

Modelado de problemas de regresión lineal con el método Cuasi-Newton

Eddy Jackeline Rodríguez

Centro de Investigación de Matemática Aplicada (CIMA). Facultad de Ingeniería.
Universidad del Zulia. Maracaibo. Venezuela.

email: eddyjackeline@yahoo.es

Recibido: 26-05-2015 Aceptado: 09-06-2015

Resumen

El método de Newton es ampliamente utilizado en las áreas afines a las matemáticas, en especial; en situaciones que involucran problemas de optimización donde el objetivo es encontrar el mínimo de una función; sin embargo; en su expresión pura, tiene como limitante el cálculo de la inversa de la matriz de los datos, es por esto, que se ha desarrollado modificaciones al método de Newton, denominándose método Cuasi-Newton. En este trabajo se aplica el método Cuasi-Newton para modelar problemas de regresión lineal, considerando como función objetivo la minimización del error para el ajuste del modelo mediante los coeficientes de regresión. Se aplica específicamente el método propuesto por Davidon-Fletcher-Powell, por tener éste la cualidad de heredar la definición positiva, lo que facilita el cálculo de la inversa de una matriz. La metodología empleada en esta investigación es teórica-práctica donde la revisión teórica parte de la consulta de autores especialistas tales como: Boulesteix A., Strimmer K. [1], Geladi P., Kowalski B. [2], Luenberger, D. [3], Hoskuldsson A. [4], Rodríguez, E. [5]. Los resultados de los ejemplos creados experimentalmente y de los reales demuestran que el modelo de problemas de regresión lineal aplicando el método cuasi Newton es eficiente.

Palabras clave: Método Cuasi-Newton, modelado de problemas de regresión, coeficientes de regresión.

Modeling linear regression problems with the Quasi-Newton method

Abstract

Newton's method is widely used in the related areas of mathematics, especially; in situations that involve optimization problems where the objective is to find the minimum of a function, but in its purest form, is limiting the calculation of the inverse array data, which is why, it has developed modifications to Newton's method, Quasi-Newton methods being called. In this paper, the Quasi-Newton method is applied to model linear regression problems, considering the objective function minimization of error for model fit by regression coefficients. Applies specifically proposed by Davidon -Fletcher -Powell method, by having it inherit the quality of positive definition, which facilitates the calculation of the inverse of a matrix. The methodology used in this research is theoretical and practical where the theoretical review of the consultation of specialists authors such as Boulesteix A., Strimmer K. [1], Geladi P., Kowalski B. [2] Luenberger, D. [3] A. Hoskuldsson [4] Rodríguez, E. [5]. The results of the examples set experimentally, and actual model show that the problems of linear regression using the quasi Newton method is efficient.

Key words: Quasi-Newton method, regression modeling problems, regression coefficients.

Introducción

Mejorar la calidad de un producto, reducir los riesgos en una planta, minimizar los costos de producción son, entre otros ejemplos, motivos del desarrollo de herramientas de optimización para los complejos problemas de operación, estructura y diseño en una planta. Para dar respuesta a esto, se crean diferentes modelos matemáticos que logran su objetivo con la ayuda de los ordenadores, con su creciente capacidad de cálculo y con el desarrollo de software que proporcionan las herramientas adecuadas para el uso de los modelos matemáticos.

En la actualidad existen diferentes técnicas que permiten ajustar de manera satisfactoria un problema de regresión, desde el más sencillo como es el de mínimos cuadrados ordinarios, Mardia et al [6], hasta los más complejos que pueden solventar los obstáculos de colinealidad, tamaño de los datos, etcétera, tales como son las técnicas de: los factores latentes fortalecidos, Bühlmann, Hothorn [7], los mínimos cuadrados parciales, Boulesteix, Strimmer [1], Geladi, Kowalski [2] y la regresión por componentes principales, Cuadras [8], entre otros. En este trabajo se desarrolla un modelo de problema de regresión lineal¹ aplicando el método Cuasi-Newton, que también supera los obstáculos ya mencionados. Este modelo parte de la construcción de una función de costo que se traduce en minimizar el error del ajuste, transformando éste en un problema de optimización sin restricciones, alcanzando la solución a través del desarrollo de un algoritmo iterativo.

Para la investigación se utilizó una metodología teórica-experimental partiendo de la consulta en fuentes bibliográficas y sitios web, presentando primeramente una visión general de la regresión lineal múltiple basado en los mínimos cuadrados ordinarios, seguido del desarrollo del método Cuasi-Newton, la formulación del método Cuasi-Newton como modelo de regresión lineal y la obtención de los coeficientes de regresión. Por último se evalúa el algoritmo de este modelo con ejemplos experimentales y real comprobándose, a través de los resultados, que modela de manera eficiente.

Regresión lineal múltiple

La regresión lineal múltiple consiste en determinar una combinación lineal entre las variables que participan en el problema a modelar. Esta combinación se realiza entre un conjunto de variables de entrada (predictores) y la variable de salida (respuesta). La relación entre estas variables se formula como un modelo lineal:

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} \quad (i = 1, 2, \dots, m) \quad (1)$$

donde $\beta_0, \beta_1, \dots, \beta_n$ son los coeficientes de la combinación lineal, llamados coeficientes de regresión. El modelo de regresión lineal de un problema se encuentra minimizando la discrepancia, es decir, minimizando la varianza residual entre el valor observado (real) y el valor estimado por el modelo:

$$\min \sum u_i^2 = \min \sum (y_i - \hat{y}_i)^2 \quad (2)$$

u_i es un error aleatorio con distribución normal que mide la discrepancia, debido a la aproximación, en la i -ésima observación.

Usando notación matricial: $U = Y - XB$

$$U^2 = U^T U = (Y - XB)^T (Y - XB)$$

¹ A los fines del presente artículo se ha preferido utilizar la denominación "modelo de problema de regresión lineal", pues la misma expresa la construcción de un modelo a un problema de regresión lineal; aunque una parte de la doctrina usa la denominación "modelo de regresión lineal".

X : Matriz de datos de las variables de entrada, de tamaño $m \times n$.

Y : Vector de variable de salida de tamaño m .

U : Vector error de tamaño m .

B : Vector de coeficientes de regresión de tamaño n .

Quedando la varianza residual como una función que depende de la matriz B así:

$$f(B) = U^T U = (Y - XB)^T (Y - XB) \quad (3)$$

la cual encuentra su mínimo cuando se cumple: $\frac{\partial f}{\partial B} = 0$

Obteniendo los coeficientes de regresión:

$$B = (X^T X)^{-1} X^T Y \quad (4)$$

siempre que $X^T X$ sea una matriz no singular.

El método descrito en esta sección para obtener los coeficientes de regresión es el de los mínimos cuadrados ordinarios (MCO) que tiene como limitante el comportamiento de la matriz $X^T X$. Es importante observar que si las variables de entrada están fuertemente correlacionadas entre sí o puede escribirse alguna como combinación lineal de las restantes, la matriz $X^T X$ tendrá como determinante un valor cercano a cero o será cero, lo que producirá una matriz singular o inestable en la solución del estimador, ocasionando un aumento en su varianza. Para superar estos obstáculos en la actualidad existen métodos tales como el de análisis de componentes principales, mínimos cuadrados parciales y factores latentes fortalecidos entre otros que se basan en la construcción de espacios que contengan sólo las variables necesarias para predecir el modelo de estudio Bühlmann, Hothorn [7], Cuadras, [8], Hoskuldsson [4], Krämer et al [9].

Más adelante, se propone y desarrolla la construcción de un modelo de regresión que obtiene los coeficientes de regresión usando métodos de optimización donde el comportamiento de la matriz $X^T X$ no representa un obstáculo.

Métodos Cuasi-Newton

Los algoritmos para la optimización no lineal sin restricciones con múltiples variables se pueden clasificar como: búsqueda sin el uso de derivadas, búsqueda con la información de la primera derivada (gradiente) y búsqueda con la información de la segunda derivada. Entre los algoritmos que usan el gradiente se encuentran los métodos Cuasi-Newton los cuales son una modificación del método de Newton, en los que se construye una aproximación de la curvatura de la función no lineal utilizando sólo información de la primera derivada, evitando la construcción de la matriz de segundas derivadas (Hessiana).

El método de Newton se basa en que una función f se aproxima a una función cuadrática para ser minimizada; para ello, cerca del punto de estudio, la función se aproxima por la serie de Taylor truncada:

$$f(x) \simeq f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k) \quad (5)$$

El lado derecho se minimiza en:

$$x_{k+1} = x_k - (H(x_k))^{-1} \nabla f(x_k)^T \quad (6)$$

Considerando las condiciones de suficiencia de segundo orden para un punto mínimo, se supone que para un punto mínimo relativo x^* , la matriz Hessiana $H(x^*)$ es definida positiva. De manera que si f tiene segundas derivadas parciales continuas, $H(x^*)$ es definida positiva cerca de x^* , por lo que el método está bien definido cerca de la solución Luenberger [3].

Es posible que el punto generado con el método de Newton no converja, debido a que la matriz Hessiana puede ser singular. Sin embargo, si el punto de inicio es cercano al punto óptimo, tal que, el gradiente en este punto es igual a cero y la matriz Hessiana evaluada en este punto es de rango completo, entonces el método de Newton está bien definido y converge al punto óptimo. Esto se explica en Luenberger [3] con el teorema: sea $f \in C^3$ en E^n , y supóngase que en el punto mínimo local x^* , el Hessiano $H(x^*)$ es definido positivo. Entonces, si se comienza suficientemente cerca de x^* , los puntos generados por el método de Newton convergen a x^* . El orden de convergencia es al menos dos.

El método de Newton puede ser modificado de manera de garantizar la convergencia desde cualquier punto de inicio. Dado un x_k , considere la dirección $d = -R\nabla f(x_k)^T$, donde $R = (\epsilon I + H)^{-1}$ es una matriz simétrica definida positiva con $H = H(x_k)$ y $\epsilon \geq 0$. La sucesión de puntos se calcula a través de $x_{k+1} = x_k + \lambda d$, donde λ es la solución óptima del problema de minimización $f(x_k + \lambda d)$ sujeto a $\lambda \geq 0$. El escalar $\epsilon \geq 0$ puede calcularse de la siguiente manera: Fije un $\delta > 0$, de manera que $\epsilon \geq 0$ sea el escalar más pequeño que haga todos los valores propios de la matriz $(\epsilon I + H)$ mayores o iguales a δ . Esta matriz es definida positiva, todos sus valores propios son positivos y tiene inversa. En particular la matriz R es también definida positiva. Ya que los valores propios de una matriz dependen continuamente de sus elementos, ϵ es una función continua de x y de aquí el mapeo punto a punto $D: E^n \rightarrow E^n$ definido por $D(x) = (x, d)$ es continuo. Esto es el algoritmo mapea $A = MD$, donde M es la búsqueda lineal usual $\{\lambda: \lambda \geq 0\}$.

Sea $\Omega = \{x^*: \nabla f(x^*) = 0\}$ y $x \notin \Omega$, mientras R sea definida positiva, $d \neq 0$ entonces, M es cerrado en (x, d) , ya que D es una función continua entonces A es cerrado sobre el complemento de Ω . Además, como $\nabla f(x)^T d = \nabla f(x)^T R \nabla f(x)^T < 0$, R es definida positiva y $\nabla f(x) \neq 0$, entonces, d es una dirección de descenso de f en x . Defínase a $y \in A(x)$ de manera que $f(y) < f(x)$, entonces f es una función de descenso y asumiendo que la secuencia generada por el algoritmo está contenida en un conjunto compacto, se sigue que el algoritmo converge. Esto demuestra que si el menor valor propio de la matriz H es mayor o igual que un δ , entonces el conjunto de puntos $\{x_k\}$ generado por el algoritmo alcanza el x^* . Es de observar que si $\epsilon_k = 0$, entonces $d_k = -H(x_k)^{-1} \nabla f(x_k)^T$, lo que nos lleva al algoritmo del método de Newton sin modificación.

La utilidad de este algoritmo presenta como obstáculo el cálculo de los valores propios de la matriz Hessiana. En la práctica se puede recurrir al procedimiento de Levenberg- Marquardt, o puede usarse la factorización de Cholesky. También existe los métodos Cuasi-Newton, que forman una aproximación a la inversa del Hessiano, cuya forma puede variar desde las más sencillas, donde la matriz permanece fija durante el proceso iterativo, hasta los más avanzados donde se construyen aproximaciones mejoradas basadas en la información del proceso.

Construcción de la Matriz Inversa

En los métodos Cuasi-Newton la inversa de la matriz Hessiana (H) o una aproximación de ella se logran por la información almacenada durante el desarrollo del proceso de descenso. Idealmente la aproximación converge a la inversa de H en el punto solución. Sea $f \in E^n$ con segundas derivadas parciales continuas, si para dos puntos x_{k+1}, x_k , se define: $p_k = x_{k+1} - x_k$ como una dirección, entonces: $\nabla f(x_{k+1}) - \nabla f(x_k) \cong H(x_k)p_k$

Si H es constante, queda: $q_k \equiv \nabla f(x_{k+1}) - \nabla f(x_k) = H p_k$

De manera que la evaluación del gradiente en dos puntos suministra información sobre H . Si se determinan “n” direcciones linealmente independientes p_k con $k = 0, 1, \dots, n - 1$, de manera que se puede definir a P y Q como matrices $n \times n$ con columnas p_k y q_k respectivamente entonces: $H = QP^{-1}$.

Si se define a F como la aproximación a la inversa del Hessiano, de manera que: $F_{k+1} q_i = p_i$ $0 \leq i \leq k$, luego de “n” pasos linealmente independientes se tendrá $F_n = H^{-1}$.

Sin embargo con $k < n$ hay infinidad de soluciones y se tendría que tener presente algunas consideraciones adicionales, tales como:

Corrección de rango uno

Como la matriz Hessiana y su inversa son simétricas, es normal exigir que la aproximación a su inversa sea simétrica y se puede definir una recursión que conserva la simetría:

$$F_{k+1} = F_k + a_k z_k z_k^T$$

donde a_k es una constante y z_k es un vector.

El segundo término de la ecuación define una matriz de rango uno con la que se actualiza la aproximación a la inversa: $p_k = F_{k+1} q_k = F_k q_k + a_k z_k z_k^T q_k$

Pre-multiplicando esta ecuación por q_k^T queda: $q_k^T p_k - q_k^T F_k q_k = a_k (z_k^T q_k)^2$ lo que lleva a:

$$F_{k+1} = F_k + \frac{(p_k - F_k q_k)(p_k - F_k q_k)^T}{q_k^T (p_k - F_k q_k)} \quad (7)$$

Luego, aplicando el método de descenso se calcula la dirección: $d_k = -F_k \nabla f_k$

Se minimiza $f(x_k + a d_k)$ respecto $a \geq 0$ y se determinan $x_{k+1} = x_k + a_k d_k, p_k = a_k d_k, \nabla f(x_{k+1})$.

Esta forma de obtener la inversa tiene por limitación que se garantice la definición positiva de (7), lo cual sucede si $q_k^T (p_k - F_k q_k) > 0$, y por implicaciones numéricas es recomendable que no sea muy cercano a cero.

Corrección de rango dos o propuesta de Davidon-Fletcher-Powell (DFP)

La inversa de la matriz Hessiana es actualizada en cada paso con la suma de matrices de rango uno. Se inicia con cualquier matriz $F(x_0)$ definida positiva simétrica, con cualquier punto x_0 , y se calculan:

La dirección $d_k = -F_k \nabla f(x_k)$

Se minimiza $f(x_k + a d_k)$ respecto a $a \geq 0$.

Se determinan $x_{k+1} = x_k + a_k d_k, p_k = a_k d_k, \nabla f(x_{k+1})$

$$\text{Se calcula } F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k} \quad (8)$$

Una propiedad interesante de esta propuesta es que hereda la definición positiva, es decir: si F_0 es definida positiva y si el paso usado por el método de DFP tiene la propiedad $p_k^T q_k > 0$, entonces F_k es definida positiva $\forall k$. (Luenberger, 1989). Esta propiedad es muy importante en los métodos Cuasi-Newton ya que obliga al algoritmo a tener siempre una dirección de descenso.

Familia de Broyden

Considera actualizaciones en la misma matriz Hessiana, en lugar de su inversa. Se denota la k -ésima aproximación de H por B_k y se busca satisfacer a: $q_i = B_{k+1} p_i \quad 0 \leq i \leq k$

$$\text{con } B_{k+1} = B_k + \frac{q_k q_k^T}{q_k^T p_k} - \frac{B_k p_k p_k^T B_k}{p_k^T B_k p_k} \quad (9)$$

Esta ecuación (9) es conocida como la actualización de Broyde-Fletcher-Goldfarb-Shanno.

Métodos Cuasi-Newton como modelo de regresión lineal

Como se vio previamente un modelo a un problema de regresión lineal se construye a partir de la premisa: $\min \sum u_i^2 = \min \sum (y_i - \hat{y}_i)^2$, lo cual puede verse como un problema de optimización sin restricciones.

El objetivo es encontrar el punto óptimo B que minimice la función: $f(B) = U^T U = (Y - XB)^T (Y - XB)$

Para ello en este trabajo se aplica y se desarrolla un modelo usando los algoritmos de los métodos Cuasi-Newton, específicamente el que corresponde a la propuesta de DFP, por sus propiedades

$$\min f(B) = \min[(Y - XB)^T (Y - XB)] = \min (Y^T Y - 2B^T X^T Y + B^T X^T XB)$$

Como $Y^T Y$ no depende de la variable B , puede escribirse $k = Y^T Y$. Además, no se altera la minimización de la función $f(B)$ si se divide entre 2.

$$\min \left(\frac{k}{2} - B^T X^T Y + \frac{1}{2} B^T X^T XB \right) \quad (10)$$

Esta nueva función para $f(B)$ puede verse como la aproximación de Taylor que hace el método de Newton en la ecuación (5), donde el gradiente de la función es $\nabla f(B) = X^T XB - X^T Y$ y la matriz Hessiana es $H(B) = X^T X$.

El mínimo se obtiene en: $B_{k+1} = B_k - (H(B_k))^{-1} \nabla f(B_k)^T$

Sustituyendo el gradiente y la matriz Hessiana:

$$B_{k+1} = B_k - (X^T X)^{-1} (X^T XB_k - X^T Y)^T \quad (11)$$

En este caso, para asegurar la convergencia en el punto generado en (11), debido a que la matriz $(X^T X)^{-1}$ puede ser singular, se encuentra el mínimo mediante una modificación al método de Newton, aplicando la propuesta de Davidon-Fletcher-Powell para encontrar la matriz inversa:

$$F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$

con, $p_k = -a_k F_k (X^T XB_k - X^T Y)$, $q_k = (X^T XB_{k+1} - X^T Y) - (X^T XB_k - X^T Y)$
 F es la aproximación a la inversa de la matriz Hessiana $(X^T X)^{-1}$.

Algoritmo de un modelo de regresión lineal con método Cuasi-Newton

Dado: B_0, F_0 (definida positiva simétrica).

Para $k = 0$ hasta n . Calcúlese:

$$\nabla f(B) = X^T XB - X^T Y$$

$d(B_k) = F_k \nabla f(XB_k)$ (dirección de descenso)

Se minimiza $f(B_k + ad(B_k))$ respecto $a \geq 0$ y se determinan:

$$B_{k+1} = B_k + a_k d(B_k)$$

$$p_k = a_k d(B_k), \nabla f(B_{k+1})$$

$$q_k \equiv \nabla f(B_{k+1}) - \nabla f(B_k)$$

$$F_{k+1} = F_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{F_k q_k q_k^T F_k}{q_k^T F_k q_k}$$

Coefficientes de Regresión con mínimos cuadrados ordinarios y Método Cuasi-Newton:

En la tabla 1, se presenta un cuadro comparativo de cómo se obtiene los coeficientes de regresión aplicando el método de mínimos cuadrados ordinarios y el método Cuasi-Newton:

Tabla 1. Obtención de los Coeficientes de Regresión (B).

Mínimos Cuadrados Ordinarios	Cuasi-Newton
$B = (X^T X)^{-1} X^T Y$	Proceso iterativo: $B_{k+1} = B_k - (X^T X)^{-1} (X^T X B_k - Y^T X)^T$ $B_{k+1} = B_k - (X^T X)^{-1} (B_k^T X^T X - X^T Y)$ $B_{k+1} = (X^T X)^{-1} X^T Y$

Se puede observar que con ambos métodos se llega a la misma fórmula para obtener los coeficientes de regresión; sin embargo, como se explicó anteriormente, el cálculo de la inversa de la matriz $(X^T X)$ falla cuando las variables de entrada presentan colinealidad, es en estos casos que el método Cuasi-Newton representa una buena alternativa ya que la inversa de la matrices actualizada en cada paso con la suma de matrices de rango uno (método de DFP) durante el proceso de iteraciones, garantizando su existencia.

Resultados y conclusiones

En este trabajo se proponen cinco ejemplos de modelado de problemas de regresión lineal que evidencian la eficiencia del algoritmo del método propuesto en esta investigación. Cuatro de los ejemplos son creados experimentalmente y uno proviene de datos reales.

Ejemplo 1: está formado de ocho variables de entrada con distribución normal (X) y una variable de salida (Y) con 100 observaciones de cada una, todas las variables de entrada son independientes. La variable respuesta se construye a partir de: $Y = XB$, con $B = (1,3,5,7,8,9,4,12)$.

Ejemplo 2: está formado de ocho variables de entrada con distribución normal (X), una variable de salida (Y) y un error aleatorio normal (u) con 100 observaciones de cada una, todas las variables de entrada son independientes. La variable respuesta se construye a partir de: $Y = XB + u$, con $B = (1,3,5,7,8,9,4,12)$,

Ejemplo 3: está formado de ocho variables de entrada con distribución normal (X) y una variable de salida (Y) con 100 observaciones de cada una, hay cuatro variables de entrada independientes X_i , con $i: 1,2,3,4$ y el resto son generadas: $X_5 = 2X_1, X_6 = X_2 + X_3, X_7 = 3X_1 + 2X_2 + X_3, X_8 = 4X_4$. La variable respuesta se construye a partir de: $Y = XB$, con $B = (1,3,5,7,8,9,4,12)$.

Ejemplo 4: está formado de nueve variables de entrada (X), una variable de salida (Y) y un error aleatorio normal (u) con 100 observaciones de cada una, como se plantea en Rodríguez [5], $Y = 5X_1 + X_2 + 7X_3 + 2X_4 + 0.5X_5 + 3X_6 + X_7 + 0.7X_8 + 3X_9 + u$ donde X_i con $i = 1, \dots, 5$ son linealmente independientes y de distribución normal y las $i = 6, \dots, 9$ son formadas por la combinación lineal de las otras variables.

Ejemplo 5: es un conjunto de datos que comprende intensidades espectrales de 60 muestras de gasolina en 401 longitudes de onda, medidos en intervalos de 2 nm de 900 nm a 1700 nm, y sus índices de octano (octanaje). Estos datos se describen en Kalivas, John H., "dos conjuntos de datos de espectros de infrarrojo cercano", Quimiometría y Laboratorio de Sistemas Inteligentes, V.37 (1997), pp.255-259.

Los datos de cada ejemplo son divididos en dos grupos, uno es usado para entrenar el algoritmo y el otro grupo para la validación del mismo.

La eficiencia del método, aquí propuesto, se comprueba a través de los criterios: Suma de los cuadrados de los errores (SSE), que refleja la variación de los datos observados alrededor de la línea de regresión. Coeficiente de determinación múltiple (R^2), que indica que proporción de la variación total de la respuesta Y es explicada por el modelo ajustado. Promedio de la suma de los cuadrados de los errores (PSSE), que muestra en promedio la diferencia de los datos observados respecto a los datos obtenidos por el modelo ajustado.

Los resultados de los ejemplos se presentan en la tabla 2, y en esta se puede observar que en base a los criterios de evaluación SSE, R^2 y PSSE el algoritmo del modelo de regresión lineal con método Cuasi-Newton, arroja excelentes resultados en los casos de los ejemplos 1 y 3, donde no hay presencia de error aleatorio normal, el ajuste con los datos de validación es de $2.8783e-012$ y $3.8275e-017$ para la suma de los cuadrados de los errores y de $5.7566e-014$ y $7.6551e-019$ para el promedio de la suma de los cuadrados de los errores respectivamente lo que indica que el error cometido es insignificante. Para los ejemplos 2 y 4, donde hay presencia de error aleatorio normal, es de notar que el error sigue siendo bajo. En el caso del ejemplo 4, se consideró el propuesto en Rodríguez [5], con el objeto de comparar la eficiencia del algoritmo planteado en este trabajo con el algoritmo PLS que es un método reconocido para el modelaje de problemas de regresión lineal cuando el conjunto de variables de entrada están correlacionadas. Los resultados se expresan en la tabla 3 y, es de notar que el algoritmo modelo de regresión lineal con método Cuasi-Newton ajusta los valores de la validación de manera satisfactoria.

Tabla 2. Resultados obtenidos con los datos de validación.

EJEMPLOS	SSE	R^2	PSSE
Ejemplo 1: las variables de entrada son independientes. No hay ruido.	$2.8783e-012$	1.0000	$5.7566e-014$
Ejemplo 2: las variables de entrada son independientes. Hay ruido normal.	$4.6132e-005$	1.0000	$9.2263e-007$
Ejemplo 3: las variables de entrada son dependientes. No hay ruido.	$3.8275e-017$	1	$7.6551e-019$
Ejemplo 4: las variables de entrada son dependientes. Hay ruido normal.	$9.2930e-016$	1	$1.8586e-017$
Ejemplo 5: datos reales.	3.4274	0.9459	0.1142

Tabla 3. comparación de los resultados del ejemplo 4.

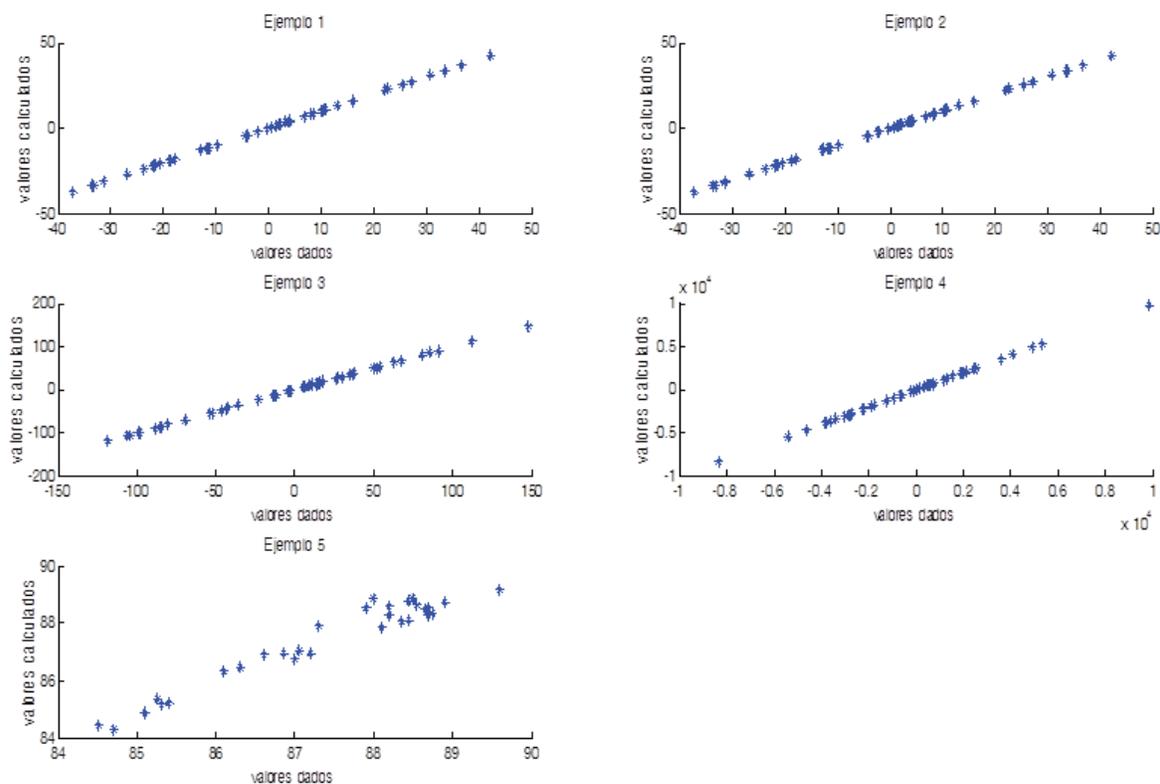
Método	SSE	R^2	PSSE
PLS ¹	$2,8904e-022$	1	$5,7808e-024$
Cuasi-Newton	$9.2930e-016$	1	$1.8586e-017$

Para el ejemplo 5, donde se trabaja con datos reales, el PSSE es de tan solo 0.1142 y la SSE es de 3.4274 con una relación lineal entre los valores dados y calculados de 0.9459, muy cercano a 1, es decir, que el 94,59% de la variación de los datos observados se explica mediante el modelo. Lo que demuestra que el ajuste del modelo es satisfactorio.

En la figura 1, se observan cinco gráficas, que corresponden a los ejemplos del 1 al 5, donde se representan la relación lineal entre la variable de salida dada y la variable de salida estimada por el modelo, aplicado al conjunto de validación; en todas ellas se destaca que existe una relación lineal de pendiente positiva de valor uno o muy cercana a uno (R^2).

Se puede concluir en base al desarrollo teórico y verificación experimental que el modelado de problemas de regresión lineal usando el método de optimización Cuasi-Newton es eficiente. Con este método se logra solventar los problemas de colinealidad entre las variables de entrada, como se verifica en los ejemplos 3 y 4, de igual manera se comprueba lo robusto del algoritmo planteado en esta investigación ante la presencia de ruido aleatorio normal, como se ve en los ejemplos 2, 4 y 5.

Figura 1. Representación de los valores dados versus los valores calculados.



² Algoritmo de mínimos cuadrados parciales (PLS) a partir del método de mayor pendiente. Rodríguez [5].

Referencias bibliográficas

1. Boulesteix A., 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.
2. Geladi P., Kowalski B. (1986). *Partial Least- Squares Regresión: A Tutorial*. Elsevier Science Publishers B V. p 1-17
3. Luenberger D. E. (1989). *Programación Lineal y no Lineal*. Addison – Wesley Iberoamericana, S. A. p. 499.
4. Hoskuldsson A. (1998). *PLS Regression Methods*. *Journal of Chemometrics*. Vol. 2. p. 211-228
5. Rodríguez, E. (2012). *Mínimos Cuadrados Parciales con el Descenso de Mayor Pendiente*. Maracai- bo, Venezuela. Fondo Editorial Biblioteca Universidad Rafael Urdaneta. N° 3.
6. Mardia K., Kent J., & Bibby J. *Multivariate analysis*. Academic Press. (1979).
7. Bühlmann P., Hothorn T. (2007). *Boosting Algorithms: Regularization, Prediction and Model Fitting*. *Statistical Science*. Vol. 22 No 4. p. 477- 505. Recuperado de <http://www.cs.princeton.edu/~schapire/boost.html>.
8. Cuadras C.M. (1991). *Métodos de Análisis Multivariante*. Promociones y Publicaciones Universita- rias, S. A. p. 1115
9. Krämer N., Boulesteix A., & Tutz G. (2008). *Penalized Partial Least Squares with Application to B-Spline Transformation and Functional Data*. p. 1-26. Recuperado de http://www.wias-berlin.de/people/kraemer/papers/chemo_pppls.pdf.