

Mínimos cuadrados parciales con el método de descenso de mayor pendiente

Eddy Jackeline Rodríguez

Centro de Investigación de Matemática Aplicada (CIMA).
Facultad de Ingeniería. Universidad del Zulia. Maracaibo, Venezuela.
email: eddyjackeline@yahoo.es

Recibido: 18-05-2012 Aceptado: 02-11-2012

Resumen

El estudio del modelado de problemas de regresión lineal es de interés en varios campos de la ciencia como en: la quimiométrica, economía, física, entre otros. Los mínimos cuadrados parciales (PLS) tienen su origen en las investigaciones realizadas por Geladi, Kowalski y Hoskuldsson. El objetivo de este trabajo es presentar un desarrollo del PLS basado en el método de descenso de mayor pendiente, por lo cual se expone previamente un estudio de estos métodos. El primero es una técnica que genera un nuevo espacio de la función regresión al reducir el espacio de la matriz de entrada, y el segundo es un método de optimización que se basa en una función a minimizar, a partir de un punto cualquiera y dirección de descenso igual al gradiente de la función. Se demuestra que es posible aplicar el método de descenso de mayor pendiente para encontrar en el PLS los vectores proyección entre las matrices de entrada y respuesta y las variables latentes. Los datos experimentales verifican que con este método de optimización el PLS modela problemas de regresión lineal de forma satisfactoria. La investigación se llevó a cabo usando una metodología descriptiva-comparativa partiendo de la consulta en fuentes bibliográficas y páginas web.

Palabras claves: PLS, método de descenso de mayor pendiente, función de costo, problema de optimización.

Partial least square with method of maximum gradient descent

Abstract

The study of problems of linear regression modeling is of great interest in various fields of Science: as in the chemometrics, economics, physics, among others. The partial least squares (PLS) have their origin in the research conducted by Geladi, Kowalski and Hoskuldsson. The objective of this paper is to present a development of the PLS based method of maximum gradient descent, which is exposed by a previous study of these, the first is a technique that generates a new regression function space by reducing the space the input matrix and the second is an optimization method based on a function to be minimized, a point from any and down direction equal to the gradient of the function. We show that it is possible to apply the method of maximum gradient descent to find the PLS projection vectors between input and response matrices and latent variables. The experimental data verified that this method of optimizing the PLS regression model problems satisfactorily. The research was carried out using a descriptive-comparative methodology on the basis of consultation in bibliographical sources and websites.

Key words: PLS, method of maximum gradient descent, function cost, optimization problem.

Introducción

Uno de los principales estudios en el análisis estadístico consiste en modelar problemas de gran dimensión, problemas que involucran un gran número de variables basados en pocas observaciones. Para estos tipos de problemas la meta es encontrar una estructura que exprese las características de los datos y los represente en un menor número de variables.

El presente trabajo aborda un estudio sobre el modelado de problemas de regresión lineal, de manera específica el llamado método de mínimos cuadrados parciales (PLS). Este método tiene la característica principal que puede ser aplicado para modelar cualquier tipo de problema aproximadamente lineal, aún cuando las variables que estén involucradas sean linealmente independientes o no. El estudio del PLS fue desarrollado inicialmente para aplicaciones quimiométricas [1], [2], [3] y [4]; sin embargo, se extendió a otras ramas de la ciencia por ser una técnica útil en el modelado de problemas de regresión y clasificación, en especial en situaciones donde la data experimental se caracteriza por tener una alta dimensionalidad, elevada multicolinealidad y pocas observaciones. Este método permite establecer estrategias de regularización evitando sobre estimación y solventando el mal condicionamiento de problemas con alta colinealidad. En este trabajo, el estudio del PLS se basa primeramente en una perspectiva geométrica que sirve de base para la formulación matemática, seguido de un criterio optimal que permite la reconstrucción de la matriz predictiva y la matriz respuesta a través de la suma de matrices de rango uno, junto con el conjunto reducido de variables latentes artificiales que se utilizan para describir las variaciones relevantes de estas matrices.

Actualmente son varios los estudios realizados sobre el desarrollo del algoritmo del PLS y sobre su uso en el ajuste de problemas de regresión [5], [6] y [7]; es por esto que en este trabajo, mediante una investigación descriptiva-comparativa, se presenta otra visión del desarrollo de la formulación del PLS planteándolo como un problema que involucra la solución de problemas de optimización sin restricción, usando el método de descenso de mayor pendiente para la búsqueda de las variables latentes. Se expone este método específicamente porque ofrece posibilidades sencillas y directas para las aplicaciones prácticas. El método de descenso de mayor pendiente conocido también como el método del gradiente, constituye una base para algoritmos avanzados que modifican esta técnica para mejorar las propiedades de convergencia y consiste en la minimización de una función, siguiendo como dirección de descenso el vector que corresponde al gradiente de la función y aplicando un proceso iterativo que lleva al óptimo de la función.

Método de cuadrados mínimos parciales (PLS)

El método de cuadrados mínimos parciales (PLS), relaciona la matriz de las variables de entrada $x \in X \subset R^m$ y la matriz de las variables de salida $y \in Y \subset R^p$ de un problema de regresión a través de sus variables latentes, las cuales se pueden definir como un conjunto de variables artificiales ortogonales que permiten hacer una reducción del espacio siguiente:

$$\begin{array}{l} X = TP^T + E \\ Y = UQ^T + F \end{array} \longrightarrow Y = \beta X \quad (1)$$

β : es el vector de coeficientes de regresión. T y U son las matrices compuestas por los vectores latentes. P y Q son las matrices compuestas por los vectores de peso. E y F son las matrices residuales.

El método PLS se basa en la proyección de las variables sobre nuevas bases cuyas direcciones se determinan usando el criterio optimal [11]:

$$\begin{array}{l} \text{Minimizar } J(w, c) = \sum_{i=1}^n \|X - Xw w^T\|^2 + \|Xw - Yc\|^2 + \|Y - Yc c^T\|^2 \\ \text{Sujeto a: } \|c\| = \|w\| = 1 \end{array} \quad (2)$$

Esta ecuación puede formularse como un problema optimal de forma cuadrática convexa con restricciones, que tiene por objetivo encontrar los vectores proyección (w, c).

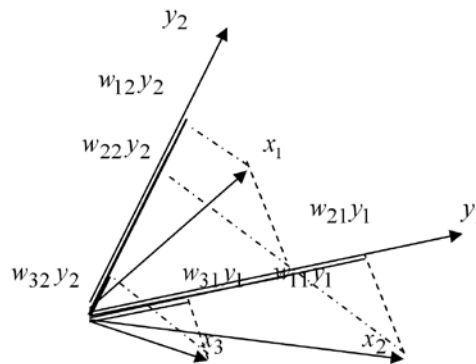
Utilizando este criterio se obtiene la reconstrucción de la matriz predictiva X y la matriz de respuestas Y como las sumas de matrices de componentes de rango uno, y un conjunto reducido de variables latentes artificiales que se utilizan para describir las variaciones más significativas de estas matrices, mientras que los residuales se van obteniendo por el proceso de reconstrucción y está relacionado con las variaciones insignificantes de ambas matrices. Esto implica buscar vectores proyección (loadings) o variables latentes (scores) que maximicen la covarianza entre los diferentes conjuntos de variables y al mismo tiempo mantengan la mayor parte de la varianza en ambos conjuntos.

El método PLS es un esquema aplicable a conjuntos multivariantes y consiste en obtener de forma iterativa direcciones ortogonales en los espacios de variables con el objeto de formar bases optimales donde proyectarlas con la características ya mencionadas, seguido de una regresión de mínimos cuadrados sobre estas proyecciones o sobre un subconjunto de las mismas.

Perspectiva geométrica del PLS

Por sencillez se representa tres vectores de entrada $X = [x_1 \ x_2 \ x_3]$ y dos vectores de salida $Y = [y_1 \ y_2]$. Se parte de una relación entre la matriz de entrada y matriz de salida, a través de la proyección ortogonal de los vectores de X sobre el espacio definido por Y , como se muestra en la figura 1.

Figura 1: proyección de los vectores x ; sobre los vectores y_j



Donde: $w_{ij}y_j$ ($i = 1,2,3; j=1,2$) vectores proyección. Con $w_j = \frac{X^T y_j}{y_j^T y_j}$

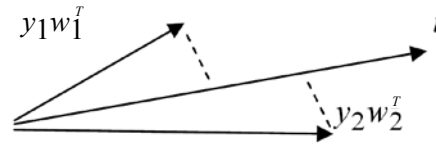
Con $j=1,2$

$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} = [w_1 \ w_2]$$

Se asume normalizados los vectores de W .

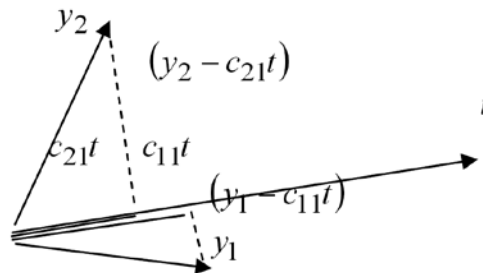
Se selecciona solo un vector (w_{ij}) de la matriz W , escogiendo aquel que proporciona una mayor relación (covarianza) entre la matriz X y la matriz Y , minimizando $\|X - YW^T\|$.

El primer vector latente $t \in T$ debe ubicarse en la dirección donde los vectores de $YW^T = y_1 w_1^T + y_2 w_2^T$ aportan su mayor dispersión, ver figura 2.

Figura 2. Representación del primer vector t 

Como $X \approx YW^T$, despejando cualquier vector de la matriz Y queda: $y = \frac{Xw}{w^T w}$, de lo que se evidencia que cualquier vector en el subespacio de Y puede escribirse como una combinación lineal: $t = \sum_{i=1}^n w_{ij}$, donde t será el primer vector de la matriz T.

El siguiente paso es proyectar de forma ortogonal los vectores de la matriz Y sobre el vector t encontrado, ver figura 3.

Figura 3: Representa la proyección de la matriz Y sobre el vector t 

Los vectores c están normalizados.

En el caso particular en que la matriz Y está formada por una sola salida, entonces el término c corresponde al escalar uno.

El vector c proporciona la mayor relación (covarianza) entre la matriz T y la matriz Y, minimizando $\|Y - tc^T\|$.

De forma similar como se procedió para encontrar al primer vector de T, se procede para encontrar al primer vector de U.

Como $Y \approx tc^T$, y despejando $t = \frac{Yc}{c^T c}$, se evidencia que cualquier vector en el subespacio de T puede escribirse como una combinación lineal: $u = \sum_{i=1}^n c_i y_i$, donde u es el primer vector de la matriz U.

Formulación matemática del PLS

El método PLS, tiene como objetivos la maximización de la covarianza muestral entre las variables latentes de cada matriz de datos, y la minimización del cuadrado del error [8], [9]. Basado en la perspectiva geométrica anterior:

Paso 1: Encuéntense los vectores “ c ” y “ w ” maximizando la covarianza entre las variables latentes $t = Xw$ y $u = Yc$, lo que en el enfoque geométrico representaba la minimización de la distancia entre los vectores de la matriz X y sus proyecciones sobre Y: $\min\|X - XW^T\|$ y la minimización de la distancia entre los vectores de la matriz Y con sus proyecciones sobre t : $\min\|Y - tc^T\|$:

$$\max(\text{cov}(t, u)) = \max(\text{cov}(Xw, Yc)) \quad (3)$$

$$\text{sujeto a: } \begin{aligned} w^T w &= 1 \\ c^T c &= 1 \end{aligned}$$

Esto consiste en un problema de optimización con restricciones, que se puede resolver aplicando el método de los Multiplicadores de Lagrange y las condiciones de Khun-Tucker, quedando:

$$\lambda_1 = w^T X^T Y c, \lambda_2 = c^T Y^T X w. \text{ Es de notar que los valores de } \lambda_1 \text{ y } \lambda_2 \text{ son iguales. Además: } w = \frac{X^T Y c}{\lambda_1},$$

$$c = \frac{Y^T X w}{\lambda_1}$$

Paso 2: Se procede a encontrar el primer vector de las matrices T y U, a partir de: $t = Xw$ y $u = Yc$

Paso 3: El siguiente paso es calcular las matrices residuales de X e Y, y si la matriz Y no es lo suficientemente pequeña, proceder a encontrar los segundos vectores latentes siguiendo el mismo procedimiento anterior. Así sucesivamente hasta que la matriz Y sea insignificante.

Para garantizar la ortogonalidad de la secuencia de variables latentes, a las matrices de datos originales se les va substrayendo en forma iterativa la parte de la información explicada por la variable latente obtenida en la presente iteración. Esto se logra restándole a las matrices residuales las matrices optimales de rango 1, basadas en la variable latente t obtenida en la presente iteración. Estas matrices se obtienen resolviendo $\min_p \|X - tp^T\|$ y $\min_q \|Y - tq^T\|$. Se puede demostrar que las soluciones están dadas por $p = X^T t$ y $q = Y^T t$ respectivamente, y que la actualización de las matrices se puede representar como $X = X - tt^T X$ y $Y = Y - tt^T Y$

Método de descenso de mayor pendiente

Este método es uno de los más antiguos usados en el estudio de minimización de una función de varias variables, también conocido como el método del gradiente [10], constituye la base de algoritmos avanzados que intentan modificar esta técnica en busca de mejores propiedades de convergencia.

Sea $f(x)$ una función continua en E^n , con primeras derivadas parciales continuas en E^n , y sea $\nabla f(x)$ el gradiente de $f(x)$ un vector fila n-dimensional. Defínase a $g(x) = \nabla f(x)^T$ como un vector columna n-dimensional, entonces el método de descenso de mayor pendiente está definido por el algoritmo iterativo:

$$x_{k+1} = x_k - \alpha_k g_k \tag{4}$$

donde α_k es un escalar no negativo que minimiza $f(x_k - \alpha_k g_k)$. En particular, todas las características importantes de convergencia local del método de descenso de mayor pendiente se observan estudiando este método a problemas cuadráticos de la forma:

$$f(x) = \frac{1}{2} x^T Q x - x^T b, \tag{5}$$

donde Q es una matriz de tamaño $n \times n$ simétrica definida positiva. El único punto mínimo de f se puede hallar directamente, igualando el gradiente a cero:

$$Qx^* = b$$

denotando a x^* como el punto óptimo.

El gradiente de f viene dado por: $g(x) = Qx - b$

Así, el método de descenso de mayor pendiente se expresa a partir de la fórmula (4) con $g_k = Qx_k - b$, donde a_k minimiza $f(x_k - a_k g_k)$:

$$f(x_k - a_k g_k) = \frac{1}{2} (x_k - a_k g_k)^T Q (x_k - a_k g_k) - (x_k - a_k g_k)^T b \tag{6}$$

obteniendo por resultado: $a_k = \frac{g_k^T g_k}{g_k^T Q g_k}$. (7)

$$\text{Es decir: } x_{k+1} = x_k - \left(\frac{g_k^T g_k}{g_k^T Q g_k} \right) g_k \quad (8)$$

con $g_k = Qx_k - b$

Algoritmo del método de descenso de mayor pendiente

Dado x_0 y para $k = 0$ hasta n :

Calcúlese:

- $g(x_k) = \nabla f(x_k)^T$, en el caso cuadrático: $g(x_k) = Qx_k - b$
- $a_k =$ es un escalar que minimiza a $f(x_k - a_k g_k)$, en el caso cuadrático:

$$a_k = \frac{g_k^T g_k}{g_k^T Q g_k}$$

$$x_{k+1} = x_k - a_k g_k$$

Desarrollo del método PLS mediante el método de optimización

El método PLS bajo la perspectiva geométrica involucra dos problemas de minimización: el primero expone a través de w la minimización de la distancia entre los vectores de la matriz X y sus proyecciones sobre la matriz Y , es decir: $\min \|X - Yw^T\|$ y el segundo expresa mediante c la minimización de la distancia entre los vectores de la matriz Y con sus proyecciones sobre T , es decir: $\min \|Y - Tc^T\|$. Estos problemas se pueden expresar como las funciones objetivos siguientes, que se llamarán funciones objetivos del PLS:

$$F_1(w) = (X - Yw^T)^T (X - Yw^T) = (X^T - wY^T) (X - Yw^T) = (X^T X - X^T Yw^T - wY^T X + wY^T Yw^T = K - 2X^T Yw^T + wY^T Yw^T$$

donde $K = X^T X$ es una constante. Se divide la ecuación entre dos y se escribe como una ecuación cuadrática:

$$F_1 = \frac{K}{2} - bw^T + \frac{1}{2} wQw^T \quad (9)$$

con: $b = X^T Y$, $Q = Y^T Y$

$$F_2(c) = (Y - Tc^T)^T (Y - Tc^T) = (Y^T - cT^T) (Y - Tc^T) = (Y^T Y - Y^T Tc^T - cT^T Y + cT^T Tc^T = K - 2Y^T Tc^T + cT^T Tc^T$$

donde $K = Y^T Y$ es una constante. Se divide la ecuación entre dos y se escribe como una ecuación cuadrática:

$$F_2 = \frac{k}{2} - bc^T + \frac{1}{2} cQc^T \quad (10)$$

con: $b = Y^T T$, $Q = T^T T$

La solución de estas funciones objetivos del PLS sustituye la forma convencional de búsqueda de los vectores w , c dadas por [1], [3] y [8].

Formulación matemática del PLS con el método de descenso de mayor pendiente

Aplicando los pasos del algoritmo del método de descenso de mayor pendiente a las funciones (9) y (10), queda:

$$\text{Para } F_1: w_{k+1} = w_k - a_k g_k$$

con gradiente: $g_k = g(w_k) = Y^T Y w_k - X^T Y$

$$\text{y } a_k = \frac{g_k^T g_k}{g_k^T Q g_k} \text{ que minimiza a } F_1(w_k - a_k g_k).$$

Como F_1 es una función cuadrática, donde Q es una matriz cuadrada simétrica definida positiva, su único punto mínimo se puede hallar directamente, al igualar el gradiente a cero

$$Y^T Y w_k - X^T Y = 0 \quad (11)$$

obteniéndose el vector proyección: $w_k = (Y^T Y)^{-1} (X^T Y)$ o $w_j = \frac{x^T y_j}{y_j^T y_j}$ por considerar los vectores que forman la matriz Y tomados uno por uno.

$$\text{Para } F_2: c_{k+1} = c_k - a_k g_k$$

con el gradiente de F_2 : $g_k = g(c_k) = T^T T c_k - Y^T T$

$$\text{y } a_k = \frac{g_k^T g_k}{g_k^T Q g_k} \text{ que minimiza a } F_2(c_k - a_k g_k).$$

Ya que F_2 es una función cuadrática, donde Q es una matriz cuadrada simétrica definida positiva, su único punto mínimo puede encontrarse al igualar el gradiente a cero:

$$T^T T c_k - Y^T T = 0 \quad (12)$$

Obteniéndose el vector proyección: $c_k = (T^T T)^{-1} (Y^T T)$ o $c_j = \frac{Y^T t_j}{t_j^T t_j}$ por considerarse los vectores de la matriz T tomados uno por uno.

Algoritmo del PLS a partir del método de mayor pendiente

Con las matrices X y Y centradas se procede con los siguientes pasos:

$$1. \text{ Obtener los vectores: } w: \text{ a partir de } F_1 = \frac{k}{2} - b w^T + \frac{1}{2} w Q w^T$$

$$c: \text{ a partir de } F_2 = \frac{k}{2} - b c^T + \frac{1}{2} c Q c^T$$

$$2. \text{ Calcular: } t = X w, \text{ hacer } t = \frac{t}{\|t\|}, u = Y c, \text{ hacer } u = \frac{u}{\|u\|}$$

$$4. \text{ Obtener las nuevas matrices: } X = X - t t^T X, \quad Y = Y - t t^T Y$$

5. Repetir hasta que se cumpla un criterio de parada preestablecido.

$$6. \text{ Calcular el coeficiente de regresión: } B = X^T U (T^T X X^T U)^{-1} T^T Y$$

$$7. \text{ Obtener el Modelo de Regresión: } Y = B X$$

Resultados y Conclusiones

Se propone un ejemplo de datos experimentales de 10 variables con 100 observaciones por variable, creados a través de la ecuación:

$$y = 5x_1 + x_2 + 7x_3 + 2x_4 + 0.5x_5 + 3x_6 + x_7 + 0.7x_8 + 3x_9 + e$$

donde: x_i con $i = 1, \dots, 5$ son linealmente independientes y de distribución normal.

x_i con $i = 6, \dots, 9$ son formadas por la combinación lineal de las otras variables.

e : Error aleatorio con distribución normal.

Los datos se dividen en dos grupos: cincuenta para entrenar el modelo y cincuenta para validar el modelo, considerando como criterios para la validación la suma de los cuadrados de los errores (SSE), el promedio de la suma de los cuadrados de los errores (PSSE) y el coeficiente de regresión lineal entre los y dados y los y estimados (R).

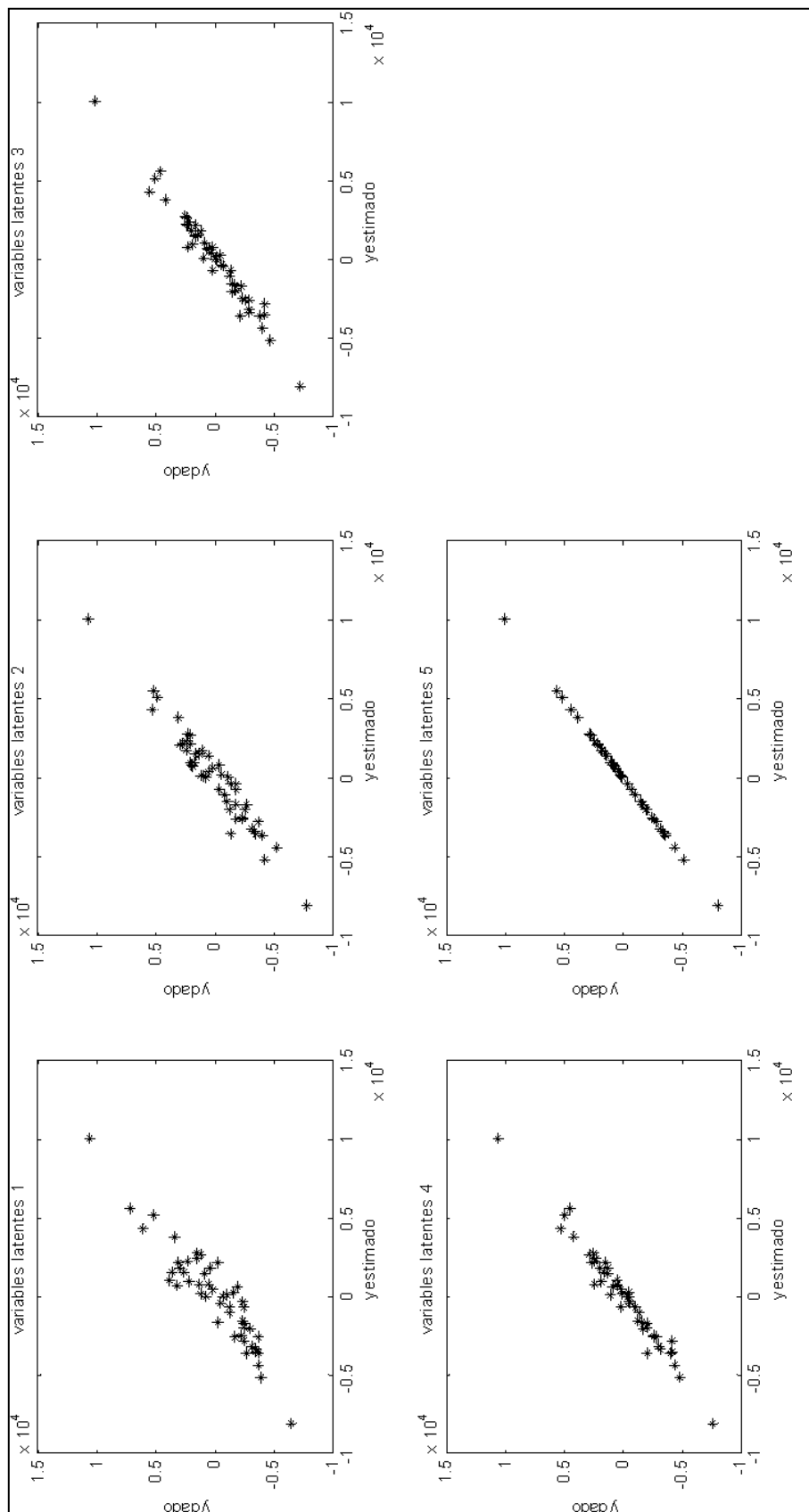
Tabla 1: Resultados obtenidos con los datos de validación.

No. De variables latentes.	SSE	PSSE	R
1	7,6837e+007	1,5367e+006	0,8337
2	3,1328e+007	6,2655e+005	0,9322
3	1,6898e+007	3,3795e+005	0,9634
4	1,6161e+07	3,2322e+005	0,9650
5	2,8904e-022	5,7808e-024	1
6	2,8904e-022	5,7808e-024	1
7	2,8904e-022	5,7808e-024	1
8	2,8904e-022	5,7808e-024	1
9	2,8904e-022	5,7808e-024	1

En la tabla 1, se muestran los resultados del modelo obtenidos con el algoritmo del PLS con el método de mayor pendiente para los datos de validación, estos se muestran con una (1) hasta nueve (9) variables latentes llegando a obtener con cinco (5) variables latentes el mejor modelo que ajusta el problema de regresión arrojando un PSSE de 5,7808e-024 y un SSE de 2,8904e-022. A partir de seis variables latentes el modelo no presenta mejoría como era de esperarse ya que el ejemplo experimental fue creado con cinco variables independientes y el resto tenía dependencia lineal con las anteriores. También puede observarse a través de los valores de R que existe en todos los casos una relación lineal entre la variable de salida dada y la variable de salida estimada por el modelo, presentándose siempre una relación positiva que se acerca y llega a uno a medida que el modelo mejora (ver figura 4).

De esta manera, se verifica lo expuesto en la teoría, donde se demostró que puede usarse el método de optimización de mayor pendiente para encontrar los vectores proyección w y c del método PLS tradicional.

Figura 4: representación de R para los datos de la validación



Referencias bibliográficas

1. Geladi P. and Kowalski B. (1986). Partial Least-Squares Regression: A Tutorial. Elsevier Science Publishers B V., p 1-17.
2. Geladi P. and Kowalski B. (1986). An Example of 2-Block Predictive Partial Least-Squares Regression with Simulated Data. Elsevier Science Publishers B V., p. 19-32.
3. Hoskuldsson A. (1998). PLS Regression Methods. *Journal of Chemometrics*. Vol. 2, p. 211-228.
4. Wold S., Sjöström M. and Eriksson L. (2001). PLS-Regressions: a Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58, p. 109-130.
5. Krämer N., Boulesteix A. and Tutz G. (2008). Penalized Partial Least Squares with Application to B-Spline Transformation and Functional Data, p. 1-26. {EN LINEA} Disponible en http://www.wias-berlin.de/people/kraemer/papers/chemo_ppls.pdf.
6. Boulesteix A. and Strimmer K. (2006). Partial Least Squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, Vol. 8, No. 1, p. 32-44.
7. Gonzalez J., Peña D. and Romera R. (2007). A Robust Partial Least Squares Method with Applications. Working paper 07-13. *Statistic and Econometric Series* 04, p. 1-20. {EN LINEA} Disponible en <http://e-archivo.uc3m.es/bitstream/10016/665/1/ws071304.pdf>.
8. Rodríguez E. (2005). Desarrollo Matemático del Modelo de Regresión No Lineal Kernel Cuadrados Mínimos Parciales. Trabajo de Ascenso para optar a la categoría de profesor asociado en la Universidad del Zulia, p. 94.
9. Bennett K. and Embrechts M. (2003). An Optimization Perspective on Kernel Partial Least Squares Regression. Chapter 11 in *Advances in Learning Theory: Methods, Models and Applications*, Suykens J. A. K. et al., Eds., NATO-ASI Series in Computer and System Sciences, IOS Press Amsterdam, The Netherlands.
10. Burden R. and Faires J. (1985). *Análisis Numérico*. México D.F. Grupo Editorial Iberoamérica, p. 721.