una función en la cual pudiera ser colocado el grado que se quiere interpolar, la cual es la que se realizó con la matriz de Vandermonde. Aunque también existe otra posibilidad, puesto que, por las facilidades que da MATLAB, podemos hacer una función que llame a otras funciones ya inscritas en MATLAB, encuentran una ecuación que se ajuste a la forma como lo es POLYFIT y otra que evalúe los puntos de la función con POLYVAL. Pero como se buscaba hacer por la definición esta no se tomo tanto en cuenta, aunque se incluye en los anexos.

V. Conclusiones

Como conclusión, viendo los gráficos, podemos decir que se tiene que buscar un grado coherente para la interpolación, porque por su misma naturaleza esta puede causar problemas o datos sin sentido como lo es mostrado en la figura 5, la cual tiene un despunte, extraño y si tratamos de aproximar datos intermedios o predecir nuevos datos, nos causará datos erróneos.

Por el lado de el ajuste lineal, considero que no realiza la misma función de una interpolación, pues creo que analiza una tendencia, buscando predecir a grandes rasgos, pudiendo colocarle una apertura al ajuste para conocer donde pueden estar nuestros futuros datos, en base a su tendencia, aunque siempre siendo poco confiable porque no ofrece más allá de un quizás.

Y por lo mostrado y lo usado que son ambas podemos decir que ambas son usadas ampliamente en varias diciplinas y sin ellas seria complicado el desarrollo científico.

Anotaciones

Este reporte fue realizado con lo que recordaba y lo re obtenido de fuentes que ya había visitado, si una información no se entiende, es porque no supe ordenar mis palabras.

Referencias

[1] SALVADOR, PEDRO. *Técnicas Computacionales*. 2007.
[2] "¿Qué Es La Regresión Lineal?" *La.mathworks.com*, la.mathworks.com/discovery/linear-regression.html.
[3] \376\377\000E\000d\000g\000a\000r\000 \000R\000o\000m\000e\000r\000o\000 \000R\000. AJUSTE de CURVAS. 2007.
[4] "Métodos de Interpolación - MATLAB & Simulink - MathWorks América Latina." La.mathworks.com, la.mathworks.com/help/curvefit/interpolation-methods.html.
[5] J Díaz López, Ernesto, et al. Matriz Vandermonde Y Ajuste Polinómico, Como Herramienta Para Estimar La Duración Del Área Foliar En El Maíz. Dec. 2014.
[6]

Bibliografia

- SALVADOR, PEDRO. *Técnicas Computacionales*. 2007.
- "¿Qué Es La Regresión Lineal?" *La.mathworks.com*, la.mathworks.com/discovery/linear-regression.html.
- Universidad de Guanajuato. "Clase Digital 2. Ajuste de Curvas Por Mínimos Cuadrados - Recursos Educativos Abiertos." Recursos Educativos Abiertos, 25 Jan. 2023, blogs.ugto.mx/rea/clase-digital-2-ajuste-de-curvas-por-minimos-cuadrados/. Accessed 1 June 2023.
- J Díaz López, Ernesto, et al. Matriz Vandermonde Y Ajuste Polinómico, Como Herramienta Para Estimar La Duración Del Área Foliar En El Maíz. Dec. 2014.
- Chapra, Steven C, and Raymond P Canale. Numerical Methods for Engineers. 7th ed., New York, Mcgraw-Hill Education, Cop, 2015.

ANEXOS

Interpolación o codigo fuente

```
Temp = [ 250 300 350 400 450 500 550 600 650 700 750
800 900 1000];
P_C = [ 1.003 1.005 1.008 1.013 1.020 1.029 1.040
1.051 1.063 1.075 1.087 1.099 1.121 1.142];
%Regresion cuadratica
n = length(Temp);
sum_x = sum(Temp);
sum_y = sum(P_C);
sum_x2 = sum(Temp.^2);
sum_xy = sum(Temp.*P_C);
sum_x3 = sum(Temp.^3);
sum_x4 = sum(Temp.^4);
sum_x2y = sum(Temp.^2.*P_C);

A = [n, sum_x, sum_x2; sum_x, sum_x2, sum_x3; sum_x2,
sum_x3, sum_x4];
B = [sum_y; sum_xy; sum_x2y];
coef = A\B;

x_reg = linspace(min(Temp), max(Temp), 100);
y_reg = coef(1) + coef(2).*x_reg + coef(3).*x_reg.^2;

plot(Temp, P_C, 'o', x_reg, y_reg);
legend('Puntos', 'Regresión cuadrática');
figure;

% Regresion lineal
n = length(Temp);
sum_x = sum(Temp);
sum_y = sum(P_C);
sum_x2 = sum(Temp.^2);
sum_xy = sum(Temp.*P_C);
m = (n*sum_xy - sum_x*sum_y) / (n*sum_x2 - sum_x^2);
b = (sum_y - m*sum_x) / n;
x_reg_1 = linspace(min(Temp), max(Temp), 100);
y_reg_1 = m*x_reg_1 + b;

plot(Temp, P_C, 'o', x_reg_1, y_reg_1);
legend('Puntos', 'Regresión Lineal');

figure;
title("Interpolación lineal");
% Interpolación lineal
i=1;
j=1;
for i = 1:length(Temp) - 1
    linea(Temp(i),P_C(j),Temp(i+1),P_C(j+1));
    i=i+1;
    j=j+1;
end
hold off;
figure;
title("Interpolación 13va");
% Interpolación 13va
i=1;
j=1;
for i=1: length(Temp) -13

treceava(Temp(i),P_C(j),Temp(i+1),P_C(j+1),Temp(i+2),
P_C(j+2),Temp(i+3),P_C(j+3),Temp(i+4),P_C(j+4),Temp(i
+5),P_C(j+5),Temp(i+6),P_C(j+6),Temp(i+7),P_C(j+7),Te
mp(i+8),P_C(j+8),Temp(i+9),P_C(j+9),Temp(i+10),P_C(j+
10),Temp(i+11),P_C(j+11),Temp(i+12),P_C(j+12),Temp(i+
13),P_C(j+13));
    i=i+1;
    j=j+1;
end
hold off;
n=6;
figure;
% interpolacion_n(Temp, P_C, n)
automatica(Temp,P_C,n);
```

FUNCION DE INTERPOLACIÓN LINEAL

```
function linea(x1, y1, x2, y2)
m = (y2 - y1) / (x2 - x1);
b = y1 - m * x1;
hold on;
plot([x1 x2], [y1 y2], 'o-')
end
```

FUNCION DE INTERPOLACIÓN 13VA

```
function treceava(x, y, x1, y1, x2, y2, x3, y3, x4, y4, x5, y5, x6, y6, x7, y7, x8,
y8, x9, y9, x10, y10, x11, y11, x12, y12, x13, y13)
    A = [x^13 x^12 x^11 x^10 x^9 x^8 x^7 x^6 x^5 x^4 x^3 x^2 x 1;
         x1^13 x1^12 x1^11 x1^10 x1^9 x1^8 x1^7 x1^6 x1^5 x1^4 x1^3 x1^2 x1 1;
         x2^13 x2^12 x2^11 x2^10 x2^9 x2^8 x2^7 x2^6 x2^5 x2^4 x2^3 x2^2 x2 1;
         x3^13 x3^12 x3^11 x3^10 x3^9 x3^8 x3^7 x3^6 x3^5 x3^4 x3^3 x3^2 x3 1;
         x4^13 x4^12 x4^11 x4^10 x4^9 x4^8 x4^7 x4^6 x4^5 x4^4 x4^3 x4^2 x4 1;
         x5^13 x5^12 x5^11 x5^10 x5^9 x5^8 x5^7 x5^6 x5^5 x5^4 x5^3 x5^2 x5 1;
         x6^13 x6^12 x6^11 x6^10 x6^9 x6^8 x6^7 x6^6 x6^5 x6^4 x6^3 x6^2 x6 1;
         x7^13 x7^12 x7^11 x7^10 x7^9 x7^8 x7^7 x7^6 x7^5 x7^4 x7^3 x7^2 x7 1;
         x8^13 x8^12 x8^11 x8^10 x8^9 x8^8 x8^7 x8^6 x8^5 x8^4 x8^3 x8^2 x8 1;
         x9^13 x9^12 x9^11 x9^10 x9^9 x9^8 x9^7 x9^6 x9^5 x9^4 x9^3 x9^2 x9 1;
         x10^13 x10^12 x10^11 x10^10 x10^9 x10^8 x10^7 x10^6 x10^5 x10^4 x10^3 x10^2
x10 1;
         x11^13 x11^12 x11^11 x11^10 x11^9 x11^8 x11^7 x11^6 x11^5 x11^4 x11^3 x11^2
x11 1;
         x12^13 x12^12 x12^11 x12^10 x12^9 x12^8 x12^7 x12^6 x12^5 x12^4 x12^3 x12^2
x12 1;
         x13^13 x13^12 x13^11 x13^10 x13^9 x13^8 x13^7 x13^6 x13^5 x13^4 x13^3 x13^2
x13 1];
    b = [y; y1; y2; y3; y4; y5; y6; y7; y8; y9; y10; y11; y12; y13];
    coef = A \ b;
    a = coef(1);
    b = coef(2);
    c = coef(3);
    d = coef(4);
    e = coef(5);
    f = coef(6);
    g = coef(7);
    h = coef(8);
    i = coef(9);
    j = coef(10);
    k = coef(11);
    l = coef(12);
    m = coef(13);
    x_reg = linspace(min([x x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13]), max([x x1
x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13]), 100);
    y_reg = a*x_reg.^13 + b*x_reg.^12 + c*x_reg.^11 + d*x_reg.^10 + e*x_reg.^9 +
f*x_reg.^8 + g*x_reg.^7 + h*x_reg.^6 + i*x_reg.^5 + j*x_reg.^4 + k*x_reg.^3 +
l*x_reg.^2 + m*x_reg + coef(14);
    hold on;
    plot([x x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13], [y y1 y2 y3 y4 y5 y6 y7 y8
y9 y10 y11 y12 y13], 'o');
    plot(x_reg, y_reg);
end
```

FUNCION DE INTERPOLACIÓN GRADO N

```
function interpolacion_n(x, y, n)
% Consultar si el grado es válido
if (n < 1 || n > length(x) - 1)
    error('Dato incorrecto.')
end

% Crear la matriz de Vandermonde
V = diag(ones(1, n)) - diag(x(1:n), -1) - diag(x(n:-
1:1), 1);

% Resolver el sistema lineal para los coeficientes
a = V\y;

% Crea la ecuación
ecu = a(1) * x^n + a(2) * x^(n - 1) + ... + a(n);

% Graficar
scatter(x, y, 'o');
hold on;
plot(x, ecu);

title(['Interpolación de grado: ' num2str(n)]);
end
```