

ESTIMACION DE LOS PARAMETROS DEL MODELO INDSCAL POR EL METODO DE SOBRECALENTAMIENTO SIMULADO

William Castillo Elizondo¹ y Jorge González Varela²

Centro de Investigación en Matemática Pura y Aplicada, Escuela de Matemática, Universidad de Costa Rica

RESUMEN

La estimación de los parámetros del modelo INDSCAL por el método CANDECOMP tiene tres desventajas: no se garantiza simetría ni la no negatividad de los pesos y, el mejor valor del *stress* que encuentra no necesariamente corresponde al óptimo global. El método SYMPRES de Ten Berge *et al* aporta una solución a los dos primeros problemas. Nosotros proponemos un nuevo método basado sobre la técnica de sobrecalentamiento simulado, llamado sslNDS, que también resuelve los dos primeros problemas y además el algoritmo converge asintóticamente a un óptimo global con probabilidad igual a uno. Los cálculos numéricos realizados muestran que con sslNDS se consigue un ajuste similar que son SYMPRES y CANDECOMP.

Palabras clave: INDSCAL, sobrecalentamiento simulado, CANDECOMP, SYMPRES

ABSTRACT

Parameter estimation of INDSCAL by CANDECOMP method has three disadvantages: finding the global minimum is not guaranteed, it can find negative weights and a diagonal matrix D such that $X = YD$ may not exist (this is known as the *symmetry problem*). To overcome to the last difficulties, Ten Berge *et al.*, (1993) proposed an algorithm called SYMPRES. We propose in this article a new method based on the simulated annealing technique, called sslNDS that solves the symmetry problem and the non-negativeness of the weights. Moreover the corresponding algorithm converges to the global optimum with probability one. From our computational experiment we obtained results which are similarity to SYMPRES and CANDECOMP methods.

Key words: INDSCAL, simulated annealing, CANDECOMP, SYMPRES

AMS: 62H99

1. EL METODO DE TEN BERGE *et al*

En esta sección se presentan algunos resultados de Ten Berge *et al.* (1991, 1993), relacionados con los problemas de la simetría y la no negatividad de los pesos, en el modelo INDSCAL. También se presenta el método SYMPRES propuesto por ellos para resolver ambos problemas.

1.1. El problema de la simetría y de la no negatividad de los pesos

En el modelo INDSCAL desarrollado por Carroll y Chang (1970) es necesario estimar la matriz X cuyas filas son las coordenadas de n puntos en \mathbb{R}^p y las matrices diagonales de pesos W_1, \dots, W_m de modo que las matrices $XW_k X^t$ aproximen lo mejor posible las matrices simétricas B_k , $k = 1, \dots, m$. Para ello se plantea minimizar la función de *stress*:

$$f(X, W_1, \dots, W_m) = \sum_{k=1}^m \|B_k - XW_k X^t\|^2$$

El algoritmo CANDECOMP, propuesto por Carroll y Chang (1970) consiste en reemplazar la matriz X^t por una matriz Y^t y minimizar la nueva función

$$g(X, Y, W_1, \dots, W_m) = \sum_{k=1}^m \|B_k - XW_k Y^t\|^2$$

¹Fax: +(506) 207 4397; E-mail: wcastill@cariari.ucr.ac.cr

²Fax: +(506) 207 4397; E-mail: jgonzale@cariari.ucr.ac.cr

Este cambio permite estimar alternativamente las matrices X y Y con mínimos cuadrados. Carroll y Chang (1970) plantearon la conjetura de que en el caso que exista convergencia, las matrices estimadas X y Y resultan con columnas proporcionales. Esto es, $Y = XD$ con D una matriz diagonal. Matrices con esta propiedad se llaman equivalentes. Si X y Y son equivalentes se redefinen las matrices W_k como: $W_k = W_k D$ obteniéndose de esta manera el óptimo para la función de Stress: $f(X, W_1, \dots, W_m)$.

A pesar de que hay comprobación práctica de esta conjetura, no se dispone a la fecha de una prueba de ella. Este problema es conocido en la literatura especializada como el problema de la simetría.

Hay además un segundo problema inherente a CANDECOMP el cual es que no se garantiza la no negatividad de los pesos definidos por las matrices W_k . Esta última dificultad ha sido sorteada por algunos autores, utilizando mínimos cuadrados no negativos (NNLS, del inglés *NonNegative Least Squares*).

Ten Berge, J.M.F.; Kiers, H.A.L.; Krijnen, W.P. (1991 - 1993), han trabajado en la discusión y solución de estos dos problemas: simetría y no negatividad de los pesos. A continuación exponemos un breve resumen de sus resultados.

1.2. Algunos resultados sobre simetría y no negatividad de los pesos

Cuando CANDECOMP se aplica con las matrices simétricas B_1, \dots, B_m , entonces las matrices $\tilde{B}_1, \dots, \tilde{B}_m$, donde $\tilde{B}_k = XW_k Y^t$, se aproximan por regresión a las primeras. Ten Berge **et al.**, demostraron lo siguiente:

1. Sean las matrices B_1, \dots, B_m semidefinidas positivas

(a) Entonces en el óptimo global de la función de *stress* f , pueden obtenerse matrices W_k con entradas negativas.

(b) No hay a la fecha prueba ni contraejemplo que permita afirmar que en el óptimo global se de la equivalencia entre X y Y.

2. Si existe alguna B_k no definida³, entonces ellos encontraron un ejemplo donde el óptimo global se alcanza para matrices X y Y no equivalentes y también para matrices equivalentes.

3. Si las matrices B_1, \dots, B_m son definidas positivas, entonces ellos muestran que en algunos óptimos locales se tienen soluciones que son no simétricas, como por ejemplo al considerar las matrices

$$B_1 = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \text{ La función } g(X, Y, W_1, W_2) \text{ alcanza un óptimo local (no simétrico)}$$

de valor 39, en $X = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, Y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, W_1 = (1), W_2 = (-1)$. El óptimo global (simétrico) es 21 y se alcanza en

$$X = Y = \begin{pmatrix} \sqrt{0.5} \\ -\sqrt{0.5} \\ 0 \end{pmatrix}, W_1 = (2) \text{ y } W_2 = (4).$$

1.3. Algoritmo SYMPRES

Ten Berge **et al.** (1993) propusieron un algoritmo que denominaron SYMPRES, el cual minimiza directamente la función Stress f , evitando el problema de la simetría y además permite adicionarle la restricción de no negatividad de los pesos definidos por las matrices W_k .

Hagamos que x_h denote la h-ésima columna de X y w_{kh} la h-ésima entrada de la matriz diagonal W_k : $h = 1, \dots, p$; $k = 1, \dots, m$. Entonces la función de Stress se escribe como:

³ Es decir, existe $x \neq 0$ tal que $x^t B_i x = 0$.

$$\begin{aligned}
f(x_h, w_{1h}, \dots, w_{mh}) &= \sum_{k=1}^m \left\| B_k - \sum_{j=1}^p x_j w_{kj} x_j^t \right\|^2 \\
&= \sum_{k=1}^m \left\| B_k - \sum_{j \neq h}^p x_j w_{kj} x_j^t - x_h w_{kh} x_h^t \right\|^2 \\
&= \sum_{k=1}^m \left\| B_{kh} - x_h w_{kh} x_h^t \right\|^2
\end{aligned} \tag{1}$$

donde $B_{kh} = B_k - \sum_{j \neq h}^p x_j w_{kj} x_j^t$.

Hagamos $x_h^t x_h = 1$ y minimicemos sobre x_h y luego sobre (w_{1h}, \dots, w_{mh}) .

1. Si se fijan w_{1h}, \dots, w_{mh} y se minimiza f sobre x_h se tiene:

$$\min_{x_h} f = \min_{x_h} \left\{ \sum_{k=1}^m \|B_{kh}\|^2 - 2 \text{tr} \left(\sum_{k=1}^m (B_{kh} x_h w_{kh} x_h^t) \right) + \sum_{k=1}^m w_{kh}^2 \right\}$$

Esto es equivalente a maximizar la función l sobre x_h :

$$\begin{aligned}
l(x_h) &= \text{tr} \left(\sum_{k=1}^m x_h^t (w_{kh} B_{kh}) x_h \right) = \text{tr} \left(x_h^t \left(\sum_{k=1}^m w_{kh} B_{kh} \right) x_h \right) \\
&= x_h^t \left(\sum_{k=1}^m w_{kh} B_{kh} \right) x_h = \langle x_h, A x_h \rangle.
\end{aligned}$$

con $A = \sum_{k=1}^m w_{kh} B_{kh}$.

Como se sabe, el máximo sobre $\|x_h\| = 1$, de $\langle x_h, A x_h \rangle$ es λ_1 , el mayor valor propio de la matriz A y se alcanza en x_h , vector propio normalizado, correspondiente a λ_1 . (2)

Si se fija x_h en (1) se tiene que el mínimo se alcanza en

$$w_{kh} = x_h^t B_{kh} x_h \quad \text{para } k = 1, \dots, m. \tag{3}$$

Si $w_h^t B_{kh} x_h < 0$ y como estamos restringidos a que $w_{kh} \geq 0$, el óptimo restringido se tiene para $w_{kh} = 0$.

La implementación del método SYMPRES sigue los pasos que se indican a continuación:

1. Se genera al azar una configuración X y las matrices de pesos W_k .
2. Para cada h , se hace lo siguiente:
 - (a) Se actualiza la columna x_h usando el resultado (2).
 - (b) Se actualizan las matrices de pesos W_k de acuerdo con la fórmula (3).
3. Repetir el paso 2. Hasta que f no decrezca más.

Como f es acotado por abajo y el algoritmo es monótono decreciente, el valor de f se reduce monótonamente hasta que se vuelve estable.

2. EL METODO ssINDS

En esta sección proponemos un nuevo método para estimar los parámetros del modelo INDSCAL, el cual se basa en la técnica de sobrecalentamiento simulado. Mediante este procedimiento se minimiza directamente la función Stress y se controla el signo de los pesos. Esto resuelve el problema de la simetría y el de la no negatividad de los pesos.

El método de sobrecalentamiento simulado (SS) es una técnica estocástica usada para optimizar la función objetivo en un problema de optimización combinatoria. Se basa en una analogía con el método físico de sobrecalentamiento (o *annealing* en inglés). Consiste en generar aleatoriamente cadenas de estados del dominio de definición del problema de optimización combinatoria, aceptando los nuevos estados (vecinos) de acuerdo con la regla de Metrópolis.

El método necesita la estimación del parámetro de temperatura inicial denotado c_0 , que corresponde a una situación de aceptación de un alto porcentaje (χ) de nuevos estados, digamos por ejemplo $\chi \in [0.85, 1[$. El parámetro de temperatura c_t se hace descender paulatinamente de acuerdo con alguna ley funcional, tal como la ley de descenso geométrico $c_{t+1} = \gamma c_t$, con, por ejemplo $\gamma \in [0.85, 1[$. Los fundamentos teóricos del método de sobrecalentamiento simulado pueden consultarse en Aarts, E. & Korts, J. (1989).

El método ha sido utilizado con éxito para optimizar el criterio en particionamiento unimodal (Trejos, J. **et al.**, 1998) y bimodal (Trejos, J. & Castillo, W., 2000). también se han propuesto métodos de optimización por sobrecalentamiento simulado, en casos de problemas no combinatorios, mediante una discretización adecuada del espacio de estados. Tal es el caso de la regresión no lineal (Trejos, J. & Villalobos, M. 1999(b)) y de *multidimensional scaling* en su versión métrica, para una sola tabla de distancias (Trejos, J. & Villalobos, M., 1999 (a) y Villalobos, M. (1998).

En este artículo proponemos un nuevo procedimiento basado en la técnica de sobrecalentamiento simulado para estimar los parámetros del modelo INDSCAL. La discretización del espacio y la generación de vecinos, es similar a la propuesta por Trejos, J. y Villalobos, M. en 1998. Este sistema de generación, consiste en definir un mallado bidimensional y una forma de perturbar las filas de una matriz para crear una cadena de nuevos estados (matrices), tal y como se describe a continuación.

2.1. Sistema de generación de estados

Sea X una configuración de tamaño $n \times p$ y W la matriz $m \times p$ cuyas filas son los pesos no negativos del modelo INDSCAL. Un estado es una matriz Z de tamaño $(n + m) \times p$ de la forma

$$Z = \begin{pmatrix} X \\ W \end{pmatrix}$$

Se considera un mallado bidimensional ($p = 2$) con paso igual a h y se define el conjunto de las direcciones canónicas del plano:

$$D = \{(1,0), (0,1), (-1,0), (0,-1)\}$$

Un estado *vecino* de Z es la matriz que se obtiene de Z , perturbando cualquiera de sus filas, como se explica enseguida:

1. Se selecciona X ó W al azar, con probabilidad $\frac{1}{2}$.
2. Se selecciona una dirección $d \in D$, al azar con probabilidad $\frac{1}{4}$.
3. Si ocurre X entonces se selecciona una fila z de la matriz X , al azar, con probabilidad $\frac{1}{n}$.
4. Si ocurre W entonces se selecciona una fila z de la matriz W , al azar, con probabilidad $\frac{1}{m}$.

5. Sea z la fila seleccionada en el paso 3 o 4, entonces la fila perturbada en \tilde{z} definida así: $\tilde{z} = z + hd$ si z es una fila de X o si z es una fila de W tal que \tilde{z} tiene todas sus entradas no negativas. En todo otro caso se define $\tilde{z}_j = \begin{cases} 0 & \text{si } (z + hd)_j < 0 \\ (z + hd)_j & \text{si no} \end{cases}$ donde $\tilde{z} = (\tilde{z}_1, \tilde{z}_2)$.

El conjunto de todos los vecinos de Z , denotado $V(Z)$, se llama *vecindario* de Z . La probabilidad de generar un vecino de Z , es $\frac{1}{8n} + \frac{1}{2m} \left(\frac{m_1}{2} + \frac{m_2}{3} + \frac{m_3}{4} \right)$ donde m_i es el número de filas de W que se pueden perturbar en $i + 1$ direcciones distintas, con $m = m_1 + m_2 + m_3$.

2.2. Fórmulas de actualización

Recordemos que w_k es la fila k ésima de la matriz W , W_k es la matriz diagonal de pesos, cuyas entradas son los elementos de w_k . Para la implementación computacional del método que proponemos es necesario encontrar una expresión sencilla de la variación del criterio cuando se pasa de un estado a su vecino. La función *stress* que usamos para la estimación de los parámetros del modelo INDSCAL es,

$$S = \sum_{k=1}^m \sum_{j=1}^m \sum_{i=1}^n (b_{kji} - x_j W_k x_i^t)^2.$$

Su variación ΔS se puede escribir así:

1. Sea $\tilde{x}_r = x_r + h d$ donde x_r es una fila de X , b_{kij} la entrada i, j de la matriz B_k , $x_i W_k x_j^t$ es el producto interno⁴ de las filas x_i y x_j de la matriz X , determinado por la matriz W_k . Entonces

$$\Delta S = \sum_{k=1}^m \sum_{i=1}^n \left[(b_{kri} - x_r W_k x_i^t)^2 - (b_{kri} - \tilde{x}_r W_k x_i^t)^2 \right] + \sum_k \left[(b_{krr} - \tilde{x}_r W_k x_r^t)^2 - (b_{krr} - x_r W_k x_r^t)^2 \right]$$

2. Similarmente, sea $\tilde{x}_r = w_r + hd$, entonces

$$\Delta S = \sum_{i=1}^n \sum_{j=1}^n \left[(b_{rij} - x_i W_r x_j^t)^2 - (b_{rij} - x_i \tilde{W}_r x_j^t)^2 \right]$$

donde \tilde{W}_r es la matriz diagonal cuyas entradas en la diagonal son los elementos de \tilde{w}_r .

2.3. El algoritmo

La implementación computacional del método se basa en el siguiente esquema algorítmico:

- Se escoge una configuración X y una matriz W , de pesos no negativos, de acuerdo con el siguiente procedimiento; cada entrada de X (resp. de W) se escoge al azar con probabilidad uniforme en $[-1, 1]$ (resp. en $[0, 2]$). Escoger un valor del paso h (por ejemplo $h = 0.1$) y calcular la temperatura inicial c_0 . Definir $t = 0$.
- Sea $t = t+1$ y $c_t = \gamma c_{t-1}$. Repetir lc veces, los pasos (a) - (c):
 - Escoger al azar la matriz X , o la matriz W , con probabilidad $\frac{1}{2}$.
 - Generar un vecino de $Z = \begin{pmatrix} X \\ W \end{pmatrix}$, digamos \tilde{Z} , de acuerdo con el sistema de generación definido en la sección 2.1.

⁴ Si la matriz W_k tiene alguna entrada igual cero en su diagonal, entonces no define un producto interno, pero ello no afecta nuestros resultados.

(c) Calcular ΔS y aceptar (o no aceptar) \tilde{Z} de acuerdo con la regla de Metrópolis⁵. Si \tilde{Z} es aceptado, definir $Z_t = \tilde{Z}$.

3. Si $c_t < \varepsilon$, terminar⁶. Si no, regresar a 2.

Este algoritmo será denotado por ssIND.

2.4. Acerca de la convergencia

Para garantizar la convergencia del algoritmo recién definido es necesario asegurar ciertas propiedades de las probabilidades de generación de nuevos estados (ver Aarts, E. & Korst, J. 1989).

Con el fin de garantizar que se tienen todas las hipótesis del teorema de convergencia del algoritmo de sobrecalentamiento simulado, se define el espacio de estados E , como el conjunto de configuraciones que incluye la configuración inicial obtenida mediante el procedimiento descrito en el paso 1 del algoritmo, así como todos los posibles nuevos estados generados por medio del paso 2 (esta definición corresponde a los estados generados por una "corrida" del algoritmo ssIND). En este contexto son válidas las dos siguientes propiedades:

Simetría: sea G_{ij} la probabilidad de generar el estado j a partir del estado i . Entonces $G_{ij} = G_{ji}$ para todo i, j . En efecto, es claro que hay diferentes rutas o cadenas para "ir" de i a j , de manera que cada paso o eslabón de la cadena consiste en generar un vecino del estado actual. Así entonces, la probabilidad de generar una cadena de i a j es igual a la probabilidad de generar la cadena inversa de j a i . Como G_{ij} es igual a la suma de las probabilidades de las distintas cadenas de i a j , es claro que $G_{ij} = G_{ji}$ para todo par de estados i, j de E .

Conexidad: para todo par de estados i, j de E , existe $\{l_0, l_1, \dots, l_p\} \subset E$, $p \geq 1$, tal que $i = l_0$, $j = l_p$ y $G_{l_s l_k} > 0$ para todo l_s, l_k . Esta propiedad es obvia puesto que cualquier cadena del estado i al estado j sirve como $\{l_0, l_1, \dots, l_p\}$.

3. COMPARACION DE LOS METODOS CANDECOMP y ssINDS

Siguiendo un procedimiento análogo al de Ten Berge *et al.* (1993), para comparar los métodos CANDECOMP y SYMPRES, hemos realizado el siguiente experimento computacional para comparar CANDECOMP y ssINDS: denotamos un grupo de m matrices de tamaño n , por $n \times n \times m$. Los resultados se obtuvieron al considerar 20 grupos de tres matrices $3 \times 3 \times 3$ y 20 grupos de matrices de $6 \times 6 \times 9$. Cada matriz B_k fue construida generando primero una matriz A con entradas tomadas al azar con distribución uniforme en $[-1, 1]$, luego se calcula AA^t , y si esta es definida positiva, se define⁷ $B_k = AA^t$.

La calidad del ajuste se mide como un porcentaje de la suma de los cuadrados de los datos. Es decir,

$$porc = 1 - \frac{\sum_k \sum_{i \leq j} (b_{kij} - x_i^t W_k x_j)^2}{\sum_k \sum_{i \leq j} b_{kij}^2}$$

De manera similar a como lo reporta Ten Berge *et al.* (1993) ambos programas se ejecutaron⁸ 10 veces para cada grupo de matrices y se escogió la mejor solución de acuerdo con el mayor valor de *porc* obtenido. Luego se hizo el promedio sobre estos mejores 20 valores. Estos porcentajes de ajuste promedio se reportan en el siguiente cuadro. También se incluyen los resultados de Ten Berge *et al.* (1993) para los algoritmos

⁵ Esta regla es: "se acepta \tilde{Z} si $\Delta S < 0$ si $\exp\left(-\frac{\Delta S}{c_t}\right) > \alpha$ con α escogido al azar con probabilidad uniforme en $[0,1]$."

⁶ El parámetro $\varepsilon > 0$ es un umbral "suficientemente" pequeño, escogido por el usuario.

⁷ Este es el mismo procedimiento usado por Ten Berge *et al.* (1993).

⁸ Para el método ssINDS, se usaron los siguientes valores de los parámetros: temperatura inicial es $c_0 = 0.6$ para el caso de matrices $3 \times 3 \times 3$, y $c_0 = 1.8$ para el caso de matrices $6 \times 6 \times 9$. En ambos casos se usó $\gamma = 0.85$ y longitud de la cadena $lc = 800$.

SYMPRES y CANDECOMP, aplicados a 100 grupos de matrices $3 \times 3 \times 3$ y de $6 \times 6 \times 9$. La columna titulada CANDE-TB corresponde a los resultados de Ten Berge *et al.* (1993).

orden	ssINDS	CANDECOMP	CANDE-TB	SYMPRES
$3 \times 3 \times 3$	90.73	88.60	87.02	87.02
$6 \times 6 \times 9$	49.90	49.97	43.23	43.23

Como se puede ver en la tabla el porcentaje obtenido con el método ssINDS es un poco mejor que con CANDECOMP; en el caso de tablas $3 \times 3 \times 3$. La diferencia es 2.13. Mientras que en el caso $6 \times 6 \times 9$, la diferencia a favor de CANDECOMP es apenas 0.07. Por otra parte, los porcentajes obtenidos por Ten Berge *et al.* son menores, quizás debido a alguna diferencia en la definición de *porc* o del *stress*, con respecto a las nuestras.

4. CONCLUSIONES

A partir de los resultados numéricos obtenidos, vemos que el comportamiento del método ssINDS es satisfactorio. Sin embargo sería necesario hacer más experimentación para tener una tendencia más confiable sobre el comportamiento de los métodos.

Comparado con CANDECOMP, una ventaja del método ssINDS es que resuelve de manera simple el problema de la no negatividad de los pesos. Esto se logra a nivel del algoritmo: si un peso es negativo no se modifica la configuración y se continúa con el proceso.

Por otra parte, el método ssINDS puede ser adaptado fácilmente para usar directamente las matrices de disimilitud, sin pasar por el intermediario de los productos escalares aproximados. Es decir, se puede usar la siguiente función de *stress* :

$$f(X, W_1, \dots, W_m) = \sum_{k=1}^m \|\delta_k - d(X, W_k)\|^2$$

donde δ_k es la k -ésima matriz de disimilitud, y $d(X, W_k)$ es la matriz de distancia entre las filas de X , con pesos definidos por W_k .

REFERENCIAS

AARTS, E. & J. KORST (1989): **Simulated Annealing and Boltzman Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computing**, John Wiley & Sons, Chichester.

CARROLL, J.D. and J.J. CHANG (1970): "Analysis of individual differences in multidimensional scaling via an n -way generalization of Eckart-Young decomposition", **Psychometrika**, 35:238-319.

TEN BERGE, J.M.F.; P.A. BEKKER and H.A.L. KIERS (1991): "Some clarifications of the tuckals2 algorithm applied to the IDIOSCAL problem", **Psychometrika**, 59:193-201.

____ (1993): "Computational solutions for the problem of negative saliences and nonsymmetry in INDSCAL", **Journal of classification**, 10:115-124.

TREJOS, J. y M. VILLALOBOS (1998): "Análisis de proximidades usando sobrecalentamiento simulado", Castillo, W. & Trejos, J. (Eds.) **Estudios de Análisis de Datos y Estadística**, Universidad de Costa Rica - Instituto Tecnológico de Costa Rica.

VILLALOBOS, M. (1998): **Optimización Estocástica para el Análisis de Proximidades**, Tesis de Maestría, Universidad de Costa Rica.

- TREJOS, J.; A. MURILLO y E. PIZA (1998): "Global stochastic optimization for partitioning", En: A. Rizzi, M. Vichi & H.H. Bock (Eds.), **Advances in Data Science and Classification**, Springer, Heidelberg, 185-190.
- TREJOS, J. and M. VILLALOBOS (1999a): "Use of simulated annealing in metric multidimensional scaling", International Conference on Large Scale Data Analysis., Köln.
-
- (1999b): "Optimización mediante recocido simulado en regresión no lineal", En: **Memorias del XIII Foro de la Asociación Mexicana de Estadística**, J. Sierra **et al.** (Eds.), Monterrey, México (en prensa).
- TREJOS, J. and W. CASTILLO (2000): "Simulated annealing optimization for two-mode partitioning", in: W. Gaul & R. Decker (Eds.) **Classification and Information Processing at the Turn of the Millenium**, Springer, Heidelberg (in press).