

OPTIMIZACION CON SOBRECALENTAMIENTO SIMULADO EN REGRESION NO LINEAL: ALGORITMO Y SOFTWARE¹

Javier Trejos Zelaya y Mario Alberto Villalobos Arias
CIMPA, Escuela de Matemática, Universidad de Costa Rica

RESUMEN

Se presenta un algoritmo que usa la técnica de de sobrecalentamiento simulado, para abordar el problema de optimización global en regresión no lineal. Se exponen ejemplos los resultados obtenidos sobre algunos ejemplos al aplicar este algoritmo a datos tomados de la literatura, se presenta también un Software computacional, para Windows 95 o superior, de fácil uso, en el que se implemento el algoritmo, además de los modelos lineales clásicos de Regresión Lineal.

Palabras clave: regresión; regresión no lineal; optimización global; sobrecalentamiento simulado; mallado; software; comparaciones.

ABSTRACT

We present an algorithm that uses the technique of simulated annealing, for the global optimization problem in non-linear regression. We show the results obtained over some examples, which are comparacle to the results obtained with classical mehotds. We have also implemented the new method in a portable software, the runs on Windows 95 or higher, very easy to use. This software contains the simulated annealing for non-linear regression as well as the classical linear regression models.

Key words: regression; non-linear regression; global optimization; simulated annealing; grid; software; comparisons.

1. LA REGRESION

En regresión se tiene las observaciones $\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_p \in \mathbb{R}^n$ y se quiere explicar \bar{y} mediante los $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$. Esta dependencia será explicada, en el caso general, por medio de una función f que depende de $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$ y que además tiene como parámetros, por determinar, $\theta_1, \theta_2, \dots, \theta_r$. Esto equivale a buscar $\hat{y} = f(X, \bar{\theta})$ que aproxime a \bar{Y} , con $X = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$ y $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_r)$.

Para medir la calidad de la aproximación de \bar{Y} por \hat{y} , en general se utiliza el criterio de mínimos cuadrados, a saber:

$$\| \bar{y} - \hat{y} \| = \min - \| \bar{y} - f(X, \bar{\theta}) \|$$

donde la norma se debe especificar. Es usual emplear la norma euclídea clásica (asociada al producto interno usual), aunque hay criterios más generales como los basados en las distancias de Minkowski. Es importante ver que en este método no es necesario que la función a optimizar sea diferenciable.

Para el caso de la Regresión Lineal Múltiple se tiene que $r = p + 1$ y

$$f(X, \bar{\theta}) = \theta_0 \bar{1} + \theta_1 \bar{x}_1 + \theta_2 \bar{x}_2 + \dots + \theta_p \bar{x}_p$$

donde $\bar{1} = (1, 1, \dots, 1)^t \in \mathbb{R}^n$, con lo que se quiere minimizar:

$$S(\bar{\theta}) = \sum_{i=1}^n [y_i - (\theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip})]^2$$

en el caso de utilizar la norma euclídea clásica. La solución de este problema es ampliamente conocida (ver, por ejemplo Albert (1972) o Tomassone **et al.** (1992)).

¹ Investigación parcialmente financiada por la Universidad de Costa Rica, proyecto 821-97-333.
Dirección electrónica de los autores: jtrejos@cariari.ucr.ac.cr y mvillalo@cariari.ucr.ac.cr.

2. REGRESION NO LINEAL USANDO SOBRECALENTAMIENTO SIMULADO

2.1. Sobrecalentamiento simulado

Con el nombre de Sobrecalentamiento Simulado (SS) se conoce una técnica iterativa de optimización combinatoria, basada en la definición de vecindarios y tal que el estado que se visita en cada interacción puede ser mejorado, en cuyo caso se cambia a ese estado, o bien puede ser peor que el anterior, de tal forma que el cambio de estado puede ser aceptado de acuerdo con una regla probabilística conocida como Regla de Metropolis. Esta regla es similar a una caminata aleatoria, con la diferencia de que la probabilidad de aceptación está controlada por un parámetro externo llamado temperatura y que denotamos con c_k . Al principio de las iteraciones c_k es grande, por lo que prácticamente todos los estados que empeoran el criterio son aceptados, mientras que al final de las iteraciones c_k es pequeño, por lo que al final es similar a un descenso del gradiente. El parámetro c_k debe ser descendido suavemente con el fin de simular el proceso físico de sobrecalentamiento, sobre el cual se basa el SS.

El SS tiene la particularidad de que se ha demostrado su convergencia asintótica hacia el óptimo global del problema de optimización combinatoria (ver por ejemplo Aarts & Korst (1990) o Laarhoven & Aarts (1988)). Esta demostración está basada en una modelación con cadenas de Markov para los estados del problema.

Una implementación adecuada del SS debe cumplir las condiciones de convergencia asintótica, esto es: debe haber reversibilidad entre los estados (si se pasa de un estado a a un estado b con probabilidad p , entonces se debe ser capaz de regresar de b a a con la misma probabilidad), conexidad (entre cualesquiera dos estados, debe haber una serie de estados intermedios de forma que se vaya de uno al siguiente con probabilidad estrictamente positiva), y además todos los vecindarios deben tener el mismo tamaño y todos los vecinos de un estado deben tener la misma probabilidad de acceso.

Para implementar el SS, se debe definir un programa de enfriamiento, esto es: (i) definir el parámetro de control inicial c_0 , (ii) definir la forma de disminuir c_k , (iii) definir el parámetro de control final, y (iv) definir el tamaño de cada iteración, asociada a la cadena de Markov que modela esa iteración y denotado L_k .

El SS ha dado excelentes resultados en diversos métodos de Análisis de Datos, como en particionamiento, Trejos, Murillo & Piza (1998); particionamiento bimodal, Castillo (1999) y Castillo & Trejos (1999); análisis de proximidades (conocido como escalamiento multidimensional o *multidimensional scaling*), Trejos & Villalobos (1999b) y Villalobos (1998); y rotaciones varimax oblicuas, Trejos (1992).

2.2. Método propuesto de regresión no lineal

Para poder aplicar el SS, se debe definir qué es un estado, la función a optimizar, qué es un vecino y cómo se genera un nuevo estado, así como el programa de enfriamiento.

El método propuesto de regresión no lineal usando sobrecalentamiento simulado está basado en la construcción de un mallado del espacio de parámetros. Un estado, denotado por I , será una r -tupla de la forma:

$$I = (\theta_1^I, \theta_2^I, \dots, \theta_r^I)$$

donde $\forall i \in \{1, 2, \dots, r\}$, $\theta_i^I \in \mathbb{R}$. Dos estados I, J son vecinos si existen $l \in \{1, 2, \dots, r\}$ y $\alpha \in \{-1, 1\}$ tales que:

$$\theta_i^J = \begin{cases} \theta_i^I & \text{si } i \neq l \\ \theta_i^I + h_l \alpha & \text{si } i = l \end{cases}$$

con $h_l \in \mathbb{R}$. Así, el espacio de los parámetros se discretiza mediante una malla de lado h_l . Consecuentemente, para generar un nuevo estado se utiliza el siguiente procedimiento, que cumple las condiciones de reversibilidad y conexidad:

Generar (J, I) (generar el estado J a partir del estado I)

1. Escoger al azar $l \in \{1, \dots, r\}$, el parámetro a perturbar será θ_l^I
2. Escoger al azar $\alpha \in \{-1, 1\}$.
3. $J := I$.
4. $\theta_l^J = \theta_l^I + \alpha h_l$.

Las escogencias al azar se realizan usando una distribución uniforme. El mallado utilizado es ilustrado en la Figura 1, para el caso de dos parámetros.

Para determinar la longitud de la cadena de Markov se toma un factor F_L del número de r , $L_k = F_L r$, y el descenso del parámetro de control se hace por medio de la fórmula $c_{k+1} = \gamma c_k$, con $0,8 < \gamma < 1$. El estado inicial I_0 se escoge de acuerdo con el modelo que se tenga, de manera empírica o por criterio experto.

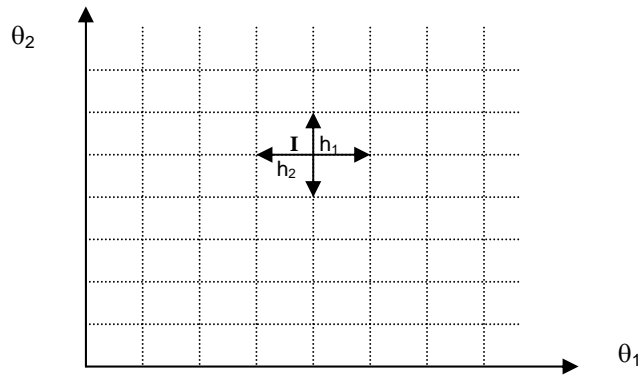


Figura 1. Mallado del espacio de parámetros.

2.3 Algoritmo propuesto

El algoritmo RNL_SS de Regresión No Lineal utilizando Sobrecalentamiento Simulado, en pseudocódigo, es el siguiente:

```

PROCEDURE RNL_SS;
BEGIN
  Inicializar ( $I_0, c_0, \gamma$ );
   $k := 0$ ;  $I := I_0$ ;
   $c_k := c_0$ ;  $L_k := F_L r$ ;
   $S_I := S(I)$ ;
   $opt := I$ ;  $S_{opt} := S_I$ ; {estado óptimo}
  REPEAT
    FOR  $L := 1$  TO  $L_k$  DO BEGIN
      Generar ( $J, I$ );
       $S_J := S(J)$ ;
      IF  $S_J \leq S_I$  THEN BEGIN
         $I := J$ ;  $S_I := S_J$ ;
        IF  $S_I < S_{opt}$  THEN BEGIN
           $opt = I$ ;  $S_{opt} = S_I$ ;
        END;
      END;
      ELSE BEGIN
        IF  $\exp\left(\frac{S_I - S_J}{c_k}\right) > \text{random}[0,1)$  THEN BEGIN
           $I := J$ ;  $S_I := S_J$ ;
        END
      END
    END;
     $k := k+1$ ;
    Calcular_Largo( $L_k$ );
     $c_k := c_k \gamma$ ;
  UNTIL Criterio_de_Parada;
END;
```

El criterio de parada puede darse por una de las siguientes opciones:

- Cuando la temperatura esté baja, es decir, cuando el parámetro de control sea casi 0.
- Cuando el óptimo encontrado, después de cada iteración, se repita un número determinado de veces.
- Después de un número determinado de iteraciones.
- Cuando la tasa de aceptación de los nuevos estados que empeoran el criterio sea casi 0.
- Cuando el usuario desee detenerlo.
- Una combinación de los anteriores.

En el programa presentado en la sección 4 se utiliza el primero de estos criterios, pero el usuario puede detenerlo cuando lo desee.

3. RESULTADOS COMPARATIVOS

Ejemplo 1: Puromycin

Los datos son tomados de Bates & Watts (1988) y se presentan en la Tabla 1. Se aplica el modelo de Michaelis-Menten, que describe la velocidad de una reacción y en función de la concentración del sustrato x :

$$f(x; M, k) = \frac{Mx}{k + x}$$

donde M es la velocidad de la reacción y k es el valor de la concentración para el cual la velocidad es igual a $M/2$, este es llamado la constante de Michaelis. Los valores iniciales son:

$$\theta_0 = M = \text{Max} \{y_i | i \in \{1, \dots, n\}\} \times 0.8$$

$$\theta_1 = k = \text{Min} \{x_i | i \in \{1, \dots, n\}\} \times 0.5.$$

Tabla 1. Datos de Puromycin

x	0.02	0.02	0.06	0.06	0.11	0.11	0.22	0.22	0.56	0.56	1.1	1.1
y	76	47	97	107	123	139	159	152	191	201	207	200

La Figura 2 representa la forma general de este modelo

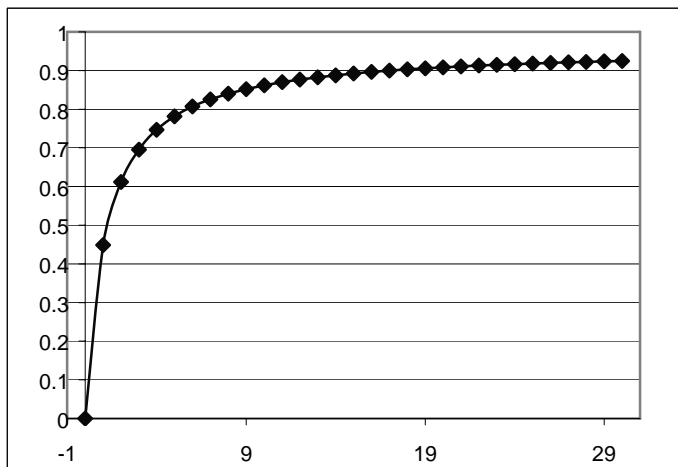


Figura 2. Modelo de crecimiento de Michaelis-Menten.

Al aplicar el método de regresión lineal, mediante linealización del modelo, se obtuvo el valor del criterio de 1920.64. En Bates & Watts (1988) se reporta que el mínimo se obtiene para los valores de los parámetros $\theta_0 = 212.7$ y $\theta_1 = 0.0641$ utilizando el método de Gauss-Newton; al evaluar el criterio de mínimos cuadrados para estos valores se obtiene el valor 1195.4549. Al aplicarle RNL_SS se obtuvo los valores $\theta_0 = 212.6902$ y $\theta_1 = 0.0641402$ que da un valor de criterio de 1195.44821, que es prácticamente el mismo que el obtenido por Bates & Watts, aunque ligeramente mejor. Sin embargo, es muy superior que el que se obtiene por linealización del modelo.

Ejemplo 2: Modelo de Crecimiento de Micherlich

Los datos mostrados en la Tabla 2 son tomados de Antoniadis *et al.* (1992) y se utiliza el modelo de crecimiento de Micherlich, dado por:

$$f(x; M, k) = M + k e^{gx}$$

con $\theta_1 = M$, $\theta_2 = k$ y $\theta_3 = g$.

Tabla 2. Datos para el modelo de crecimiento de Micherlich.

x	1	2	3	4	5	6	7	8	9	10	11	12	13
y	3.183	3.059	2.871	2.622	2.541	2.184	2.11	2.075	2.018	1.903	1.77	1.762	1.55

En este ejemplo, Antoniadis **et al.** (1992) reportan el valor del criterio igual a 0.0534535 para los valores de los parámetros $\theta_0 = 0.96312$, $\theta_1 = 2.519$ y $\theta_2 = -0.10305$. Con RNL_SS se obtuvieron resultados similares, con $\theta_0 = 0.95622$, $\theta_1 = 2.52619$ y $\theta_2 = -0.102529$, para un valor del criterio de 0.053465.

Ejemplo 3: Velocidad de Reacción

Los datos son utilizados por Tomassone **et al.** (1992) y se presentan en la Tabla 3. Aquí se utiliza el modelo de Michaelis-Menten.

Tabla 3. Datos de Tomassone **et al** (1992) para la velocidad de reacción

x	0.3	0.5	1	2	3	4	5	10
y	0.17	0.27	0.43	0.65	0.73	0.78	0.79	0.81

En Tomassone **et al.** (1992) se reportan los valores de los parámetros mostrados en la Tabla 4, con sus respectivos valores del criterio de mínimos cuadrados.

Tabla 4. Valores reportados por Tomassone **et al.** (1992) para los datos de cinética enzimática.

Método	M	k	Valor del Criterio
Gauss-Newton	0.96	1.14	0.0082556
Lineweaver-Burk	0.92	1.58	0.0660704
Woolf	0.92	1.00	0.0101543
Eadie	1.01	1.37	0.0109522
Semi-paramétrico	1.00	1.34	0.0104194

Con RNL_SS se obtuvieron los siguientes valores: $M = 0.963902$, $k = 1.13899$, para un valor del criterio de 0.0082017, mejorándose el mejor resultado de Tomassone **et al.** (1992) en 1 %. Por otro lado al linealizar el modelo y aplicando el método de regresión lineal usual, solo se tiene un valor del criterio de 0.020194, que se aleja mucho del valor que se obtuvo con RNL_SS.

Otros modelos

En Trejos & Villalobos (1999a) se presentan los resultados obtenidos con RNL_SS aplicados a modelos de crecimiento de cultivos, en particular el crecimiento del Ñame alado. Para los distintos tejidos analizados (tubérculos, tallos, hojas, hojas secas) se obtuvieron resultados comparables a los obtenidos con el método de Gauss-Newton, usando modelos de crecimiento logístico y tipo campana. Ambos modelos se encuentran en el software que se presenta aquí.

Recuérdese que el modelo de crecimiento logístico está dado por la función:

$$f(x; M, \kappa, g) = \frac{M}{1 + \exp(-\kappa(x - g))}$$

donde M es el valor máximo asintótico, κ es el control de la velocidad de crecimiento (curvatura) y g es el punto de inflexión, cuando el crecimiento pasa de ser acelerado a desacelerado. El gráfico correspondiente se presenta en la Figura 3. Se ve que en este caso se tienen 3 parámetros: $\theta_0 = M$, $\theta_1 = \kappa$, $\theta_2 = g$. En este caso los valores iniciales son:

$$\theta_0 = M = \text{Max} \{y_i / i \in \{1, \dots, n\}\}$$

$$\theta_1 = \kappa = 1,$$

$$\theta_2 = g = (\text{Max} \{x_i\} + \text{Min} \{x_i\})/2, i \in \{1, \dots, n\}.$$

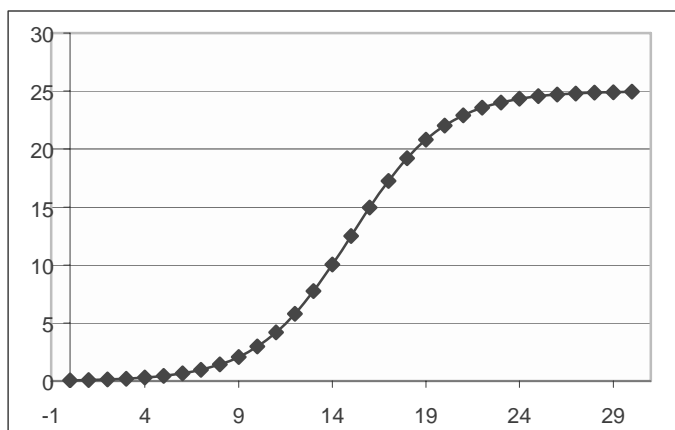


Figura 3. Modelo de crecimiento logístico

Por su parte, el modelo de crecimiento tipo campana (llamado así por ser el gráfico de una campana de Gauss truncada) describe el crecimiento de un tejido que llega a un máximo y luego desciende un poco en su etapa final:

$$f(x; M, \kappa, g) = M e^{-\kappa(x-g)^2}$$

donde M es el valor máximo, κ es el control de la velocidad de crecimiento (curvatura) y g es el punto donde se alcanza el máximo. El gráfico correspondiente se muestra en la Figura 4. Para este caso los valores iniciales son:

$$\theta_1 = M = \text{Max} \{y_i / i \in \{1, \dots, n\}\}$$

$$\theta_2 = \kappa = 1,$$

$$\theta_3 = g = 2/3 \text{Max} \{x_i\} + 1/3 \text{Min} \{x_i\}, i \in \{1, \dots, n\}.$$

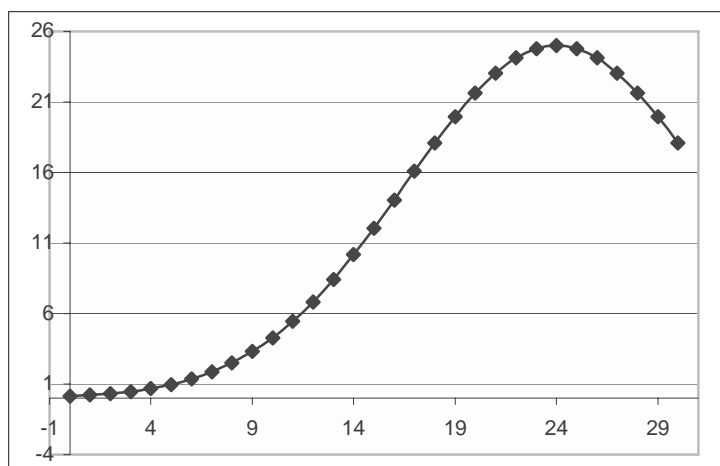


Figura 4. Modelo de crecimiento tipo campana truncada.

4. SOFTWARE DE REGRESION LINEAL Y NO LINEAL

Con el algoritmo que se presenta en este trabajo se está desarrollando un Software de fácil manejo para el usuario. Se presenta en este trabajo la primera versión del mismo. Está diseñado en Visual Basic 6.0 para trabajar bajo la plataforma de Windows 95 o superior. Se espera que este software forme parte del Paquete Integrado de Métodos de Análisis de Datos PIMAD, desarrollado en la Universidad de Costa Rica.

Entre las principales ventajas que tiene se encuentran las siguientes:

- Tiene implementada una hoja electrónica para editar los datos o importarlos, incluyendo formatos como Formula 1, Excel, Excel 4.0, texto o texto con tabulaciones. Esta hoja electrónica permite todas las manipulaciones de copia, corte y pegue de las aplicaciones Windows. Así mismo, permite realizar operaciones mediante fórmulas y la aplicación de funciones.

- Permite que el usuario seleccione los datos marcándolos en la hoja electrónica, lo cual da mucha versatilidad.
- Permite realizar regresiones lineales y los casos no lineales que se convierten en lineales mediante una transformación algebraica: polinomial (del tipo x^n), exponencial, logarítmico, asintótico y logístico.
- Permite realizar una regresión no lineal usando sobrecalentamiento simulado. Es de destacar que el usuario puede ver gráficamente la evolución de la curva solución.

Como se puede apreciar en la Figura 5, el software puede ser usado mediante botones o mediante menú, con las facilidades de una hoja de cálculo tipo Excel.

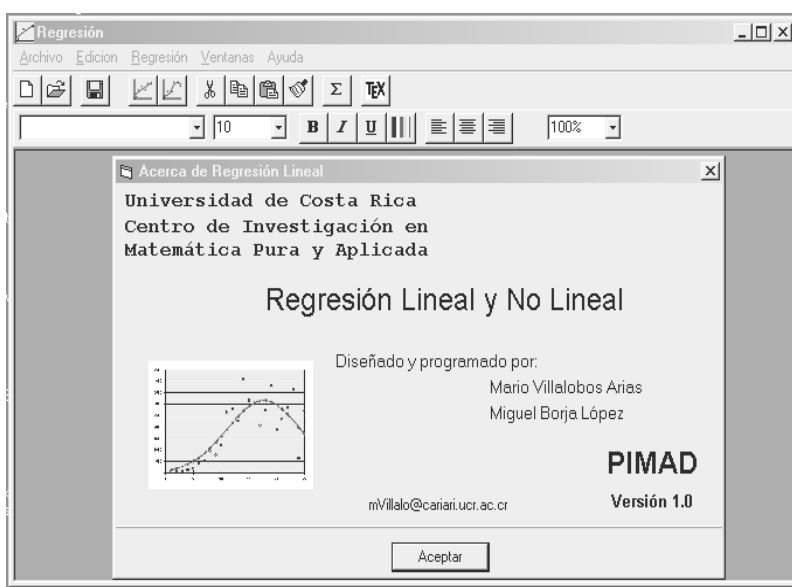







Figura 5. Pantalla inicial del Software Regresión Lineal y No Lineal.

Los tres primeros botones  se usan para manejar los archivos de datos: crear un nuevo archivo, abrir uno existente y para guardar un archivo. Los dos botones que siguen  sirven para escoger entre regresión lineal y regresión no lineal. Los cuatro botones que siguen  son para la edición de los datos (cortar, copiar, pegar y copiar formato, respectivamente). El botón siguiente sirve para hacer sumas de filas o columnas. El botón  genera un archivo en formato LaTeX de la gráfica que se tenga seleccionada.

La siguiente barra es para el aspecto de los datos en la ventana, como son: tipo de letra, tamaño, negrita, itálica, subrayado, color y alineamiento del texto y porcentaje de vista.



En la Figura 6 se ilustra la forma de seleccionar los datos al oprimir el botón de regresión lineal . Puede verse que los datos son marcados directamente con el ratón. En otro ventana semejante se piden los datos que corresponderán a los valores de y .

Una vez que se han seleccionado los datos se procede a seleccionar el modelo que se quiere ejecutar, lo cual se puede apreciar en la Figura 7.

Luego de esto se oprime terminar. Si se desea se puede devolver para cambiar las escogencias oprimiendo el boton correspondiente. Inmediatamente el programa despliega tres ventanas en las que aparecen el grafico de la aproximación, el gráfico los errores estandarizados y en una hoja de cálculo los datos resultantes de la regresión. En caso de que se escogiera la logística el grafico se vería como el de la Figura 8.

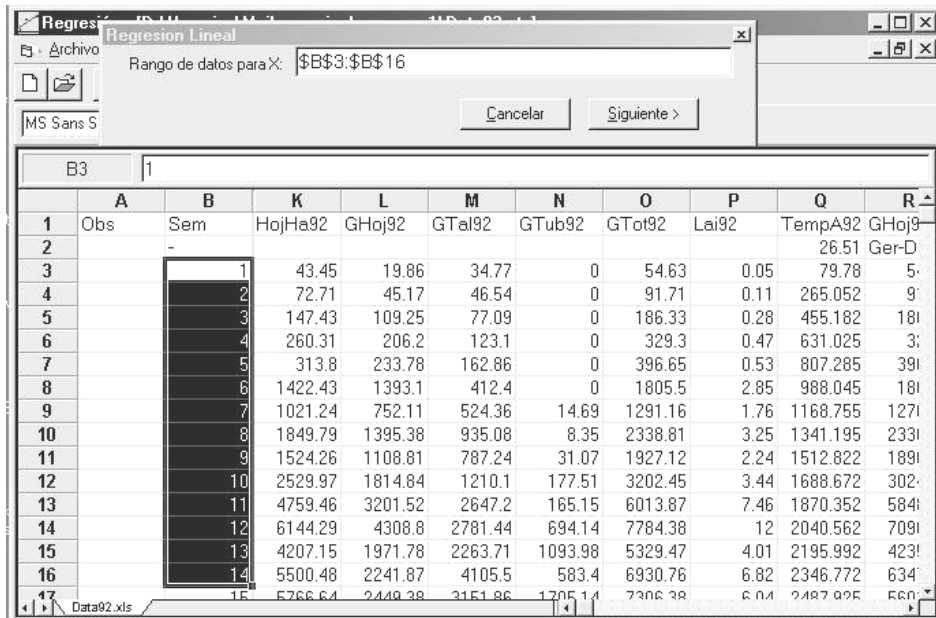


Figura 6. Ilustración de la operación de marcar en la hoja electrónica para la escogencia de los datos.

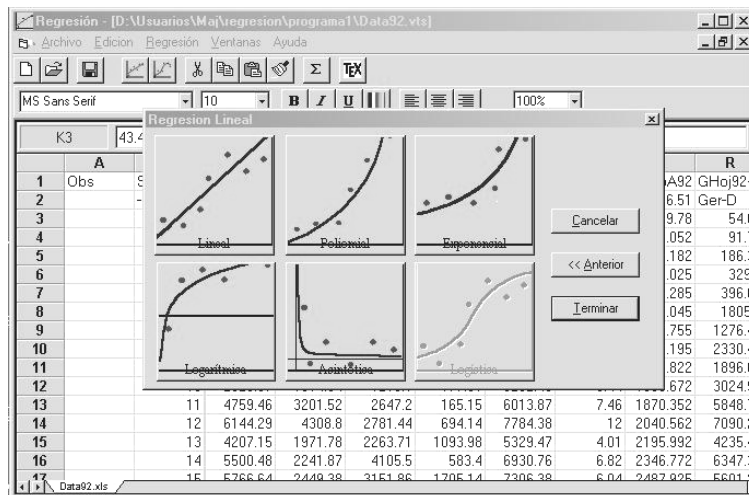


Figura 7. Selección del modelo de regresión lineal o del modelo linealizado (se ilustra la escogencia del modelo logístico).

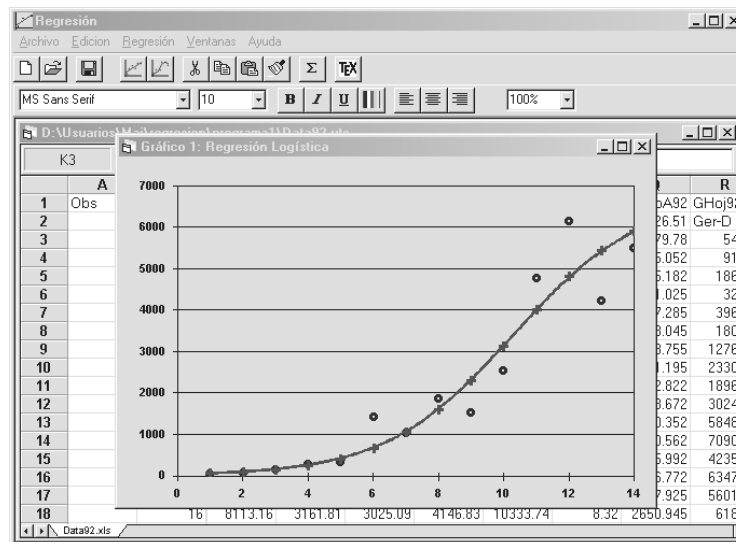


Figura 8. Resultado de una regresión logística usando linealización.

Las ventanas se pueden ver en forma de mosaico como la Figura 9, en la que se muestran la ventana de los datos, los gráficos de solución y de error, y una ventana con la solución numérica. Las ventanas de gráfico se pueden salvar en el formato gráfico que se desee.

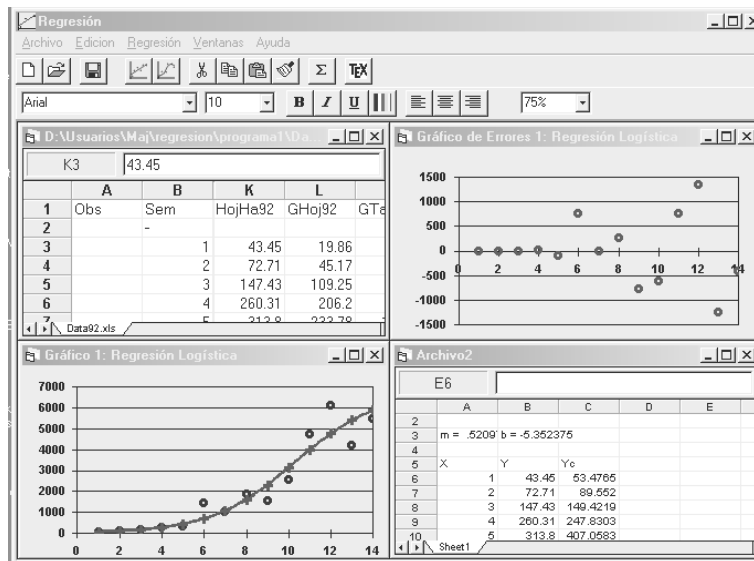




Figura 9. Presentación de las ventanas en forma de mosaico.

En el caso en que se escoja la regresión no lineal  se escogen los datos tal como se hace con la regresión lineal. Entre los modelos que estarán disponibles, se encuentran los aquí expuestos y muchos de los que aparecen en Bates & Watts (1988). La próxima versión del software contará con la opción de escoger el modelo deseado.

Después de escoger los datos se presenta una ventana con el gráfico donde se muestra como la aproximación se va logrando; el botón  permite detener la ejecución de la regresión. Aquí también se generan las tres ventanas como en el caso lineal. La figura 10 muestra el resultado obtenido para los datos de velocidad de reacción.

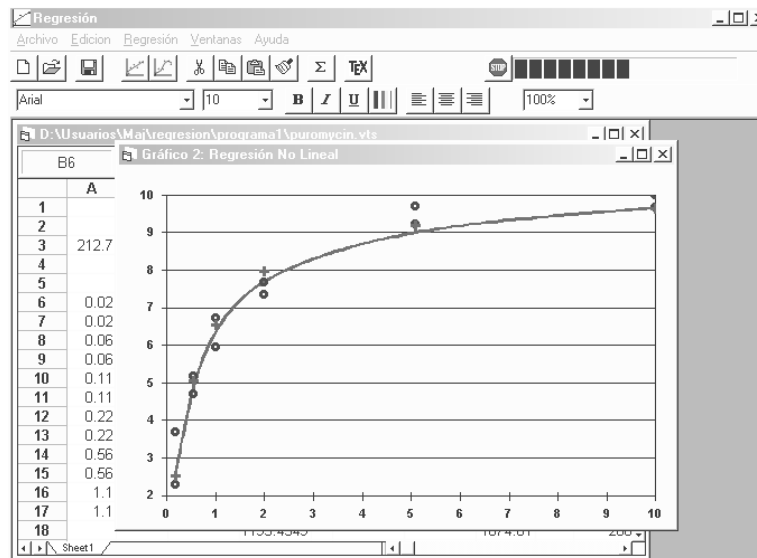


Figura 10. Ventana que muestra el resultado de la aproximación para el modelo de Michaelis-Merten.

REFERENCIAS

1. AARTS, E. and J. KORST (1990): "Simulated Annealing and Boltzmann Machines. A Stochastic Approach to Combinatorial Optimization and Neural Computing", John Wiley and Sons, Chichester.

2. ALBERT, A. (1972): "Regression and the Moore-Penrose Pseudoinverse", Academic Press, New York.
3. ANTONIADIS, A. **et al.** (1992): "Regresion Non Linéaire et Applications", Economica, Paris.
4. BATTES, D. and D. WATTS (1988): "Nonlinear Regression Analysis and its Aplications", John Wiley & Sons, New York.
5. CASTILLO, W. (1999) Métodos de Particionamiento Bimodal y Trimodal, Tesis de Maestría (en preparación), Universidad de Costa Rica, San Pedro.
6. CASTILLO, W. y J. TREJOS (1999): "Nuevos resultados en clasificación bimodal jerárquica y por particiones", Preprint, CIMPA, Universidad de Costa Rica.
7. LAARHOVEN, P.J.M. van and E.M. AARTS (1988): "Simulated Annealing: Theory and Applications", Kluwer Academic Publ., Dordrecht.
8. PIZA, E. (1998): "Clasificación automática: particionamiento mediante sobrecalentamiento simulado", **Investigación Operacional** 19(2-3):152-163.
9. TREJOS, J. (1993): "A simulated annealing implementation for oblique varimax rotations", in **Applied Stochastic Models and Data Analysis**, Vol. II, J. Janssen & C.H. Skiadas (Eds.), World Scientific, Singapur: 981-989.
10. _____ (1999): "Stochastic optimization applied in data analysis", in **Applied Stochastic Models and Data Analysis**, H. Bacelar-Nicolau, F.C. Nicolau & J. Janssen (Eds.), Instituto Nacional de Estadística, Lisboa: 305-310.
11. TREJOS, J.; A. MURILLO and E. PIZA (1998): "Global stochastic optimization for partitioning", in **Advances in Data Science and Classification**, A. Rizzi **et al.** (Eds.), Springer-Verlag, Berlin: 185-190.
12. TOMASSONE, R. **et al.** (1992): "La Régression: Nouveaux Regards sur une Ancienne Méthode Statistique", Masson, Paris.
13. VILLALOBOS, M. (1998): Optimización Estocástica para el Análisis de Proximidades. Tesis de Maestría, Universidad de Costa Rica, San Pedro.
14. TREJOS, J. and M. VILLALOBOS (1999a) "Optimización mediante recocido simulado en regresión no lineal", Memoria del *XIII Foro Nacional de Estadística*, Monterrey, México.
15. _____ (1999b) "Une implémentation du recuit simulé en analyse des proximités", *XXXI Journées de Statistique*, Grenoble: 111-114.