

APLICACIÓN DE LA BÚSQUEDA TABÚ EN REGRESIÓN NO LINEAL

MARIO A. VILLALOBOS* JAVIER TREJOS† SERGIO DE LOS COBOS‡

Recibido/Received: 30 Jul 2005 — Aceptado/Accepted: 27 Abr 2006

Resumen

Se presenta una implementación de la técnica de optimización combinatoria conocida como búsqueda tabú a la estimación de parámetros de un modelo de regresión no lineal dado. La implementación está basada en una discretización del espacio de parámetros y en el recorrido de este espacio a través de una malla, para la generación de los estados vecinos de un estado actual. Se presentan algunos resultados comparativos sobre datos reales o simulados.

Palabras clave: regresión, estimación de parámetros, optimización combinatoria, mínimos locales.

Abstract

We implement the combinatorial optimization technique known as tabu search in the parameter estimation problem in a given non linear model. For the generation of neighbors, the implementation is based on a discretization of the parameter space, which is covered by a mesh. We present some comparative results on real or simulated data.

Keywords: regression, parameter estimation, combinatorial optimization, local minima.

Mathematics Subject Classification: 62J02, 90C27.

*CIMPA–Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. E-Mail: mvallalo@cariari.ucr.ac.cr

†CIMPA–Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica. E-Mail: jtrejos@cariari.ucr.ac.cr

‡Departamento de Ingeniería Eléctrica, Universidad Autónoma Metropolitana – Iztapalapa, Av. Michoacán y La Purísima s/n, Col. Vicentina, México D.F. CP 09340, México. E-Mail: cobos@xanum.uam.mx

1 Introducción

En regresión no lineal se encuentra el problema, tan frecuente en distintos métodos estadísticos, de la obtención de mínimos locales [8, 16]. En efecto, el método clásico de regresión no lineal es el conocido como de Gauss–Newton, el cual está basado en una aproximación de la función de ajuste mediante el uso de un polinomio de Taylor de primer orden, y en un procedimiento de búsqueda de un óptimo local [4, 8]. Por ello, es razonable pensar en la implementación de técnicas de optimización global. Entre estas técnicas, podemos citar el recocido simulado, la búsqueda tabú y los algoritmos genéticos [7], las cuales han sido ampliamente usadas en distintos problemas estadísticos, de investigación operacional o de ingeniería.

En particular, en Análisis Estadístico de Datos han sido usados en particionamiento para el análisis de conglomerados con resultados muy superiores a los de métodos clásicos, como el de las k -medias y el de Ward, en clasificación bimodal, en escalamiento multidimensional, en rotaciones varimax oblicuas y en la obtención de conjuntos aproximados (conocidos como *rough sets* en inglés). Todas estas implementaciones de las técnicas de optimización mencionadas han mejorado o al menos igualado los resultados conocidos, y los detalles pueden consultarse en [7]. En regresión no lineal ya se ha aplicado la técnica de recocido simulado (ver [18]) con resultados satisfactorios.

En el presente trabajo hacemos una presentación de la regresión no lineal como problema de optimización, y describimos algunos de los métodos más conocidos para llevarla a cabo. Enseguida se presenta la implementación efectuada de la búsqueda tabú para este problema. Finalmente, damos algunas conclusiones y perspectivas a este trabajo. Conviene señalar que nos restringimos al caso de la regresión con una variable explicativa, pudiendo generalizarse fácilmente el trabajo al caso de varias variables explicativas; además, no abordamos los problemas relativos a la estimación estadística, como la obtención de intervalos de confianza para los parámetros.

2 Regresión no lineal

Dadas dos variables \mathbf{x} y \mathbf{y} observadas sobre n objetos, donde \mathbf{x} es una variable explicativa y \mathbf{y} es una variable a explicar que depende de \mathbf{x} , se quiere describir la relación de dependencia de \mathbf{y} respecto a \mathbf{x} mediante una función f ; es decir, se quiere establecer la relación funcional $\mathbf{y} = f(\mathbf{x}) + \epsilon$, donde ϵ es un término de error¹. La función f depende generalmente de ciertos parámetros, cuyo vector denotaremos $\vec{\theta}$, por lo que escribiremos a la función de regresión $f_{\vec{\theta}}$. Se quiere utilizará el criterio de mínimos cuadrados, el cual mide la calidad de la aproximación funcional propuesta:

$$S(\vec{\theta}) = \|\mathbf{y} - f(\mathbf{x})\|^2 = \sum_{i=1}^n [y_i - f_{\vec{\theta}}(x_i)]^2 \quad (1)$$

donde $\mathbf{x} = (x_1, \dots, x_n)^t$ y $\mathbf{y} = (y_1, \dots, y_n)^t$ son los vectores de las observaciones de las variables, y $\|\cdot\|$ es la norma Euclídea usual. Hacemos notar que para el trabajo

¹En este trabajo no supondremos que ϵ sigue alguna distribución en particular.

con las heurísticas de optimización combinatoria (recocido simulado, búsqueda tabú) que usaremos, podemos usar otros criterios, como el de la norma L_1 : $S_1(\vec{\theta}) = \|\mathbf{y} - f(\mathbf{x})\|_1 = \sum_i |y_i - f_{\vec{\theta}}(x_i)|$. Esta extensión no presenta ninguna dificultad ya que nuestros métodos no necesitan ningún tipo de diferenciabilidad del criterio. Resultados comparativos usando estos otros criterios serán presentados en otro trabajo. Notemos también que en el criterio (1) se pueden hacer ponderaciones tanto de los objetos como de las variables. Sin embargo, para no hacer pesada la presentación, nos restringiremos aquí al caso en que los objetos tienen el mismo peso y las variables también.

Finalmente, mencionemos otros dos tipos de regresión no lineal: por maximización de la función de verosimilitud (o minimización de la log-verosimilitud), estudiada ampliamente en [15], y el uso de las ideas de la estadística bayesiana [3], haciendo una estimación del riesgo mínimo.

Salvo cuando $f_{\vec{\theta}}$ es una función lineal, no se conoce una solución general a este problema. Debe notarse que en algunos casos la función $f_{\vec{\theta}}$ no es en sí lineal pero el problema de regresión se puede *linealizar*. Por ejemplo, el modelo $\mathbf{y} = \theta_1 e^{\theta_2 \mathbf{x} + \epsilon}$ es linealizable aplicando logaritmo natural, obteniéndose $\mathbf{y}' = \theta_3 + \theta_2 \mathbf{x} + \epsilon$, donde $\mathbf{y}' = \ln y$ y $\theta_3 = \ln \theta_1$. Sin embargo, el modelo $\mathbf{y} = \theta_1 e^{\theta_2 \mathbf{x}} + \epsilon$ no es linealizable mediante la función logaritmo, ya que el error ϵ es aditivo. En este trabajo no nos ocuparemos de los modelos que son linealizables, ya que éstos se resuelven fácilmente mediante la regresión lineal clásica.

Algunos ejemplos de modelos que no son linealizables, son los siguientes:

- crecimiento logístico: $\mathbf{y} = \frac{\theta_1}{1 + \exp[-\theta_2(\mathbf{x} - \theta_3)]} + \epsilon$
- crecimiento con decaimiento: $\mathbf{y} = \theta_1 \exp[-\theta_2(\mathbf{x} - \theta_3)^2] + \epsilon$
- [8]: $\mathbf{y} = \frac{\theta_1}{\theta_1 - \theta_2} [e^{-\theta_2 \mathbf{x}} - e^{-\theta_1 \mathbf{x}}] + \epsilon$

2.1 Método de Gauss–Newton

El método más usado en regresión no lineal es el de Gauss-Newton, que se basa en una aproximación lineal de la función $f_{\vec{\theta}}$ cuando ésta es derivable. La aproximación de $f_{\vec{\theta}}(\mathbf{x})$ se basa en el polinomio de Taylor de primer orden alrededor del punto $\vec{\theta}^0 \in \mathbb{R}$:

$$f_{\vec{\theta}}(x_i) = f_{\vec{\theta}^0}(x_i) + \sum_{j=1}^p \left[\frac{\partial f_{\vec{\theta}}(x_i)}{\partial \theta_j} \right]_{\vec{\theta}=\vec{\theta}^0} (\theta_j - \theta_j^0). \quad (2)$$

Poniendo $f_i^0 = f_{\vec{\theta}^0}(x_i)$, $\beta_j^0 = \theta_j - \theta_j^0$ y $z_j = \left[\frac{\partial f_{\vec{\theta}}(x_i)}{\partial \theta_j} \right]_{\vec{\theta}=\vec{\theta}^0}$, entonces se puede ver que el modelo no lineal se aproxima por uno lineal de la forma:

$$\mathbf{y} - f^0 = \sum_{j=1}^p \beta_j^0 z_j^0 + \epsilon,$$

por lo que se puede obtener una estimación \mathbf{b}^0 de $\vec{\beta}^0$ usando regresión lineal múltiple: $\mathbf{b}^0 = (\mathbf{Z}^{0'}\mathbf{Z}^0)^{-1}\mathbf{Z}^{0'}(\mathbf{y}^0)$ donde \mathbf{Z}^0 y \mathbf{y}^0 son las matrices que contienen como componentes z_j y $y_i - f_i^0$, respectivamente.

Se procede usando un valor inicial $\vec{\theta}^0$, y por iteraciones sucesivas se construye una sucesión de valores de los parámetros de la regresión $\vec{\theta}^0, \vec{\theta}^1, \vec{\theta}^2, \dots$, usando $\vec{\theta}^k$ para estimar $\vec{\theta}^{k+1}$ durante la $k + 1$ -ésima iteración en la aproximación (2), hasta que se converge a un valor estable. Claramente esta convergencia² puede ser hacia un óptimo local del criterio $S(\vec{\theta})$ en (1), ya que no se garantiza la convergencia a un óptimo global siendo la búsqueda local en los contornos elipsoidales de los distintos puntos del proceso iterativo. Por ello se ha pensado en el uso de técnicas de optimización combinatoria que traten de evitar esos mínimos locales del criterio, entre las que está el recocido simulado que se describe a continuación y la búsqueda tabú que usamos en el presente trabajo.

2.2 Uso del sobrecalentamiento simulado

El sobrecalentamiento simulado es una técnica iterativa de optimización combinatoria basado en la generación de vecinos de un estado del problema de optimización y en el uso de un parámetro externo de temperatura c_t que controla la probabilidad de aceptación de un nuevo estado: los estados que mejoran el criterio son aceptados siempre, los estados que empeoran el criterio son aceptados con cierta probabilidad que depende de c_t [1], la cual decrece conforme se avanza en las iteraciones. Una característica importante del sobrecalentamiento simulado es que se puede probar que, asintóticamente, converge en probabilidad al óptimo del problema, lo cual lo convierte en una de las mejores técnicas de optimización combinatoria por su relativamente sencilla manera de implementar.

En el caso de la regresión no lineal, se considera el espacio de los parámetros $\vec{\theta}$, siendo un estado el dato de los parámetros. Por ejemplo, en el caso del crecimiento logístico, un estado es el vector tridimensional $\vec{\theta} = (\theta_1, \theta_2, \theta_3)$. Un nuevo estado es generado por el desplazamiento de una de las componentes de $\vec{\theta}$ en alguna de las direcciones paralelas a alguno de los ejes de coordenadas, usando un paso dado h_t (el cual disminuye conforme c_t decrece). Es decir, se está discretizando el espacio de los parámetros mediante un mallado rectangular, y se recorre este mallado de manera paralela a los ejes de coordenadas.

De acuerdo con el algoritmo usual de sobrecalentamiento simulado, un nuevo estado $\vec{\theta}'$ es aceptado a partir del estado $\vec{\theta}$ si $S(\vec{\theta}') < S(\vec{\theta})$, o bien, cuando $\Delta S = S(\vec{\theta}') - S(\vec{\theta}) > 0$ entonces se puede aceptar con probabilidad $\exp[-\Delta S/c_t]$.

En [18] se hace una descripción más amplia del uso del sobrecalentamiento simulado en regresión no lineal y se reportan resultados similares a los obtenidos con el método de Gauss-Newton sobre datos reales de crecimiento de cultivos.

2.3 Otros métodos

Una técnica ampliamente usada en optimización es la llamada de descenso del gradiente. Se trata de un método iterativo que busca la dirección de máximo descenso en cada punto de la iteración. En el caso de la regresión, se debe mover una estimación $\vec{\theta}^k$ de θ en

²En realidad, esta convergencia no está ni siquiera garantizada, ver [8, pp.464–465].

la dirección del vector $\left(-\frac{\partial S(\vec{\theta})}{\partial \theta_1}, \dots, -\frac{\partial S(\vec{\theta})}{\partial \theta_p}\right)$. Debe notarse que si bien teóricamente el método converge, esta convergencia puede ser muy lenta.

Marquardt [11] propuso un método que lleva su nombre y que trata de mejorar los defectos de los métodos de Gauss-Newton y de descenso de gradiente. El método está basado en una interpolación entre las direcciones que escogen esos dos métodos en cada iteración.

3 Regresión no lineal usando búsqueda tabú

3.1 Búsqueda tabú

La Búsqueda Tabú (BT) es una técnica iterativa de optimización combinatoria basada, al igual que el sobrecalentamiento simulado, en el examen de los vecinos de un estado actual [9], [7]. Tiene la particularidad de que hace uso extensivo de la memoria, lo cual no tiene el sobrecalentamiento simulado. De esta forma, trata de evitar los vecindarios donde pueden encontrarse los mínimos locales y también trata de evitar los ciclos.

Las principales características de la BT son:

- El estado actual posee un vecindario \mathcal{V}_i el cual puede ser generado en su totalidad en cada iteración, o bien se puede generar únicamente una muestra de vecinos.
- Se maneja una lista tabú T tal que contiene un número $|T|$ de vecinos que no pueden ser accesados durante un cierto número de iteraciones. Debe notarse que la longitud de esta lista puede ser dinámica.
- Se procede a escoger el vecino $j^* \in \mathcal{V}_i$ tal que su valor de la función objetivo es el mejor entre todos los elementos de \mathcal{V}_i (o aquéllos generados, en caso de usar solamente una muestra), excepto si j^* pertenece a T .
- Se aplica un criterio llamado de *aspiración* para acceder a un estado j^* si es un estado en la lista tabú pero con el mejor valor de la función objetivo que se haya encontrado durante las iteraciones. Este criterio de aspiración tiene aplicación cuando los nuevos estados son generados mediante un tipo de movimiento y la codificación de los estados en T se hace mediante el movimiento inverso, y no usando explícitamente cada estado. Existen otros tipos de criterios de aspiración, como puede consultarse en [7].
- Se puede codificar de alguna forma la región que se ha explorado, con el fin decidir posteriormente explorar regiones del espacio de estados no exploradas aún.
- Se itera durante un número m de veces, dado por el usuario.

Así, la búsqueda tabú usa principalmente dos parámetros: la longitud $|T|$ de la lista tabú y el número máximo de iteraciones m .

La BT tiene múltiples aplicaciones en investigación operacional [7]. En Análisis Multivariado de Datos, también ha sido usada en el problema de clasificación por particiones para el análisis de conglomerados [17], así como en el escalamiento multidimensional [19], con excelentes resultados.

3.2 Detalles de la implementación

Al igual que para el uso del sobrecalentamiento simulado, discretizaremos el espacio de parámetros mediante una malla de ancho h_k , siendo h_k más fina conforme se avance en las iteraciones. Así, un estado es un vector $\vec{\theta}$ del espacio de parámetros. Un vecindario se genera por el desplazamiento en la malla hacia alguno de los estados contiguos, mediante la modificación por h_k en alguno de los parámetros del vector $\vec{\theta}$.

Por ejemplo, en el caso del crecimiento logístico, donde un estado está dado por el vector tridimensional $\vec{\theta} = (\theta_1, \theta_2, \theta_3)$, entonces los vecinos son los seis estados:

$\vec{\theta}^1 = (\theta_1 + h_k, \theta_2, \theta_3)$	$\vec{\theta}^2 = (\theta_1, \theta_2 + h_k, \theta_3)$	$\vec{\theta}^3 = (\theta_1, \theta_2, \theta_3 + h_k)$
$\vec{\theta}^4 = (\theta_1 - h_k, \theta_2, \theta_3)$	$\vec{\theta}^5 = (\theta_1, \theta_2 - h_k, \theta_3)$	$\vec{\theta}^6 = (\theta_1, \theta_2, \theta_3 - h_k)$

Así, si el estado $\vec{\theta}$ es un vector p dimensional, entonces tiene $2p$ vecinos.

El movimiento consiste entonces en la escogencia de uno de los parámetros y en la escogencia de una dirección. De esta forma, si a la l -ésima coordenada de $\vec{\theta}$ se le añade h_k , lo cual podemos denotar como el movimiento $(l, +1)$, entonces codificaremos en la lista tabú el movimiento inverso $(l, -1)$ con el fin de impedir el regreso a la posición anterior de $\vec{\theta}$. Usaremos una lista tabú de tamaño p y emplearemos el criterio de aspiración antes descrito.

4 Resultados comparativos

4.1 Ejemplo didáctico

Considérese el siguiente ejemplo didáctico [2], en el que se quiere ajustar el modelo $y = \beta_1 e^{-\beta_2 x}$ para el conjunto de datos siguiente:

x	-2.5	-1	1	2
y	1	1.1	-1.1	0.2

La figura 1 muestra las curvas de nivel del criterio $S(\vec{\beta})$ para este conjunto de datos, y se puede apreciar la presencia de dos mínimos locales de $S(\vec{\beta})$: uno al frente y otro al fondo de la figura.

Estos mínimos locales son $\vec{\beta}^* = (0.669, 0.214)$, con un valor de $S(\vec{\beta}^*) = 1.968$, y $\vec{\beta}^1 = (-0.764, -0.0298)$ con un valor de $S(\vec{\beta}^1) = 3.436$.

4.2 Datos de Puromycin

Los datos son tomados de Bates & Watts (1988) [4] 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 \mathbf{x} :

$$f_{M,k}(\mathbf{x}) = \frac{M\mathbf{x}}{k + \mathbf{x}}$$

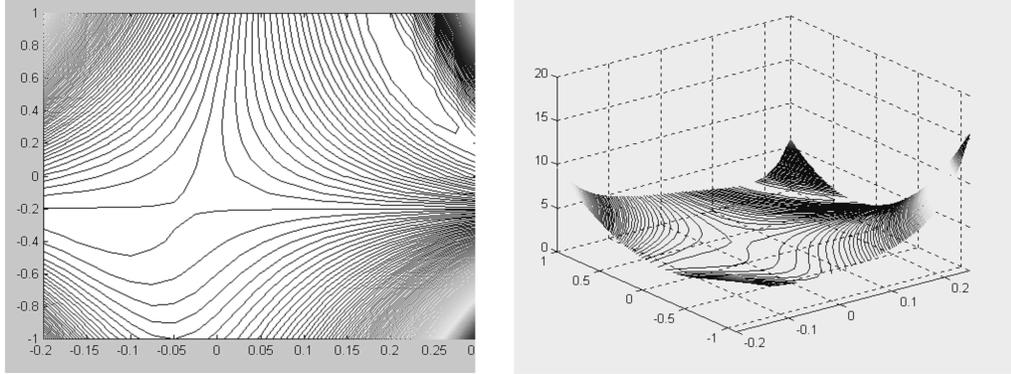


Figura 1: Ilustración de la presencia de mínimos locales.

donde $\beta_1 = M$ es la velocidad de la reacción y $\beta_2 = k$ es el valor de la concentración para el cual la velocidad es igual a $M/2$, éste es llamado la constante de Michaelis. Los valores iniciales son:amsterdam

$$\begin{aligned} b_1 &= M = \max\{y_i/i \in 1, \dots, n\} \times 0.8 \\ b_2 &= k = \min\{x_i/i \in 1, \dots, n\} \times 0.5. \end{aligned}$$

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

Tabla 1: Datos de Puromycin.

La figura 2 representa la forma general de este modelo.

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 $b_1 = 212.7$ y $b_2 = 0.0641$ 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 $b_1 = 212.6902$ y $b_2 = 0.0641402$ que da un valor de criterio de 1195.44821, mientras que con RNL_Tabu los parámetros obtenidos son $b_1 = 212.7594$ y $b_2 = 0.063797$ para un criterio de 1196.097. Estos valores del criterio son prácticamente iguales al obtenido por Bates & Watts. Sin embargo, es muy superior que el que se obtiene por linealización del modelo, que es 1920.64.

4.3 Modelo de crecimiento de Micherlich

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

$$f_{M,k,g}(x) = M + ke^{gx}$$

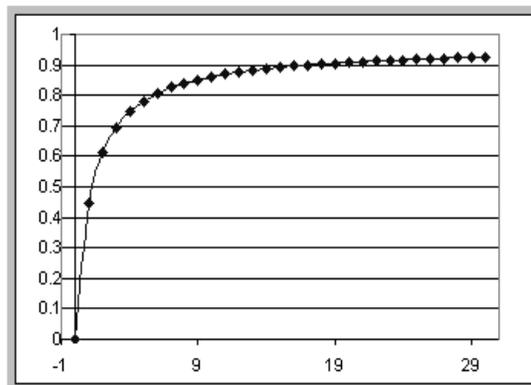


Figura 2: Modelo de crecimiento de Michaelis-Menten.

con $b_0 = M$, $b_1 = k$ y $b_2 = g$.

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

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

En la figura 3 se muestra el gráfico de estos datos de crecimiento.

Para este ejemplo, en [2] —usando el método de Gauss–Newton— se reporta el valor del criterio igual a 0.0534535 para los valores de los parámetros $b_1 = 0.96312$, $b_2 = 2.519$ y $b_3 = -0.10305$. Con RNL_SS se obtuvieron resultados similares, con $b_1 = 0.95622$, $b_2 = 2.52619$ y $b_3 = -0.102529$, para un valor del criterio de 0.053465, y con RNL_Tabu los parámetros obtenidos son $b_1 = 0.96239$, $b_2 = 2.524$, y $b_3 = -0.10238$, para un criterio de 0.054075.

4.4 Velocidad de reacción en cinética enzimática

Los datos son utilizados por Tomassone et al. (1992) [16] y se presentan en la tabla 3.

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

Tabla 3: Datos de Tomassone et al. (1992) para la velocidad de reaccin.

En Tomassone et al. (1992) [16] se reportan los valores de los parámetros mostrados en la tabla 4, con sus respectivos valores del criterio de mínimos cuadrados. En primer

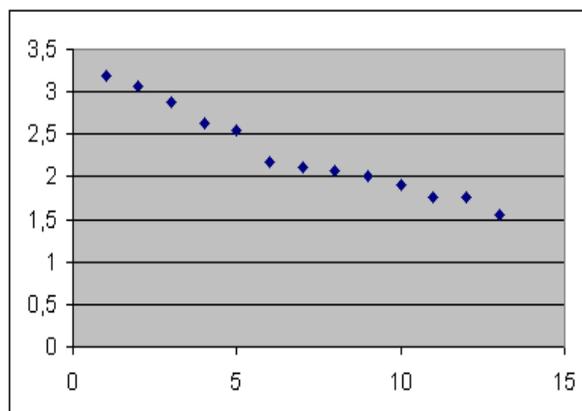


Figura 3: Gráfico de datos para el modelo de crecimiento de Michaelis.

lugar, se presentan los resultados utilizando el modelo de Michaelis–Menten y el método de Gauss–Newton. Además, se muestran los resultados de las linealizaciones siguientes:

- $1/y$ en función de $1/x$ (transformación de Lineweaver–Burk),
- x/y en función de x (transformación de Woolf),
- y en función de y/x (transformación de Eadie).

Finalmente, se presentan los resultados para un método semiparamétrico (basado en la estimación de las medianas para los parámetros).

Modelo	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
Semiparamétrico	1.00	1.34	0.0104194
RNL_SS	0.963902	1.13899	0.0082017
RNL_Tabu	0.96	1.12	0.008325

Tabla 4: Valores obtenidos por distintos métodos para los datos de velocidad de reacción.

Con RNL_SS se obtuvieron los siguientes valores: $M = 0.963902$, $k = 1.13899$, para un valor del criterio de 0.0082017, mejorándose ligeramente 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.

4.5 Datos de crecimiento de cultivos

En [19] se presentan los resultados obtenidos con RNLSS aplicados a modelos de crecimiento de cultivos, en el caso 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.

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

$$f_{M,\kappa,g}(\mathbf{x}) = \frac{M}{1 + \exp(-\kappa(\mathbf{x} - g))} \quad (3)$$

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 4. Se ve que en este caso se tienen 3 parámetros: $\beta_1 = M$, $\beta_2 = \kappa$, y $\beta_3 = g$.

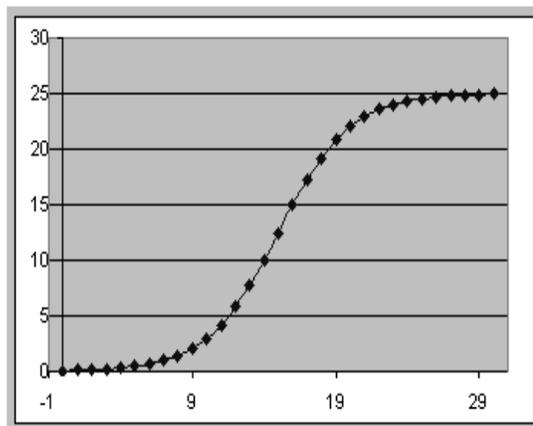


Figura 4: Modelo de crecimiento logístico

En este caso los valores iniciales son:

$$\begin{aligned} b_1 &= M = \max\{y_i / i \in \{1, \dots, n\}\}, \\ b_2 &= \kappa = 1, \\ b_3 &= g = (\max\{x_i\} + \min\{x_i\}) / 2, i \in \{1, \dots, n\}. \end{aligned}$$

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:

$$f_{M,\kappa,g}(\mathbf{x}) = M e^{[-\kappa(\mathbf{x} - g)]} \quad (4)$$

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 5. Para este caso los valores iniciales son:

$$\begin{aligned} b_1 &= M = \max\{y_i/i \in \{1, \dots, n\}\}, \\ b_2 &= \kappa = 1, \\ b_3 &= g = 2/3 \max\{x_i\} + 1/3 \min\{x_i\}, i \in \{1, \dots, n\}. \end{aligned}$$

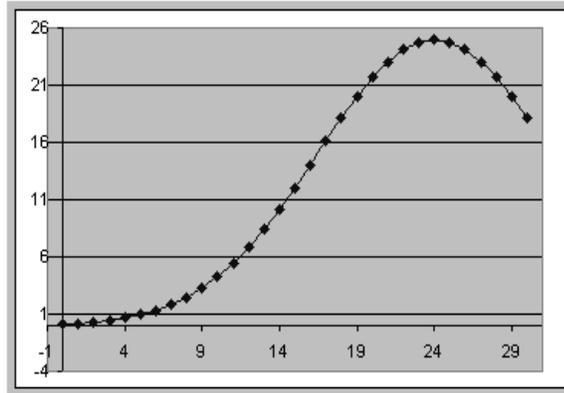


Figura 5: Modelo de crecimiento tipo campana truncada.

A manera de ilustración, presentamos los resultados sobre los tubérculos, usando un modelo logístico (3) y sobre los tallos, usando un modelo de campana truncada (4).

En la tabla 5 se presentan los resultados obtenidos para los tubérculos usando el modelo logístico, comparándolos con los obtenidos con el método de Gauss-Newton y para dos mallados diferentes. En la tabla 6 se presentan los resultados obtenidos para los tallos usando el modelo de campana truncada. Se puede apreciar que los resultados obtenidos son prácticamente los mismos, y que la precisión mejora un poco si el mallado es más fino. A manera de ilustración diremos que en el primer caso el coeficiente de determinación es de 96.37% mientras que en el segundo caso es de 86.25% (aunque prevenimos al lector que este coeficiente, por ser lineal, tiene poco sentido en regresión no lineal).

4.5.1 Otros ejemplos

En Draper & Smith (1966) [8] se analizan diversas series de datos para diferentes modelos no lineales, alguno de regresión múltiple, con el método de Gauss-Newton. Nosotros hemos comparado esas series de datos con el método RNL_Tabu, obteniendo los mismos resultados.

1. Sean t : temperatura, y : presión en vapor, y el modelo

$$\hat{y} = \beta_1(10)^{\frac{\beta_2 t}{\beta_3 + t}}.$$

	G–N (SAS [©])	RNL_SS (100)	RNL_SS (1000)
M	9726.58654	9809.8803	9735.37488
κ	0.069572	0.06845	0.069464
g	283.295855	283.630	283.379
S	84305832.1	84350442.8	84307573.5
$S/\ y\ ^2$	3.6273%	3.6293%	3.6274%

Tabla 5: Resultados comparativos para los tubérculos con el modelo logístico entre la regresión no lineal usando el método de Gauss–Newton (G–N) implementado en SAS[©] y usando sobrecalentamiento simulado (RNL_SS), para mallas de 100 y 1000 puntos en cada dimensión.

	G–N (SAS [©])	RNL_SS (100)	RNL_SS (1000)
M	3079.70691	3085.8357	3080.60547
κ	0.000262	0.00026627	0.00026194
g	280.861285	281.12	280.869
S	64772971	64796860.4	64773042
$S/\ y\ ^2$	13.73878%	13.74685%	13.73879

Tabla 6: Resultados comparativos para los tallos con el modelo de campana truncada entre la regresión no lineal usando el método de Gauss–Newton (G–N) implementado en SAS[©] y usando sobrecalentamiento simulado (RNL_SS), para mallas de 100 y 1000 puntos en cada dimensión.

Método	b_1	b_2	b_3	S
Gauss–Newton	5.2673	8.5651	294.9931	1718.21086
RNL_Tabu	5.2573	8.5651	294.9931	1718.21086

2. Sean \mathbf{x}_1 : tiempo de reacción (min.), \mathbf{x}_2 : temperatura (Kelvin), \mathbf{y} : fracción que queda del material original, y el modelo

$$\hat{\mathbf{y}} = \exp \left\{ -\beta_1 \mathbf{x}_1 \exp \left[-\beta_2 \left(\frac{1}{\mathbf{x}_2} - \frac{1}{620} \right) \right] \right\}.$$

Método	b_1	b_2	S
Gauss–Newton	1.00376	27.539	0.004293
RNL_Tabu	1.00375	27.539	0.004298

3. Sean \mathbf{x}_1 , \mathbf{x}_2 : reacciones parciales, \mathbf{y} : tasa de reacción, y el modelo

$$\hat{\mathbf{y}} = \frac{\beta_1 \beta_3 \mathbf{x}_1}{1 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2}.$$

Método	b_1	b_2	b_3	S
Gauss-Newton	3.57	12.77	0.63	0.00788
RNL_Tabú	3.65	12.68	0.63	0.00795

5 Conclusiones y perspectivas

Se ha visto que la búsqueda tabú tiene una fácil y práctica implementación en el problema de regresión no lineal. Los resultados son comparables a los obtenidos con sobrecalentamiento simulado y con el método de Gauss-Newton. Además, son superiores a los resultados que producen las linealizaciones de modelos no lineales. Se debe aún ahondar en la calibración de los parámetros (la longitud de la lista tabú y el número máximo de iteraciones). También debe explorarse sobre algunos problemas con múltiples mínimos locales, para probar la eficiencia del método en casos donde los métodos de búsqueda local no puedan encontrar

Referencias

- [1] Aarts, E.M.; Korst, J. (1989) *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, Chichester.
- [2] Antoniadis, A.; Berruyer, J.; Carmona, R. (1992) *Régression Non Linéaire et Applications*. Economica, Paris.
- [3] Bard, Y. (1974) *Nonlinear Parameter Estimation*. Academic Press, New York.
- [4] Bates, D.M.; Watts, D.G. (1988) *Nonlinear Regression Analysis and its Applications*. John Wiley & Sons, New York.
- [5] Borowiak. *Model Discrimination for Nonlinear Regression Models*.
- [6] Daniel; Wood. *Fitting Equations to Data*.
- [7] de los Cobos, S.; Trejos, J. (2001) *Optimización Combinatoria a través de Métodos Estocásticos*. Libro en preparación.
- [8] Draper, N.R.; Smith, H. (1968) *Applied Regression Analysis*. John Wiley & Sons, New York.
- [9] Glover, F.; Taillard, E., de Werra, D. (1993) “A user’s guide to tabu search”, *Annals of Ops.Res.* **41**.
- [10] Glover, F.; Laguna, M. (1995) “Tabu search”, in: C. Reeves (Ed.) *Modern Heuristic Techniques for Combinatorial Problems*. McGraw-Hill, London: 70–150.
- [11] Marquardt, D.W. (1963) “An algorithm for least squares estimation of nonlinear parameters”, *Journal of the Society for Industrial and Applied Mathematics* **2**: 431–441.

- [12] Ratkowsky, D.A. (1983) *Nonlinear Regression Modeling. A Unified Practical Approach*. Marcel Dekker, New York.
- [13] Reeves, C. (Ed.) (1995) *Modern Heuristic Techniques for Combinatorial Problems*. McGraw-Hill, London.
- [14] Rodríguez, W.; Trejos, J. (1995) *Análisis y Modelado del Crecimiento del Ñame Alado (Dioscorea alata)*. Informe final de investigación, Universidad de Costa Rica.
- [15] Ross, G. (1990) *Nonlinear Estimation*. Springer-Verlag, New York.
- [16] Tomassone, R.; Audrain, S.; Lesquoy, E.; Millier, C. (1992) *La Régression. Des Nouveaux Regards sur une Ancienne Méthode Statistique*. Masson, Paris.
- [17] Trejos, J.; Murillo, A.; Piza, E. (1998) “Global stochastic optimization for partitioning”, In: A. Rizzi et al. (Eds.) *Advances in Data Science and Classification*, Springer-Verlag, Berlin.
- [18] Trejos, J.; Villalobos, M. (1999) “Optimización mediante recocido simulado en regresión no lineal”, *Memorias del XII Foro Nacional de Estadística*, Monterrey: 183–190.
- [19] Trejos, J.; Villalobos, M. (2000) “Optimización con sobrecalentamiento simulado en regresión no lineal: algoritmo y software”, *Investigación Operacional* 21(3): 236–246. La Habana, Cuba.
- [20] Venables, W.N.; Ripley, B.D. (1994) *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.