(#396). INTERPOLACIÓN (IV): MÉTODO DE SPLINE CÚBICO

[MONOTEMA] Los métodos de interpolación vistos hasta ahora no permitían aproximar a la función por tramos. El método de spline sí que lo puede hacer, y es muy potente para conseguir buenos resultados en cualquier tipo de análisis de datos. Seguiremos a Burden, Faires & Burden (2017), y lo enfocaremos, como siempre, de manera simplificada.

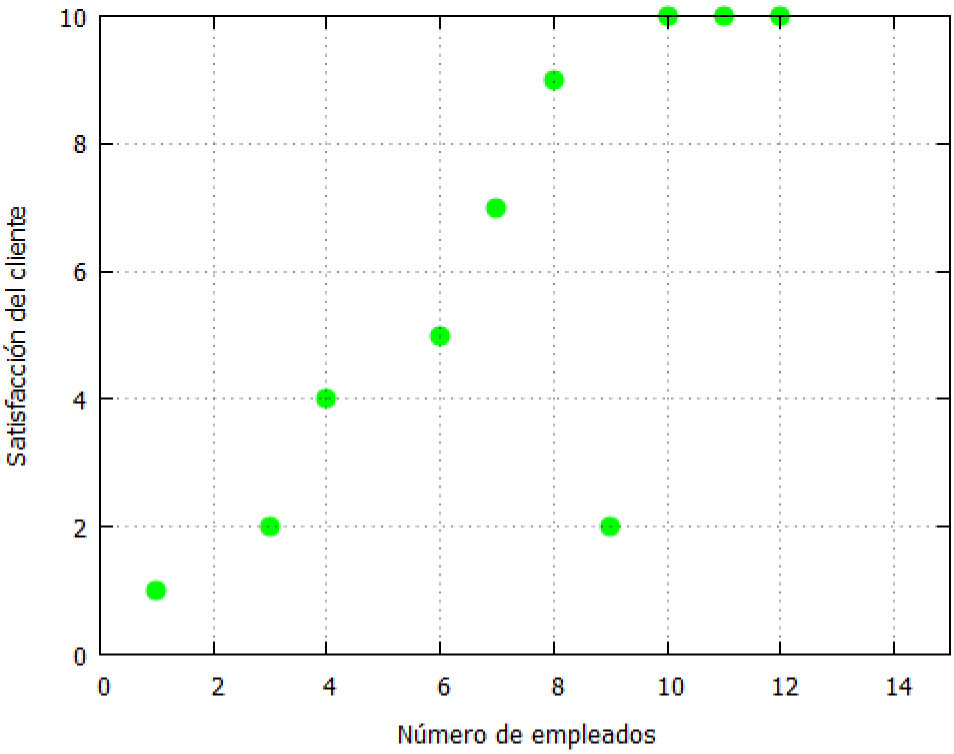
Datos de partida

Usaremos los mismos datos que en el método de Lagrange:

x	f(x)
1	1
3	2
4	4
6	5
7	7
8	9
9	2
10	10
11	10
12	10

Estos datos relacionan la cantidad de empleados utilizados en un gran supermercado (x) con la saltisfacción del cliente $y = f(x)$. La satisfacción del cliente se mide en una escala de 0 a 10, donde 0 es el valor mínimo y 10 el valor máximo.

```
x_:[1,3,4,6,7,8,9,10,11,12];
fx_:[1,2,4,5,7,9,2,10,10,10];
plot2d([discrete, x_, fx_],
[x,0,15],[y,0,10], [style, points], [color,green],
[xlabel, “Número de empleados”],
[ylabel, “Satisfacción del cliente”], [legend, false]);
```



Método de spline cúbico

(1) Objetivo: Aproximarse numéricamente a la función $f(x)$, de la que conocemos sólo ciertos datos por medio de varios polinomios que pasen por los *i* puntos conocidos, es decir, calcularemos polinomios por tramos de datos. Emplearemos, además, splines cúbicos, es decir, polinomios de orden 3.

(2) Condiciones: La función $f(x)$ debe ser continua en el intervalo $[a,b]$ considerado y además tiene que ser doblemente diferenciable.

(3) Descripción rápida: Partimos de la serie de datos $(x_i, f(x_i))$, los cuales es conveniente ordenar de manera creciente por los nodos x_i, y escogemos los *j* subintervalos

para los cuales vamos a calcular splines $S_j(x)$.

La condiciones que deben satisfacerse son las siguientes:

a) $S_j(x)$ son los *j* polinomios que se calculan para los *i*=*j*+1 nodos.

b) $S_j(x_j) = f(x_j)$ y $S_j(x_{j+1}) = f(x_{j+1})$ para cada *j*. Es decir, el spline cúbico debe pasar por todas las imágenes de los datos, y esto hace que dos splines consecutivos pasen por un nodo común, por lo que al evaluar ambos splines en el mismo punto el resultado deba ser el mismo.

c) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ y $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ para cada *j*=1…*n*-2. No sólo las evaluaciones de los splines, sino también sus derivadas primera y segunda deben coincidir en los nodos de conexión.

d) Si el spline es natural (frontera libre), entonces: $S''(x_1) = S''(x_n) = 0$

Los splines de frontera condicionada no los vamos a comentar ya que es necesario conocer los valores de la derivada de la función original en los extremos, algo de lo que carecemos, debido a que sólo tenemos un conjunto de datos brutos y sus imágenes.

Estas condiciones suponen unas restricciones en la resolución de un sistema de ecuaciones con 4x*j* incógnitas. Es decir, 4 incógnitas por cada spline calculado. Así, para *n*=3 puntos obtendríamos *j*=2 splines cúbicos, de la forma:

$$S_1(x) = a_1 + b_1(x-1) + c_1(x-1)^2 + d_1(x-1)^3$$

$$S_2(x) = a_2 + b_2(x-3) + c_2(x-3)^2 + d_2(x-3)^3$$

Ahora tenemos que determinar esas constantes usando las restricciones comentadas anteriormente.

b) $a_1 = f(x_1)$, $f(x_2) = a_1 + 2b_1 + 4c_1 + 8d_1$, $f(x_2) = a_2$, $a_2 + b_2 + c_2 + d_2 = f(x_3)$

c) $S_1'(x_2) = S_2'(x_2)$, por lo que $b_1 + 4c_1 + 12d_1 = b_2$

y $S_1''(x_2) = S_2''(x_2)$, por lo que $2c_1 + 12d_1 = 2c_2$

d) $S_1''(x_1) = 0$, por lo que $2c_1 = 0$

y $S_1''(x_3) = 0$, por lo que $2c_2 + 6d_2 = 0$

Ya sólo queda resolver el sistema de ecuaciones anterior, es decir, y considerando los nodos [1,3,4] e imágenes [1,2,4]:

$$\begin{aligned} a_1 &= 1 \\ a_1 + 2b_1 + 4c_1 + 8d_1 &= 2 \\ a_2 &= 2 \\ a_2 + b_2 + c_2 + d2 &= 4 \\ b_1 + 4c_1 + 12d_1 &= b_2 \\ 2c_1 + 12d_1 &= 2c_2 \\ 2c_1 &= 0 \\ 2c_2 + 6d_2 &= 0 \end{aligned}$$

Implementación en Maxima

Vamos a realizar nuestra estimación en Maxima. Supongamos, como en el caso del método de Lagrange, que queremos aproximarnos a un valor que desconocemos, cuando el número de empleados es 2. Para ello vamos a coger los 3 primeros puntos, con la programación siguiente:

```
x_:[1,3,4];
fx:[1,2,4];
S1: a1+b1*(x-x_[1])+c1*(x-x_[1])^2+d1*(x-x_[1])^3;
S2: a2+b2*(x-x_[2])+c2*(x-x_[2])^2+d2*(x-x_[2])^3;
fx1:ev(S1,x=x_[1]);
fx2:ev(S1,x=x_[2]);
fx2_:ev(S2,x=x_[2]);
fx3:ev(S2,x=x_[3]);
S1prima: diff(S1,x);
S1prima_ev:ev(S1prima,x=x_[2]);
S2prima: diff(S2,x);
S2prima_ev:ev(S2prima,x=x_[2]);
S1primaprima: diff(S1prima,x);
S1primaprima_ev:ev(S1primaprima,x=x_[2]);
S2primaprima: diff(S2prima,x);
S2primaprima_ev:ev(S2primaprima,x=x_[2]);
S1primaprimainicio_ev:ev(S1primaprima,x=x_[1]);
S2primaprimafin_ev:ev(S2primaprima,x=x_[3]);
e1: fx1-fx[1];
e2: fx2-fx[2];
e3: fx2_-fx[2];
e4:fx3-fx[3];
e5:S1prima_ev-S2prima_ev;
e6:S1primaprima_ev-S2primaprima_ev;
e7:S1primaprimainicio_ev-0;
e8: S2primaprimafin_ev-0;
algsys ([e1, e2, e3, e4,e5,e6,e7,e8],
[a1,b1,c1,d1,a2,b2,c2,d2]);
```

Aunque a priori parece un código algo complejo, lo cierto es que sigue paso a paso el algoritmo que deviene de las restricciones planteadas. Para formular el sistema de ecuaciones a resolver, hay que recordar que debemos escribir esas ecuaciones en Maxima de la forma x-a=0 y no de la forma x=a.

Con el comando “algsys” resolvemos el sistema de ecuaciones, y obtenemos la siguiente solución:

$a_1 = 1$
$b_1 = 0$
$c_1 = 0$
$d_1 = \frac{1}{8}$
$a_2 = 2$
$b_2 = \frac{3}{2}$
$c_2 = \frac{3}{4}$
$d_2 = -\frac{1}{4}$

Por tanto, podemos construir las dos splines:

$$S_1(x) = 1 + \frac{1}{8}(x-1)^3, x \in [1,3]$$

$$S_2(x) = 2 + \frac{3}{2}(x-2) + \frac{3}{4}(x-2)^2 - \frac{1}{4}(x-2)^3, x \in [3,4]$$

Y ahora hacemos la comprobación con Maxima. Esas splines evaluadas en los nodos tienen que dar sus imágenes correspondientes:

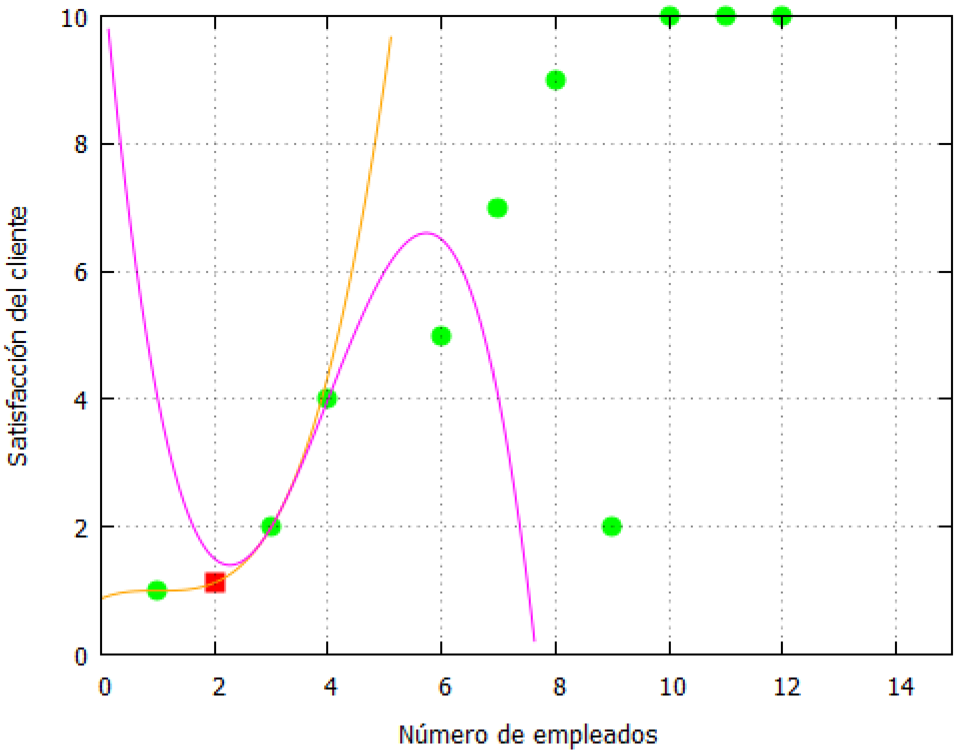
```
S1solucion: 1+(1/8)*(x-x_[1])^3;
S2solucion: 2+(3/2)*(x-x_[2])+(3/4)*(x-x_[2])^2-(1/4)*(x-x_[2])^3;
Comprobacionnodo1: ev(S1solucion,x=x_[1]);
Comprobacionnodo2: ev(S1solucion,x=x_[2]);
Comprobacionnodo2_: ev(S2solucion,x=x_[2]);
Comprobacionnodo3: ev(S2solucion,x=x_[3]);
```

Como vemos, los resultados coinciden, lo que indica que hemos hecho bien la programación y ya tenemos las 2 splines cúbicas.

Podemos dibujar las splines y estimar el valor para el cual queríamos hacer la interpolación, es decir, ¿cuál sería el valor de satisfacción del cliente con 2 empleados en el establecimiento?.

Para ello empleamos la primea spline que es la que pasa por los dos nodos que contienen al valor en cuestión, y su resultado es 1.125. Con Maxima lo podemos hacer fácilmente, y además representarlo en un gráfico:

```
x_:[1,3,4,6,7,8,9,10,11,12];
fx:[1,2,4,5,7,9,2,10,10,10];
x_interpolar:2;
f_S1solucion: ev(S1solucion,x=x_interpolar), numer;
plot2d([[discrete, x_, fx], [discrete,
[[x_interpolar,f_S1solucion]]],S1solucion,S2solucion],
[x,0,15],[y,0,10], [style, points, points, lines,lines],
[color,green, red, orange,magenta],
[xlabel, “Número de empleados”],
[ylabel, “Satisfacción del cliente”], [legend, false]);
```



La curva naranja representa la primera spline y la curva rosa la segunda.

Podemos hacer una comprobación con el comando de Maxima “cspline” y veremos que nos da el mismo resultado:

```
S1solucion: 1+(1/8)*(x-x_[1])^3, expand;
S2solucion: 2+(3/2)*(x-x_[2])+(3/4)*(x-x_[2])^2-(1/4)*(x-
x_[2])^3,expand;
load (interpol);
p:[[1,1],[3,2],[4,4]];
cspline(p);
```

Conclusión

Las splines cúbicas son una muy interesante forma de realizar interpolaciones, probablemente una de las que más ventajas conlleva y que permite realizar aproximaciones más fiables. Su procedimiento en Maxima es un poco laborioso, pero si seguirmos ordenadamente los pasos que surgen de las restricciones del método, no debemos tener problema.

Todos los posts relacionados

(#428) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIc)

(#427) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIb)

(#426) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIa)

(#424) TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (I)

(#422). JUGADORES ACB AL NIVEL DE NBA EN LA TEMPORADA 2018/19

(#421). LOS JUGADORES MÁS PRODUCTIVOS DE LA ACB 2018/19

(#420). LA PARADOJA DE CHEVALIER DE MERE Y LOS ESTUDIANTES DE MARKETING

(#412). LOS JUGADORES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#411). LOS ROOKIES MÁS PRODUCTIVOS DE LA NBA 2018/19

(#410) ¿QUIÉN HA SIDO EL MEJOR ANOTADOR DE LA FASE REGULAR DE LA EUROLIGA 2018/19?

(#409). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (II)

(#408). ALGUNAS CONSIDERACIONES EN LA ESTIMACIÓN DEL RENDIMIENTO POR MINUTO EN BALONCESTO (I)

(#407). AVANCE DE LOS NOVATOS MÁS PRODUCTIVOS DE LA NBA 2018/19

(#398). AJUSTE DE FUNCIONES (II): CAPACIDAD EXPLICATIVA Y PARTICIÓN DE LA FUNCIÓN

(#397). AJUSTE DE FUNCIONES (I): SPLINES VS MÍNIMOS CUADRADOS

(#396). INTERPOLACIÓN (IV): MÉTODO DE SPLINE CÚBICO

(#395). INTERPOLACIÓN (III): MÉTODO DE DIFERENCIAS DIVIDIDAS

(#394). INTERPOLACIÓN (II): MÉTODO DE NEVILLE

(#393). INTERPOLACIÓN (I): POLINOMIO DE LAGRANGE

(#391). BÚSQUEDA DE SOLUCIONES (V): MÉTODO DE LA POSICIÓN FALSA

(#390). BÚSQUEDA DE SOLUCIONES (IV): MÉTODO DE LA SECANTE

(#389). BÚSQUEDA DE SOLUCIONES (III): METODO DE NEWTON

(#388). BÚSQUEDA DE SOLUCIONES (II): METODO DEL PUNTO FIJO

(#387). BÚSQUEDA DE SOLUCIONES (I): MÉTODO DE LA BISECCIÓN

(#384). MEJOR UN SÓLO ITEM QUE VARIOS PARA MEDIR ACTITUDES

(#381). TABLAS DE TUMORES TOTALES EN EL ESTUDIO DEL NTP SOBRE EFECTOS DE LA RADIACIÓN DE TELÉFONO MÓVIL

(#380). ÍNDICES APROXIMADOS FLEXIBLES EN ECUACIONES ESTRUCTURALES

(377). AVANZANDO EN LA COMPLEJIDAD; DOBLE DEPENDENCIA

(#376). CAMPOS DIRECCIONALES. COMENZANDO A REPRESENTAR LA COMPLEJIDAD

(#373). PREDICCIONES CON BUENAS APROXIMACIONES

(#366). MODELO DE HOMOGENEIDAD EN TABLAS DE 2X2

(#365). BÁSICOS DE ECUACIONES ESTRUCTURALES (V): MATRIZ DE DATOS BRUTOS

(#362). ¿HORMESIS ENTRE EL VOLUMEN DE ENTRENAMIENTO Y LA HIPERTROFIA?

(#358). BÁSICOS DE ECUACIONES ESTRUCTURALES (IV): REGRESIÓN LINEAL SIMPLE CON FIABILIDADES DIVERSAS

(#354). BÁSICOS DE ECUACIONES ESTRUCTURALES (III): COVARIANZA Y CORRELACIÓN ENTRE VARIABLES LATENTES

(#353). BÁSICOS DE ECUACIONES ESTRUCTURALES (II): VARIABLES LATENTES Y FIABILIDAD

(#352). TESTOSTERONA Y CONSUMO CONSPICUO EN HOMBRES

(#350). BÁSICOS DE ECUACIONES ESTRUCTURALES (I): COVARIANZAS Y DESVIACIONES SOBRE LA MEDIA

(#346). MODELO NO LINEAL EN LAS VARIABLES PERO LINEAL EN LOS PARÁMETROS CON STATA Y MAXIMA

(#337). EXPLORANDO LA PREDICCIÓN LINEAL EN MARKETING CON MAXIMA

(#336) EXPLORANDO EL DETERMINISMO Y EL AZAR CON C Y MAXIMA

(#335). EXPLORANDO EL COMPORTAMIENTO CAÓTICO CON MAXIMA

(#334). EXPLORANDO LA NO LINEALIDAD CON MAXIMA

(#331). LOS CONSTRUCTOS FORMATIVOS NO REFLEJAN ESTADOS PSICOLÓGICOS

(#323). EL ATAQUE CON MICROONDAS A LA EMBAJADA DE EEUU EN MOSCÚ DURANTE LA GUERRA FRÍA

(#314). REDUCIR EL SESGO EN ESTUDIOS SOBRE EFECTIVIDAD DE VACUNAS

(#313). EL SISTEMA PLANETARIO DE TRAPPIST-1

(#310). MÁS ALLÁ DEL REDUCCIONISMO EN BIOLOGÍA

(#307). LA CRISIS DE LA INVESTIGACIÓN EN PSICOLOGÍA

(#302). CONSTRUCTOS EN PSICOLOGÍA E IMPLICACIONES ESTADÍSTICAS

(#301). CENSURA EN EL LENGUAJE CIENTÍFICO

(#298). ¿CAMBIAR DE ENTRENADOR A MITAD DE TEMPORADA?

(#294). NEUROLOGÍA DE LA MALDAD; PISTAS SOBRE LAS RAÍCES DE LA MALDAD INDIVIDUAL

(#287). MEDIAS VERDADES Y OCULTACIÓN; ACEPTACIÓN DEL ENGAÑO DE CORPORACIONES

(#283). INTERPRETACIÓN FRECUENTISTA DE LA PROBABILIDAD; EL PAPEL DE LOS MODELOS

(#276). LA IRRACIONALIDAD, SCHRÖDINGER, Y EL EFECTO TÚNEL

(#265). DEFENDIENDO EL P-VALOR A TRAVÉS DE SU CORRECTA INTERPRETACIÓN

(#264). MÁS ALLÁ DE LOS ENSAYOS CONTROLADOS ALEATORIZADOS

(#263). CAUSALIDAD EN CASOS INDIVIDUALES; ALGUNAS RECOMENDACIONES PARA SU ESTUDIO

(#260). PROPONEN SER 10 VECES MÁS EXIGENTES CON EL P-VALOR

(#257). ERRORES DE CODIFICACIÓN DE EMBOLIAS PULMONARES AL USAR EL ICD-10

(#252). LA PREGABALINA NO SE ASOCIA A MALFORMACIONES EN EL EMBARAZO

(#243). LOS DEPORTISTAS DE ÉLITE VIVEN MÁS, PERO CON MATICES

(#232). LA OTRA HISTORIA DE NIKE

(#191). LA TRIADA OSCURA, LOS ANTICONCEPTIVOS Y LA PREFERENCIA SEXUAL

(#185). VENTA DE ENTRADAS EN LA NBA Y POTENCIAL DE LOS EQUIPOS

(#163). CRECEN LAS DUDAS SOBRE PLS (PARTIAL LEAST SQUARES)

(#159). ENTENDER LA NATURALEZA DE LOS DATOS Y DE LOS TEST ESTADÍSTICOS

(#128). INDICADORES FORMATIVOS EN ECUACIONES ESTRUCTURALES

(#127). AJUSTE EXACTO EN ECUACIONES ESTRUCTURALES

(#104). CORRUPCIÓN Y DESIGUALDAD ENTRE LOS MUY RICOS

(#100). DESCREÍDOS DE LA EPIDEMIOLOGÍA

(#90). REDES EN PSICOPATOLOGÍA Y EN MARKETING

(#76). EL 85-90% DE LOS NUEVOS MEDICAMENTOS SON INNECESARIOS

(#33). CLUSTERS DE CÁNCER PEDIÁTRICO

(#31). TRÁFICO Y CÁNCER DE PULMÓN

(#26). FOTOVOLTAICAS HIPERGEOMÉTRICAS

(#24). ROMPER EL CRISTAL

(#22). COMPLEJIDAD DE LOS FENÓMENOS SOCIALES

(#20). DECISIONES HEURÍSTICAS

(#17). RADICALES DE LAS VACUNAS

(#13). 24755 CACAS EN EL PARQUE

(#12). ESTADÍSTICAS SIN ESTADÍSTICA

(#10). EL DESCUBRIMIENTO DEL ARGÓN

(#6). ¿VIVIMOS MÁS Y CON MEJOR SALUD?

(#5). ¿ES LA ECONOMÍA UNA CIENCIA?

(#3). HEZONJA Y LAS PREDICCIONES ESTADÍSTICAS

(#1) ¿CÓMO JUGARÁ MARIO HEZONJA SU PRÓXIMO PARTIDO?