



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelos de Regresión Multivariado Aplicado en Variables Educativas

Jaider Albeiro Figueroa Flórez

Universidad Nacional de Colombia
Facultad de Ciencias Exactas y Naturales
Maestría en Ciencias Matemáticas Aplicadas
Manizales - Caldas, Colombia
2014



UNIVERSIDAD NACIONAL DE COLOMBIA

Multivariate Regression Models Applied in Educational Variables

Jaider Albeiro Figueroa Flórez

Universidad Nacional de Colombia
Facultad de Ciencias Exactas y Naturales
Maestría en Ciencias Matemáticas Aplicadas
Manizales - Caldas, Colombia
2014

Modelos de Regresión Multivariado Aplicado en Variables Educativas

Jaider Albeiro Figueroa Flórez

Tesis presentada como requisito parcial para optar al título de:
Magister en Ciencias - Matemática Aplicada

Director:

Ph.D. Carlos Daniel Acosta Medina

Grupo de Investigación:

Modelamiento Matemático y Computación Científica
cdacostam@unal.edu.co

Codirector:

Germán Albeiro Castaño Duque

Grupo de Investigación:

Fortalecimiento y Desarrollo de los Grupos Cultura Organizacional, Gestión Humana y
Cultura en la Calidad de la Educación
gacastanod@unal.edu.co

Universidad Nacional de Colombia
Facultad de Ciencias Exactas y Naturales
Maestría en Ciencias Matemáticas Aplicadas
Manizales - Caldas, Colombia
2014

Dedicatoria

Dedico la realización de este trabajo a Dios, como una forma de cualificación personal para seguir aportando en la enseñanza del mundo de las Matemáticas.

Agradecimientos

Agradezco a Dios por darme la oportunidad de luchar y alcanzar otro de mis sueños. A mi familia por su apoyo incondicional. A todos mis amigos que me animaron a seguir escalando en conocimiento.

Índice general

1. Preliminares y Fundamentación	13
1.1. La Técnica PLS y el Algoritmo NIPALS	13
1.2. PLS Regression (PLSR)	14
1.2.1. Regresión PLS1 (PLS1R)	15
1.2.2. Regresión PLS2(PLS2R)	17
1.3. PLS No Lineal	18
1.3.1. Kernel PLS Regression (KPLSR)	19
2. Metodología	23
2.1. Elección de variables y recopilación de datos	23
2.2. Adecuación de la Base de Datos	26
2.3. Construcción y Estructuración de los Modelos	26
2.4. Implementación, Ajuste y Validación de los Modelos	26
3. Experimentación y Resultados	29
3.1. Construcción de Bases de Datos y Modelos Lineales	29
3.1.1. Administración de Empresas (N)	29
3.1.2. Administración de Empresas (D)	38
3.1.3. Administración en Sistemas Informáticos	39
3.1.4. Arquitectura	40
3.1.5. Ingeniería Civil	40
3.1.6. Ingeniería Eléctrica	41
3.1.7. Ingeniería Electrónica	42
3.1.8. Ingeniería Física	43
3.1.9. Ingeniería Industrial	43
3.1.10. Ingeniería Química	44
3.1.11. Matemáticas	45
3.2. Construcción de Modelos de Regresión Multivariado Usando técnicas No lineales	46
3.2.1. Administración de Empresas (N)	46
3.2.2. Administración de Empresas (D)	49
3.2.3. Administración en Sistemas Informáticos	49
3.2.4. Arquitectura	50
3.2.5. Ingeniería Civil	50
3.2.6. Ingeniería Eléctrica	50
3.2.7. Ingeniería Electrónica	51

3.2.8. Ingeniería Física	51
3.2.9. Ingeniería Industrial	51
3.2.10. Ingeniería Química	52
3.2.11. Matemáticas	52
3.2.12. Una aplicación de la técnica KPLS en una función real no lineal, $N =$ $L = 1$	52
4. Discusión y Conclusiones	55
4.1. Sobre los modelos de regresión basados en las técnicas PLS1 y LS	55
4.2. Sobre los modelos de regresión usando la técnica PLS2	57
4.3. Sobre los modelos de regresión usando la técnica PLS2 Linealizado	59
4.4. Sobre los modelos de regresión no lineales aplicando la técnica KPLS (KPLSR)	59

Resumen

Se realiza un estudio sobre el comportamiento de variables educativas asociadas a indicadores de alta calidad en los programas de pregrado de la Universidad Nacional de Colombia - Sede Manizales, con el objeto de crear modelos de regresión multivariado que permitan proyectar comportamientos futuros y establecer relaciones entre las variables asociadas a los factores estudiantes, docencia, procesos académicos e investigación y las relacionadas con garantía, reconocimiento y aseguramiento de la calidad. Para la construcción y estructuración de los modelos se utiliza la técnica PLSR y Kernel PLSR, haciendo los ajustes y validaciones pertinentes. Se obtienen modelos que contribuyen al mejoramiento de los aspectos predictivo y explicativo conjuntamente, obteniendo información interesante para efectos de toma de decisiones en el ámbito académico y administrativo.

Palabras Claves: Regresión en mínimos cuadrados parciales, Modelos explicativos, Modelos predictivos, Deflactor, Validación cruzada, Porcentaje de predicción, Métodos Kernel.

Abstract

We study the behavior of educational variables associated to high-quality indicators of the undergraduate programs offered by the Universidad Nacional de Colombia – Sede Manizales, with aims at elaborating multivariate regression models that enable us to plan future behavior, and establish relations between the variables associated to the factors students, teaching, academic processes and research, and the ones related to guarantee, recognition and assurance of quality. We use the PLS and Kernel PLS techniques, with the appropriate adjustments and validations. We obtain models that contribute to the improvement of predicting and explicative aspects jointly, obtaining relevant information to be taken into account when making academic or administrative decisions.

Keywords: Partial least square regression, Explanatory models, Predictive models, Deflate, Cross-validation, Prediction percentage, Kernel methods.

Introducción

Los modelos de regresión multivariado(MRM) son muy usados en estudios de investigación en cualquier campo de las ciencias (Exactas, Naturales, Sociales,) cuando estos ponen en consideración dos tipos o conjunto de variables (las denominadas explicativas, regresoras o de entrada y las clasificadas como dependientes, de respuesta o salida) y se desea establecer una relación matemática entre ellas. El investigador busca con estos modelos básicamente dos objetivos, el primero es poder estimar el grado de influencia de las variables de entrada en el comportamiento o variabilidad de las variables de salida (Modelos Explicativos) y el segundo con el animo de predecir comportamientos a largo plazo en las variables de respuesta una vez se tenga información sobre las variables explicativas (Modelos Predictivos).

Existen muchas técnicas para la construcción de MRM con la intención de mejorar los aspectos explicativos y predictivos, entre ellas destacamos: La técnica de Mínimos Cuadrados (LS y OLS), Análisis en Componentes Principales(PCA), Análisis basado en sistemas de ecuaciones estructurales y otras más recientes como: La técnica Mínimos Cuadrados Parciales (PLS) y PLS- Path Modeling; para el caso no lineal encontramos las técnicas: Vectores de maquinas de soporte (SVM), Procesos Gaussianos y Métodos Splines, métodos Kernel, entre otros.[11, 17]

En este trabajo se presenta la construcción, estructuración, implementación, ajustes y validación de MRM aplicando las técnicas PLS y KPLS, en un estudio sobre el comportamiento de variables educativas asociadas a indicadores de calidad en los programas de pregrado de la Universidad Nacional - Sede Manizales, se eligen aquellos modelos que expliquen de mejor manera la relación entre las variables estudiantes, docencia, procesos académicos e investigación y las relacionadas con garantía, reconocimiento y aseguramiento de la calidad educativa en educación superior. Trabajo que surge por la necesidad de priorizar mejoramiento en los factores de más impacto en calidad y tomar decisiones adecuadas para el fortalecimiento de procesos académicos, administrativos y financieros en los programas evaluados.

Existen varias razones por las cuales se decide trabajar con los modelos de Regresión PLS (PLS-R) y Regresión KPLS (KPLS-R), entre ellas destacamos [7, 10, 11, 15, 17]:

- Trabajar con base de datos de las que se desconoce el tipo de distribución probabilística que tienen asociada sus variables.
- Permitir trabajar con base o bloques de datos en las que el número de variables es mayor que el número de observaciones.

- Eliminar el problema de multicolinealidad entre las variables explicativas y entre las variables de respuesta, a partir de la construcción de variables latentes ortogonales.
- Las bondades de la técnica PLS para establecer relaciones explicativas entre las variables de entrada y las de respuesta, gracias al criterio de optimización que utiliza (máxima covarianza entre las variables latentes de entrada y las latentes de salida).
- El ofrecer un mejoramiento en capacidad de predicción de los modelos sobre aquellas variables de respuesta cuya variabilidad se aleja del comportamiento lineal (Caso KPLS).

A continuación se exponen las teorías más relevantes que han permitido el desarrollo, validez y avances de la técnica PLS, los modelos PLSR y KPLSR. En el capítulo 2 se explica la metodología usada en los procesos de elección de variables, recopilación y adecuación de base de datos. El capítulo 3 describe las fases de experimentación y resultados (construcción, estructuración implementación, ajustes y validación de los modelos y los algoritmos usados). En el capítulo 4 se presenta una discusión sobre los resultados obtenidos y las conclusiones.

Capítulo 1

Preliminares y Fundamentación

1.1. La Técnica PLS y el Algoritmo NIPALS

La técnica PLS fue desarrollada por el suizo Helman Ole Andreas Wold en el año 1966, como una alternativa para eliminar el problema de multicolinealidad. La técnica permite calcular componentes principales mediante secuencias iterativas de regresiones simples usando el método OLS (mínimos cuadrados ordinarios) y además calcular correlaciones canónicas con secuencias iterativas de regresiones múltiples usando OLS [11]. De aquí que los resultados obtenidos con PLS para bases de datos completas son iguales a los obtenidos con la técnica PCA, la ventaja está cuando la base de datos es incompleta o cuando su aplicación pasa al plano de la regresión.

Considere X la matriz de datos de tamaño $(n \times N)$ centrada y rango a , la idea básica es descomponer la matriz X en la forma $X = TP^T + E$ donde T está formada por columnas de vectores latentes o componentes ortogonales (scores) y es de tamaño $(n \times a)$, P la matriz de vectores de peso (loadings) de tamaño $(N \times a)$ y E es la matriz de residuales.

Así :

$$X = [X_{*1}|X_{*2}|X_{*3}|\dots|X_{*N}] = T_{*1}P_{*1}^T + T_{*2}P_{*2}^T + \dots + T_{*a}P_{*a}^T \quad (1.1)$$

Donde T_{*k} y P_{*k}^T son las columnas de las matrices T y P^T respectivamente. $k = 1, \dots, a$.

El primer problema que resuelve PLS es el de multicolinealidad entre las columnas de variables que componen X , entonces cada componente scores o variable latente se construye de la forma $T_{*k} = Xw$, donde w es un vector adecuado de pesos y al deflactar X se garantiza que las $T_{*k}'s$ sean ortogonales [12].

Consideremos una primera aproximación de X con su primera componente principal ($k = 1$), de (1.1) observamos que la columna j -ésima de X , tiene la forma: $X_{*j} = p_{1j}T_{*1}$, de donde

$$p_{1j} = X_{*j}^T T_{*1} \quad (1.2)$$

Una nueva aproximación de X tomando dos componentes principales, nos da que la

columna j -ésima de X , tiene la forma: $X_{*j} = p_{1j}T_{*1} + p_{2j}T_{*2}$ ó bien $X_{*j} - p_{1j}T_{*1} = p_{2j}T_{*2}$, así:

$$p_{2j} = (X_{*j} - p_{1j}T_{*1})^T T_{*2} \quad (1.3)$$

De modo que para la componente k -ésima, la columna j -ésima de X , tiene la forma:

$$X_{*j} - \sum_{l=1}^{k-1} p_{1j}T_{*l} = p_{kj}T_{*k}$$

de donde

$$p_{kj} = (X_{*j} - \sum_{l=1}^{k-1} p_{1j}T_{*l})^T T_{*k} \quad (1.4)$$

Las características de los T_{*k} 's, y la consecución de los p_{k*} por medio de (1.4) a medida que se consideran nuevas componentes scores, permite construir las matrices T y P , obteniendo una aproximación de la matriz original X en la forma $X \approx TP^T$.

El comportamiento descrito en los líneas anteriores constituyen la base para la construcción del algoritmo NIPALS (Non-linear Iterative Partial Least Squares) (Wold, 1973) y el desarrollo de los algoritmos PLS-R.

Algoritmo NIPALS

1. $X_0 = X_h$, X centrada o estandarizada.
2. Para $h = 1, \dots, a$ (a es el rango de X).
 - 2.1 t_h : Inicial¹
 - 2.2 $p_h = X_{h-1}^T t_h / (t_h^T t_h)$
 - 2.3 Normar p_h a 1
 - 2.4 $t_h = X_{h-1} p_h$
 - 2.5 Deflactar X , $X_h = X_{h-1} - t_h p_h^T$
 - 2.6 Repetir pasos 2.2 a 2.5 hasta convergencia

1.2. PLS Regression (PLSR)

En este caso consideremos dos bloques o matrices de variables X y Y , donde X es la matriz de variables de entrada o explicativas de tamaño $(n \times N)$ y Y la matriz de variables de salida o respuesta de tamaño $(n \times L)$, consideremos además el problema de multicolinealidad.

¹Se recomienda iniciar tomando el primer t como la primera columna de X_0 [12]. Otros autores toman t aleatorio [14, 17].

El método PLS sugiere construir componentes ortogonales (vectores latentes, vectores scores) en X y Y de la forma $\mathbf{t} = X\mathbf{w}$ y $\mathbf{u} = Y\mathbf{c}$, respectivamente, donde \mathbf{w} y \mathbf{c} son vectores de peso adecuados de norma 1. Cada vector columna \mathbf{t} genera la matriz T , de manera análoga cada vector columna \mathbf{u} genera la matriz U , permitiendo la descomposición de X y Y , en la forma[14, 15]:

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \quad (1.5)$$

Donde T y U son matrices de tamaño $(n \times a)$ cuyas columnas son las componentes ortogonales anteriormente descritas. La matriz P de tamaño $(N \times a)$ y Q de tamaño $(L \times a)$ representan las matrices de pesos(loadings) y las matrices E y F de tamaño $(n \times N)$ y $(n \times L)$ respectivamente, son matrices residuales.

La gran virtud del método PLS es considerar el problema explicativo entre X y Y , a partir de las nuevas variables latentes representativas $\mathbf{t} = X\mathbf{w}$ y $\mathbf{u} = Y\mathbf{c}$. La técnica considera un modelo de regresión adecuado aquel que además de hacer reducción de dimensionalidad garantice la relación entre estas nuevas variables, por lo que se propone maximizar el cuadrado de la covarianza entre ellas, es decir, se propone resolver el problema[12, 19]:

$$\max_{\|\mathbf{w}\|=\|\mathbf{c}\|=1} [\text{cov}(X\mathbf{w}, Y\mathbf{c})]^2 \quad (1.6)$$

Encontrado los \mathbf{w} y \mathbf{c} adecuados, que formarán las matrices W y C y generando las matrices T y U obtenemos la matriz de coeficientes B , de tal modo que el modelo de regresión PLS toma la forma:

$$Y = XB + \varepsilon \quad (1.7)$$

1.2.1. Regresión PLS1 (PLS1R)

Consideremos el caso univariado ($L = 1$), es decir, una sola variable de salida Y . En este caso se busca obtener un modelo de la forma:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_N X_N + \epsilon \quad (1.8)$$

en las que X_1, \dots, X_N pueden estar altamente correlacionadas. Construimos entonces las componentes ortogonales $\mathbf{t} = X\mathbf{w}$, realizando en el nuevo espacio de variables regresiones de la forma:

$$Y = c_1 \mathbf{t}_1 + c_2 \mathbf{t}_2 + c_3 \mathbf{t}_3 + \cdots + c_a \mathbf{t}_a + \delta \quad (1.9)$$

Para resolver (1.6), consideremos \mathbf{v} el vector de covarianzas entre X y Y , es decir, $\mathbf{v} = X^T Y$, con X y Y centradas o estandarizadas, entonces :

$$[\text{cov}^2(X\mathbf{w}, Y)] = [\mathbf{w}^T \text{cov}(X, Y)]^2 = [\mathbf{w}^T \mathbf{v}]^2 = \mathbf{w}^T \mathbf{v} \mathbf{v}^T \mathbf{w} \quad (1.10)$$

La función lagrangiana que maximiza (1.10) sujeta a $\|\mathbf{w}\| = 1$, es:

$$\phi(\mathbf{w}, \lambda) = [\mathbf{w}^T \mathbf{v} \mathbf{v}^T \mathbf{w}] - \lambda(\mathbf{w}^T \mathbf{w} - 1)$$

Derivando respecto a \mathbf{w} e igualando a cero tenemos:

$$\frac{\partial \phi}{\partial \mathbf{w}} = 2\mathbf{v}\mathbf{v}^T \mathbf{w} - 2\lambda \mathbf{w} = 0$$

Por tanto,

$$\mathbf{v}\mathbf{v}^T \mathbf{w} = \lambda \mathbf{w} \quad (1.11)$$

quedando un problema de valores y vectores propios, donde λ y \mathbf{w} son el autovalor y auto-vector de $\mathbf{v}\mathbf{v}^T$ respectivamente.

Premultiplicando (1.11) por \mathbf{w}^T e igualmente por \mathbf{v}^T , se deduce de λ que el \mathbf{w} buscado tiene la forma $\mathbf{w} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{X^T y}{\|X^T y\|}$ y corresponde al vector de covarianzas normalizado[12]. Este resultado se aprovecha en el desarrollo del algoritmo iterativo para PLSR, tomando como base el algoritmo NIPALS en la construcción de las componentes scores.

Existen muchas versiones de estos algoritmos, a continuación se presenta una de ellas derivada de la versión SIMCA-P²

Algoritmo PLS1R

Usaremos en este caso \mathbf{y} en vez de Y .

1. $X_0 = X, \mathbf{y}_0 = \mathbf{y}$, X e \mathbf{y} centradas o estandarizadas.
2. Para $h = 1, \dots, a$ (a es el rango de X).
 - 2.1 $w_h = X_{h-1}^T \mathbf{y}_{h-1} / \|X_{h-1}^T \mathbf{y}_{h-1}\|$
 - 2.2 $t_h = X_{h-1} w_h$, componente h de X
 - 2.3 $p_h = X_{h-1}^T t_h / (t_h^T t_h)$
 - 2.4 $X_h = X_{h-1} - t_h p_h^T$, Deflactamos X
 - 2.5 $c_h = \mathbf{y}_{h-1}^T t_h / (t_h^T t_h)$
 - 2.6 $u_h = \mathbf{y}_{h-1} / c_h$, componente h de y
 - 2.7 $\mathbf{y}_h = \mathbf{y}_{h-1} - t_h c_h$, Deflactamos \mathbf{y}
 - 2.8 end h

De (1.9), la formula de regresión para \mathbf{y} queda :

$$\mathbf{y} \approx \hat{\mathbf{y}} = c_1 \mathbf{t}_1 + c_2 \mathbf{t}_2 + c_3 \mathbf{t}_3 + \dots + c_N \mathbf{t}_N \quad (1.12)$$

pero por construcción de los t_h y su relación con los w_h , p_h y los c_h , podemos expresar $\hat{\mathbf{y}}$ en terminos de las variables originales, asi:

$$\hat{\mathbf{y}} = XW(P^T W)^{-1} \mathbf{c} = XB \quad (1.13)$$

donde $B = W(P^T W)^{-1} \mathbf{c}$, en este caso es el vector de los coeficientes de regresión PLS de \mathbf{y} sobre X utilizando h componentes [12, 14, 17].

²SIMCA-P (The standard in Multivariate Data Analysis): Es un software desarrollado por la compañía Umetrics que trabaja basicamente con métodos PCA y PLSR.

1.2.2. Regresión PLS2(PLS2R)

Consideremos ahora el caso $L > 1$, salida múltiple. X y Y centradas o estandarizadas

De las relaciones expuestas en (1.5), se puede mostrar que (1.6) sujeto a $\|\mathbf{w}\| = 1$ y $\|\mathbf{c}\| = 1$ genera el problema de autovalores y autovectores para \mathbf{w} [12, 14]

$$X^T Y Y^T X \mathbf{w} = \lambda \mathbf{w} \quad (1.14)$$

Con \mathbf{w} óptimo de la forma $\mathbf{w} = X^T \mathbf{u}$. De manera análoga, por sustituciones sucesivas se pueden obtener para \mathbf{t} , \mathbf{c} y \mathbf{u} , los problemas (Tenenhaus, 1998):

$$X X^T Y Y^T \mathbf{t} = \lambda \mathbf{t} \quad (1.15)$$

$$Y^T X X^T Y \mathbf{c} = \lambda \mathbf{c} \quad (1.16)$$

$$Y Y^T X X^T \mathbf{u} = \lambda \mathbf{u} \quad (1.17)$$

Aprovechando la capacidad del algoritmo NIPALS para encontrar en forma iterativa estas componentes y resolver los problemas de autovalores, generalizamos el algoritmo de PLS1R en la forma:

Algoritmo PLS2R

Entradas: X , Y centradas o estandarizadas.

1. Iniciamos con \mathbf{u} aleatorio³.
2. $\mathbf{w} = X^T \mathbf{u}$
3. $\mathbf{t} = X \mathbf{w}$, $\mathbf{t} \leftarrow \mathbf{t} / \|\mathbf{t}\|$
4. $\mathbf{c} = Y^T \mathbf{t}$
5. $\mathbf{u} = Y \mathbf{c}$, $\mathbf{u} \leftarrow \mathbf{u} / \|\mathbf{u}\|$
6. Repetir pasos 2 a 5 hasta convergencia.
7. Deflactar X , Y : $X \leftarrow X - \mathbf{t} \mathbf{t}^T X$, $Y \leftarrow Y - \mathbf{t} \mathbf{t}^T Y$

El modelo de regresión PLS descrito en (1.7), puede expresarse en la forma [12, 14, 15]

$$Y \approx \hat{Y} = X B$$

Donde $B = W(P^T W)^{-1} C^T$, es la matriz de coeficientes de regresión. La matriz $P^T W$ es triangular superior y por tanto invertible.

³Recomendable iniciar con una combinación lineal de las columnas de Y , es la primera componente principal de Y .

Usando el hecho que $\mathbf{t}^T \mathbf{t} = 0, \forall i \neq j$ y $\mathbf{t}^T \mathbf{u} = 0, \forall j > i$, derivamos las siguientes igualdades

$$\begin{aligned} W &= X^T U \\ P &= X^T T (T^T T)^{-1} \\ C &= Y^T T (T^T T)^{-1} \end{aligned}$$

Haciendo sustituciones en B de la expresión (1.7), podemos escribirla B nuevamente en la forma [12, 14, 15]

$$B = X^T U (T^T X X^T U)^{-1} T^T Y \quad (1.18)$$

de donde

$$Y \approx \hat{Y} = X B = T T^T Y = T C^T \quad (1.19)$$

1.3. PLS No Lineal

Hay ocasiones en que los modelos lineales PLS ($Y = X B + \epsilon$), quedan cortos para explicar y predecir los comportamientos de Y con relación a los cambios en X , entonces debemos recurrir a modelos no lineales de la forma $Y = f(X) + \epsilon$ con f bien definida en X , que mejoren estos inconvenientes.

Siguiendo las líneas de la técnica PLS, podemos asumir dos maneras de modelar relaciones de datos no lineales, a partir, de la técnica PLSR [14]:

1. La primera estrategia consiste en considerar la reformulación de la relación lineal $U = T D + H$ (entre los vectores o componentes scores \mathbf{t} de X y las componentes \mathbf{u} de Y)⁴ dada en PLSR, por la relación no lineal

$$\mathbf{u} = f(\mathbf{t}) + \epsilon = f(X, \mathbf{w}) + \epsilon \quad (1.20)$$

$f(\cdot)$ es una función continua que modela relaciones no lineales. Funciones polinomiales, splines y otros métodos de suavizado se han usado en la construcción de $f(\cdot)$.

2. Una segunda estrategia consiste en usar métodos Kernel en espacios de Hilbert Reprodutor del Kernel (RKHS) (N. Aronszajn, 1950, B. Shölkopf, 2002), aprovechando que los espacios donde se mueven los bloques de variables originales de entrada y salida son subconjuntos de \mathbb{R}^N y \mathbb{R}^L respectivamente. La idea básica es asumir una transformación no lineal de las variables de entrada $\{\mathbf{x}\}_{i=1}^n$ en un espacio de características \mathcal{F} , esto es, considerar una aplicación

$$\Phi : \mathbf{x}_i \in \mathbb{R}^N \rightarrow \Phi(\mathbf{x}_i) \in \mathcal{F} \quad (1.21)$$

y luego construir un modelo PLSR en \mathcal{F} [17].

Uno de los fundamentos técnicos para aplicar métodos kernel surge del tratamiento que se da a la técnica PLSR cuando tenemos más observaciones que variables ($n \gg N$),

⁴Recuerde que esta relación lineal es esencial en PLS2R para la determinación de B descrito en (1.18)

donde la deflacción no se hace sobre X sino sobre la matriz XX^T de tamaño $(n \times n)$, que es entre otras una matriz kernel \mathbf{K} definida por $\mathbf{K} = k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ [17]. Este comportamiento da pie para generalizar la aplicación de PLSR a cualquier matriz kernel \mathbf{K} definida en \mathcal{F} .

A continuación se describe con más detalle las características, bondades y construcción de este tipo de modelos.

1.3.1. Kernel PLS Regression (KPLSR)

Los recientes desarrollos de la teoría o de los métodos kernel, también han sido aplicados en la técnica PLS. La metodología Kernel PLS se propuso para establecer relaciones entre conjunto de variables de las cuales se pretenden hacer estudios, en problemas de regresión y clasificación [14].

Tal como se comentó en (1.20), la técnica KPLSR se basa en mapeos del espacio original de datos en un espacio de alta dimensión \mathcal{X} , luego aplicando el truco kernel la estimación de PLS en el espacio \mathcal{F} se reduce a cálculos de álgebra lineal tan simples como en PLS lineal [14]. Como es sabido el truco kernel permite conocer el producto punto entre dos elementos $\Phi(\mathbf{x})$, $\Phi(\mathbf{y})$ en \mathcal{F} y construir la matriz kernel \mathbf{K} sin necesidad de conocer quien es en realidad Φ , esto es (N. Aronszajn, 1950, B. Shölkopf, 2002)

$$k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X} \quad (1.22)$$

Se define así la gran matriz \mathbf{K} de productos puntos entre todos los mapeos de los puntos de datos $\mathbf{K} = \Phi\Phi^T$, donde Φ denota la matriz de mapeo de los elementos o datos del espacio \mathcal{X} , $\{\Phi(\mathbf{x}_i) \in \mathcal{F}\}_{i=1}^n$.

Motivados por la teoría de RKHS y lo descrito anteriormente, se deriva el algoritmo para KPLS como una modificación del algoritmo para PLS2R en los pasos 2 y 3 [14, 17]

Algoritmo KPLS

Entradas: \mathbf{K} , Y centradas.

1. Iniciamos con \mathbf{u} aleatorio.
2. $\mathbf{t} = \Phi\Phi^T \mathbf{u} = \mathbf{K}\mathbf{u}$, $\mathbf{t} \leftarrow \mathbf{t}/\|\mathbf{t}\|$
3. $\mathbf{c} = Y^T \mathbf{t}$
4. $\mathbf{u} = Y\mathbf{c}$, $\mathbf{u} \leftarrow \mathbf{u}/\|\mathbf{u}\|$
5. Repetir pasos 2 a 4 hasta convergencia.
6. Deflactar $\Phi\Phi^T = \mathbf{K}$: $\Phi\Phi^T \leftarrow (\Phi - \mathbf{t}\mathbf{t}^T\Phi)(\Phi - \mathbf{t}\mathbf{t}^T\Phi)^T$, $Y \leftarrow Y - \mathbf{t}\mathbf{t}^T Y$ ó bien

$$\mathbf{K} \leftarrow (\mathbf{I}_n - \mathbf{t}\mathbf{t}^T)\mathbf{K}(\mathbf{I}_n - \mathbf{t}\mathbf{t}^T) = \mathbf{K} - \mathbf{t}\mathbf{t}^T\mathbf{K} - \mathbf{K}\mathbf{t}\mathbf{t}^T + \mathbf{t}\mathbf{t}^T\mathbf{K}\mathbf{t}\mathbf{t}^T \quad (1.23)$$

Para centrar la matriz \mathbf{K} que contiene el mapeo de datos en el espacio \mathcal{F} , se aplica el procedimiento (schölkopf,1998, Wu, 1997)

$$\mathbf{K} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{K}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T) \quad (1.24)$$

De manera análoga a lo expuesto en PLS2R, la matriz de coeficientes de regresión B descrita en (1.18), para el caso KPLS tiene la forma[17]

$$B = \Phi^T U (T^T \mathbf{K} U)^{-1} T^T Y \quad (1.25)$$

y la predicción sobre los datos de entrenamiento, esta dada por

$$\begin{aligned} \hat{Y} &= \Phi B \\ &= \Phi \Phi^T U (T^T \mathbf{K} U)^{-1} T^T Y \\ &= \mathbf{K} U (T^T \mathbf{K} U)^{-1} T^T Y = \mathbf{K} \mathbf{D} \\ &= T T^T Y \end{aligned} \quad (1.26)$$

Las últimas igualdades se tienen de: $\mathbf{D} = U (T^T \mathbf{K} U)^{-1} T^T Y$, y $T = \Phi R$, donde $R = \Phi^T U (T^T \mathbf{K} U)^{-1}$ (de Jong, 1993, Helland, 1998).

Para hacer predicción sobre los puntos de prueba $\{\mathbf{x}_i\}_{i=n+1}^{n+n_t}$, se utiliza la matriz de coeficientes de regresión B dada en (1.25), de modo que la aproximación de las predicciones queda[17]

$$\hat{Y}_t = \Phi_t B = \mathbf{K}_t U (T^T \mathbf{K} U)^{-1} T^T Y \quad (1.27)$$

Sobre la interpretación de los modelos KPLS (caso $L=1$), podemos recurrir al teorema de representación en un RKHS \mathcal{H} .

Considere un conjunto de entrenamiento $\{\mathbf{x}_i\}_{i=1}^n$ un subconjunto de un conjunto compacto $\mathcal{X} \subset \mathbb{R}^N$, $\{\mathbf{y}_i\}_{i=1}^n \in \mathbb{R}$ el conjunto de salidas correspondientes. La solución al problema⁵

$$\min_{[f \in \mathcal{H}] R_{reg}(f)} = \frac{1}{n} \sum_{i=1}^n V(\mathbf{y}_i, f(\mathbf{x}_i)) + \xi \|f\|_{\mathcal{H}}^2 \quad (1.28)$$

$\xi \in \mathbb{Z}^+$ (coeficiente de regularización) y $\|\cdot\|^2$ la norma definida en \mathcal{H} (estabilizador), esta dada por el teorema

Representer Theorem (caso simple): Considere el funcional $V(\mathbf{y}_i, f)$ el cual depende de f mediante los $\{f(\mathbf{x}_i)\}_{i=1}^n$ (valores de f sobre el conjunto de datos). Entonces cualquier solución al problema (27) tiene una representación de la forma

$$f(\mathbf{x}) = \sum_{i=1}^n c_i K(\mathbf{x}_i, \mathbf{x}) \quad (1.29)$$

⁵Esta solución fue dada por Kimeldorf y Wahba(1971), Wahba(1999)

Donde $\{c_i\}_{i=1}^n \in \mathbb{R}$.

Basado en (1.28) podemos entonces representar la solución de la regresión KPLS como [15,17,19]

$$f(\mathbf{x}, \mathbf{d}) = \sum_{i=1}^n d_i K(\mathbf{x}, \mathbf{x}_i) \quad (1.30)$$

donde $\mathbf{d} = U(T^T \mathbf{K}U)^{-1}T^T Y$, es un vector $n \times 1$.

Podemos también interpretar los modelos KPLS como un modelo de regresión lineal de la forma (Garthwaite, 1994, Höskuldsson, 1998)

$$f(\mathbf{x}, \mathbf{c}) = c_1 \mathbf{t}_1(\mathbf{x}) + c_2 \mathbf{t}_2(\mathbf{x}) + \dots + c_a \mathbf{t}_a(\mathbf{x}) = \mathbf{c}^T \mathbf{t}(\mathbf{x}) \quad (1.31)$$

donde los $\{\mathbf{t}_i(\mathbf{x})\}_{i=1}^a$ son la proyecciones de \mathbf{x} en las a componentes extraídas y \mathbf{c} es el vector de pesos descrito como C en PLS2R.

Capítulo 2

Metodología

Se describe a continuación la manera en que se abordaron los temas de: Elección de variables y recopilación de datos, Adecuación de la Base de Datos, Construcción y Estructuración de los Modelos, e Implementación, Ajuste y Validación de los Modelos.

2.1. Elección de variables y recopilación de datos

Se hizo una revisión bibliográfica sobre variables educativas asociadas a procesos de acreditación de alta calidad para programas de pregrado, se seleccionaron aquellas consideradas de alto impacto y medibles en forma cuantitativa. Luego se inició el proceso de recolección de datos sobre estas variables seleccionadas en las diferentes dependencias de la Universidad Nacional - sede Manizales. Los datos recopilados corresponden a los periodos académicos 2009-I a 2012-II de los programas: Administración de Empresas Nocturna, Administración de Empresas Diurna, Admon en sistemas informáticos, Arquitectura, Ingeniería civil, Ingeniería Eléctrica, Ingeniería Electrónica, Ingeniería Física, Ingeniería Industrial, Ingeniería Química y Matemáticas.

La elección de las variables de entrada y de salida se realizaron fundamentados en los Lineamientos para la acreditación de programas académicos de pregrado en Colombia establecidos por el CNA [1, 2, 3, 4, 5], bajo la premisa de tener en cuenta como principios de calidad **factores iniciales o propios** como: Condiciones académicas, disponibilidad de recursos físicos y financieros y pertinencia social y profesional , y **factores de garantía, reconocimiento y aseguramiento de la calidad** como: Desempeño de egresados, Producción intelectual de docentes, reconocimiento de la sociedad, Impacto obtenido en el medio, calidad de procesos pedagógicos, entre otros.

En este sentido, se decide trabajar por programa con las siguientes variables de entrada¹

¹Para la recolección de datos relacionados con estas variables se recurrieron a las siguientes fuentes: Plataforma SARA, Sistema de Información de Evaluación Educativa (ICFES), Estadísticas Unal - <http://portal.manizales.unal.edu.co/planeacion/index.php/estadisticas>, Información suministrada por las Dependencias: Planeación, DIMA, Facultades, ORI, CEUNAL, Bienestar universitario, Oficina de registro y matrícula y Dirección Académica.

Para los programas de Admon de Empresas (D), Admon de empresas (N):

- x_{*1} = Tasa de Absorción.
- x_{*2} = Número de estudiantes matriculados provenientes de Instituciones públicas.
- x_{*3} = Número de estudiantes matriculados provenientes de Instituciones privadas.
- x_{*4} = Resultado en prueba de admisión de los estudiantes.
- x_{*5} = Número de Estudiantes que reciben apoyo institucional (prestamos, alojamientos, alimentación)
- x_{*6} = Número de Estudiantes en intercambio con universidades nacionales o extranjeras.
- x_{*7} = Número de estudiantes realizan prestamos en biblioteca
- x_{*8} = Número de estudiantes que acceden a consulta en las bases de datos Sinab.
- x_{*9} = Número de Estudiantes vinculados como monitores.
- x_{*10} = Número de Estudiantes que reservan cupo.
- x_{*11} = Número de Estudiantes que cancelan semestre.
- x_{*12} = Número de docentes con título de maestría y doctorado vinculados al programa
- x_{*13} = Número de docentes con título de doctorado vinculados al programa.
- x_{*14} = Número de docentes con dedicación exclusiva.
- x_{*15} = Número de convenios del programa establecidos con el sector productivo para efectos de prácticas.
- x_{*16} = Número de semilleros de investigación en el programa.
- x_{*17} = Número de estudiantes vinculados a semilleros de investigación
- x_{*18} = Número de docentes que lideran semilleros de investigación
- x_{*19} = Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.
- x_{*20} = Productividad Académica de docentes.

Para el programa de Sistemas Informáticos:

- x_{*14} = Número de docentes con dedicación medio tiempo y TC.

Para el programa de Arquitectura:

- x_{*14} = Número de docentes con dedicación medio tiempo y TC.
- x_{*15} = Número de docentes con dedicación exclusiva.

Para el programa de Ingeniería Civil:

- x_{*14} = Número de docentes con dedicación medio tiempo y TC.
- x_{*15} = Número de docentes con dedicación exclusiva.
- x_{*18} = Número de docentes que lideran semilleros de investigación & Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.

Para los programas de Ing. Eléctrica, Ing. Electrónica:

- x_{*6} = Número de estudiantes realizan prestamos en biblioteca
- x_{*7} = Número de estudiantes que acceden a consulta en las bases de datos Sinab.
- x_{*8} = Número de Estudiantes vinculados como monitores.
- x_{*9} = Número de Estudiantes que reservan cupo.

- x_{*10} = Número de Estudiantes que cancelan semestre.
- x_{*11} = Número de docentes con título de maestría y doctorado vinculados al programa
- x_{*12} = Número de docentes con título de doctorado vinculados al programa
- x_{*13} = Número de docentes con dedicación de medio tiempo y tiempo completo.
- x_{*14} = Número de docentes con dedicación exclusiva.
- x_{*15} = Número de semilleros de investigación en el programa.
- x_{*16} = Número de estudiantes vinculados a semilleros de investigación.
- x_{*17} = Número de docentes que lideran semilleros de investigación.
- x_{*18} = Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.
- x_{*19} = Productividad Académica de docentes.

Para el programa de Ing. Física:

- x_{*6} = Número de estudiantes realizan prestamos en biblioteca
- x_{*7} = Número de estudiantes que acceden a consulta en las bases de datos Sinab.
- x_{*8} = Número de Estudiantes que reservan cupo.
- x_{*9} = Número de Estudiantes que cancelan semestre.
- x_{*10} = Número de docentes con título de maestría y doctorado vinculados al programa
- x_{*11} = Número de docentes con título de doctorado vinculados al programa
- x_{*12} = Número de docentes con dedicación de medio tiempo y tiempo completo.
- x_{*13} = Número de docentes con dedicación exclusiva.
- x_{*14} = Número de semilleros de investigación en el programa.
- x_{*15} = Número de estudiantes vinculados a semilleros de investigación.
- x_{*16} = Número de docentes que lideran semilleros de investigación.
- x_{*17} = Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.
- x_{*18} = Productividad Académica de docentes.

Para el programa de Ingeniería Industrial:

- x_{*14} = Número de docentes con dedicación de medio tiempo y tiempo completo.
- x_{*15} = Número de docentes con dedicación exclusiva.

Para el programa de Ingeniería Química:

- x_{*15} = Número de semilleros de investigación en el programa.
- x_{*16} = Número de estudiantes vinculados a semilleros de investigación.
- x_{*17} = Número de docentes que lideran semilleros de investigación.
- x_{*18} = Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.
- x_{*19} = Productividad Académica de docentes.

Para el programa de Matemáticas:

- x_{*6} = Número de estudiantes realizan prestamos en biblioteca
- x_{*7} = Número de estudiantes que acceden a consulta en las bases de datos Sinab.
- x_{*8} = Número de Estudiantes vinculados como monitores.

- x_{*9} = Número de Estudiantes que reservan cupo.
- x_{*10} = Número de Estudiantes que cancelan semestre.
- x_{*11} = Número de docentes con título de maestría y doctorado vinculados al programa
- x_{*12} = Número de docentes con título de doctorado vinculados al programa
- x_{*13} = Número de docentes con dedicación de medio tiempo y tiempo completo.
- x_{*14} = Número de docentes con dedicación exclusiva.
- x_{*15} = Número de semilleros de investigación en el programa & Número de docentes que lideran semilleros de investigación.
- x_{*16} = Número de estudiantes vinculados a semilleros de investigación.
- x_{*17} = Número de grupos de investigación reconocidos & Número de docentes que lideran grupos de investigación en el programa.
- x_{*18} = Productividad Académica de docentes.

Las variables de respuesta para todos los programas son:

- y_{*1} = Resultados de estudiantes en pruebas saber pro.
- y_{*2} = Promedio académico estudiantil.
- y_{*3} = Número de estudiantes que desertan del programa.

2.2. Adecuación de la Base de Datos

Se realizan ajustes sobre los datos y las observaciones dependiendo de las problemáticas presentadas en la recolección inicial: Ajustes de escala y cambio de representación numérica, duplicación de datos y eliminación de datos faltantes. De modo que para todos los programas se lograron adecuar 8 observaciones, es decir, se estudiarán los periodos comprendidos desde 2009I a 2012II.

2.3. Construcción y Estructuración de los Modelos

Para el caso $L = 1$, para cada programa se presentan modelos iniciales de regresión lineal múltiple y se comparan capacidad de predicción, los dos modelos presentados corresponden a las técnicas de mínimos cuadrados (LS) y PLS1R.

Para el caso $L > 1$, En cada programa se presentan modelos multivariados de regresión lineal (técnica PLS2R) y no lineal (KPLSR)

Se usan algoritmos de Matlab (como plsregression) y las versiones NIPALS construidos por los autores en Matlab.

2.4. Implementación, Ajuste y Validación de los Modelos

Se implementan los modelos descritos anteriormente para cada caso y se realizan procesos de validación usando el método de validación cruzada dejando uno por fuera (Cross Validation

- Leave one out (LOO)), se seleccionan los de mejor comportamiento predictivo y explicativos.

Para el caso salida múltiple se ajustan aquellos modelos cuya comportamiento predictivo no es el deseado, a partir, de técnicas de linealización y se comparan dichos resultados con los modelos iniciales. Además se presentan modelos de regresión no lineales (KPLSR) que mejoran los procesos predictivos y explicativos.

Capítulo 3

Experimentación y Resultados

3.1. Construcción de Bases de Datos y Modelos Lineales

Las bases de datos usadas para la construcción de los modelos en cada programa se exponen en **ANEXOS**.

Se presentan los resultados obtenidos en la implementación, ajuste y validación de los modelos para cada programa académico:

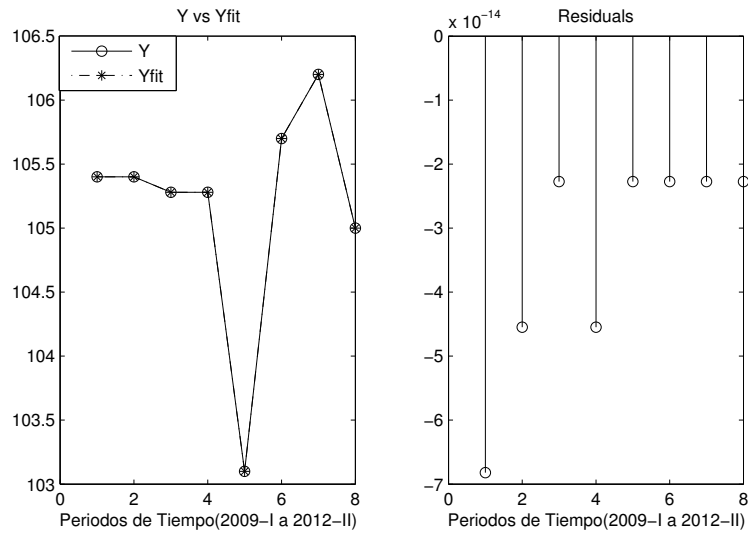
3.1.1. Administración de Empresas (N)

Los modelos con salida única tienen en cuenta como variable de salida (Resultados en pruebas SABER-PRO y los de salida múltiple las variables (Resultados en pruebas SABER-PRO, Promedio Académico, Deserción).

Modelo LS

Aplicando el método de mínimos cuadrados se obtiene el modelo:

$$\hat{Y} = \begin{bmatrix} -6,28E-02 \\ 1,63E-01 \\ -3,85E-02 \\ -3,83E-02 \\ -7,06E-02 \\ 9,51E-02 \\ -5,12E-02 \\ 9,02E-02 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_7 \\ x_8 \\ x_{13} \\ x_{16} \\ x_{17} \\ x_{20} \end{bmatrix} \quad (3.1)$$

Gráfica 3.1: Gráfica de Y vs \hat{Y} (Yfit) y Residuales - con LS

Gráfica 3.2: Cross Validation LOO y Errores de Predicción - con LS

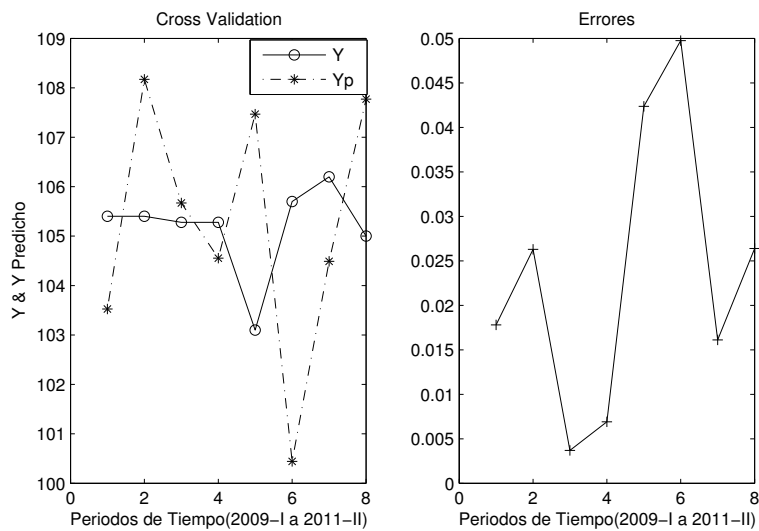


Tabla 3.1: Predicciones del modelo y Errores de Predicción - con LS

OBS.	Real	Predicción	Errores
1	105,400	103,524	1,78E-02
2	105,400	108,171	2,63E-02
3	105,280	105,669	3,70E-03
4	105,280	104,553	6,90E-03
5	103,100	107,468	4,24E-02
6	105,700	100,442	4,97E-02
7	106,200	104,489	1,61E-02
8	105,000	107,770	2,64E-02

Método Mínimos Cuadrados Parciales - PLS1

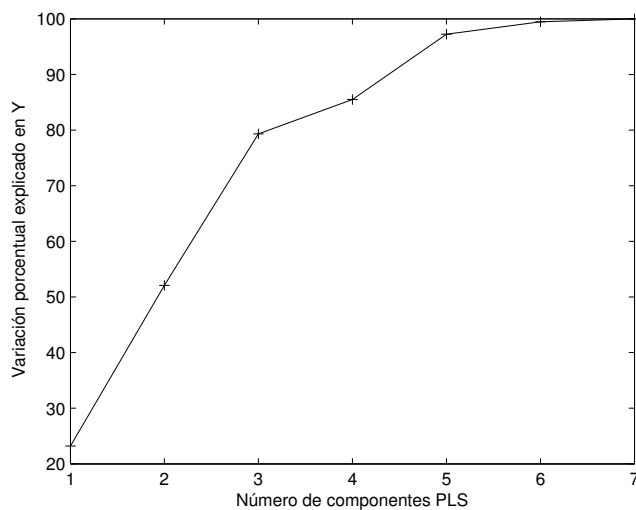
Aplicando el método de Regresión PLS1, se obtiene el modelo:

$$\hat{Y} = 118,7403 + \begin{bmatrix} -5,204E-04 \\ 2,116E-02 \\ 2,914E-02 \\ -1,087E-02 \\ -1,238E-02 \\ -1,077E-02 \\ -2,846E-02 \\ -3,020E-02 \\ -2,814E-02 \\ -3,731E-02 \\ 5,826E-02 \\ -5,426E-03 \\ -7,738E-03 \\ -6,187E-04 \\ -2,608E-02 \\ 3,115E-03 \\ -4,258E-02 \\ -2,094E-03 \\ -9,450E-04 \\ -1,665E-02 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{20} \end{bmatrix} \quad (3.2)$$

Tabla 3.2: Influencia de las variables latentes o componentes PLS sobre Y

	Número de componentes						
	1	2	3	4	5	6	7
Indiv.	0,232146	0,288534	0,272523	0,061885	0,117173	0,02243	0,00531
%Acum.	23,2146	52,06797	79,32025	85,50875	97,22603	99,46905	100

Gráfica 3.3: Influencia de las variables latentes o componentes PLS sobre Y



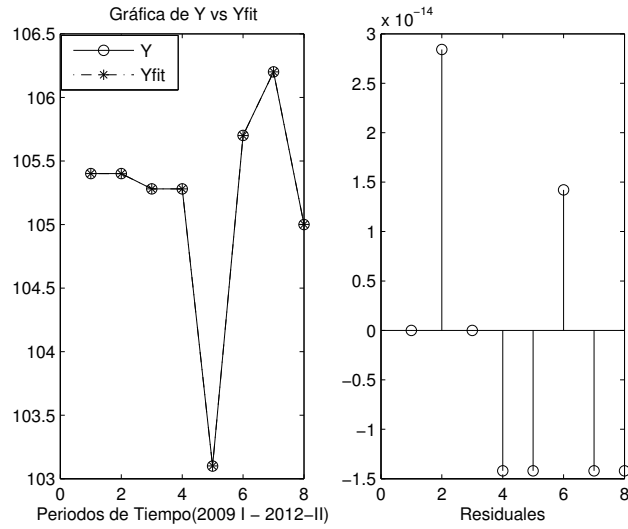
Gráfica 3.4: Gráfica de Y vs \hat{Y} (Yfit) y Residuales

Tabla 3.3: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	105,400	103,927	1,40E-02
2	105,400	106,320	8,73E-03
3	105,280	105,240	3,78E-04
4	105,280	105,838	5,30E-03
5	103,100	105,517	2,34E-02
6	105,700	102,899	2,65E-02
7	106,200	105,241	9,03E-03
8	105,000	105,702	6,69E-03

Modelo de Regresión PLS2

Se pretende dar solución al problema de construir un modelo con la intención de predecir el comportamiento de las variables: Saber - Pro, Promedio Académico y Deserción, a partir, del comportamiento de las Variables de entrada: Estudiantes, Docentes, Procesos Académicos e Investigación.

Aplicando el método PLS2 Regression (PLS2R), se obtuvo el modelo:

$$\hat{Y} = \begin{bmatrix} 118,7403213 \\ 3,026418071 \\ -157,3675828 \end{bmatrix} + B^T [x_1 \ x_2 \ x_3 \ \dots \ x_{20}]^T \quad (3.3)$$

Donde B esta dada por la matriz:

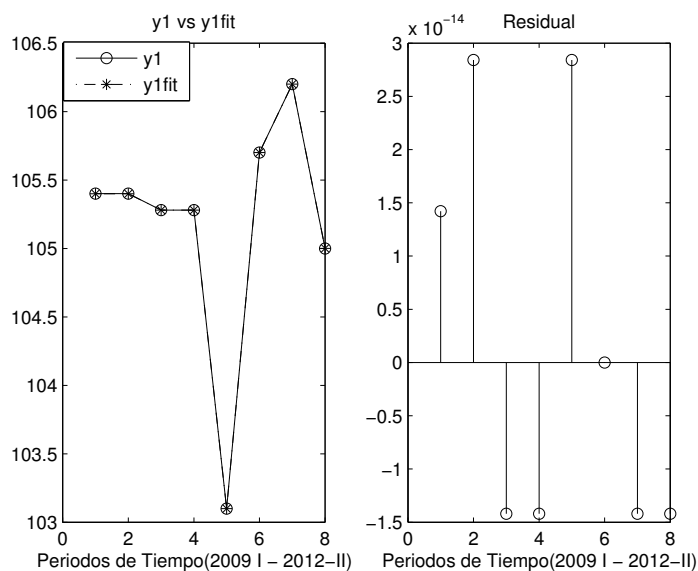
$$B = \begin{bmatrix} -5,204E-04 & -2,398E-05 & 6,969E-03 \\ 2,116E-02 & 8,862E-04 & -4,913E-01 \\ 2,914E-02 & 1,161E-03 & -2,675E-01 \\ -1,087E-02 & 6,716E-04 & 2,750E-01 \\ -1,238E-02 & -1,565E-03 & 8,860E-01 \\ -1,077E-02 & 2,994E-05 & -7,642E-02 \\ -2,846E-02 & 4,305E-04 & 3,173E-02 \\ -3,020E-02 & -7,472E-04 & 6,597E-01 \\ -2,814E-02 & -5,922E-04 & 3,846E-01 \\ -3,731E-02 & -1,484E-04 & 4,855E-02 \\ 5,826E-02 & 8,864E-04 & -6,212E-01 \\ -5,426E-03 & -5,124E-05 & 1,208E-02 \\ -7,738E-03 & -3,307E-04 & 1,963E-01 \\ -6,187E-04 & 1,681E-05 & -1,909E-03 \\ -2,608E-02 & -6,095E-04 & 3,162E-01 \\ 3,115E-03 & 4,163E-04 & -1,412E-01 \\ -4,258E-02 & -1,296E-03 & 6,138E-01 \\ -2,094E-03 & 2,809E-04 & -7,363E-02 \\ -9,450E-04 & 1,180E-04 & -3,080E-02 \\ -1,665E-02 & -2,899E-04 & 2,459E-01 \end{bmatrix}$$

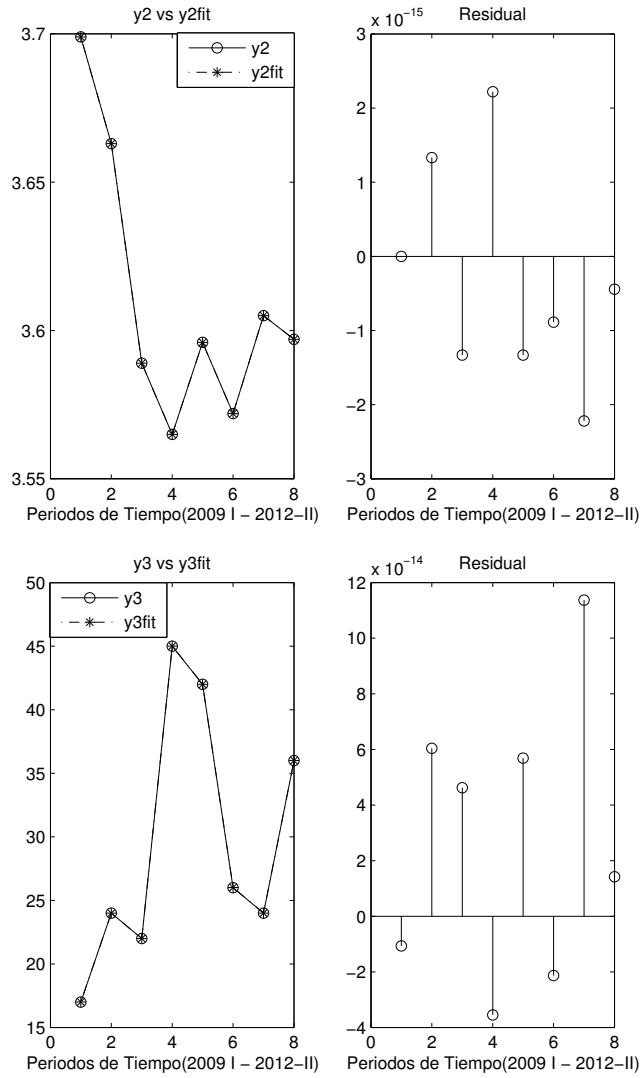
Tabla 3.4: Influencia de las Variables Latentes o Componentes PLS Sobre Y

	Número de componentes						
	1	2	3	4	5	6	7
Indiv.	0,115471	0,144676	0,26569	0,154173	0,214038	0,104494	0,001458
% Acum.	11,5471	26,01472	52,58376	68,00103	89,40479	99,85419	100

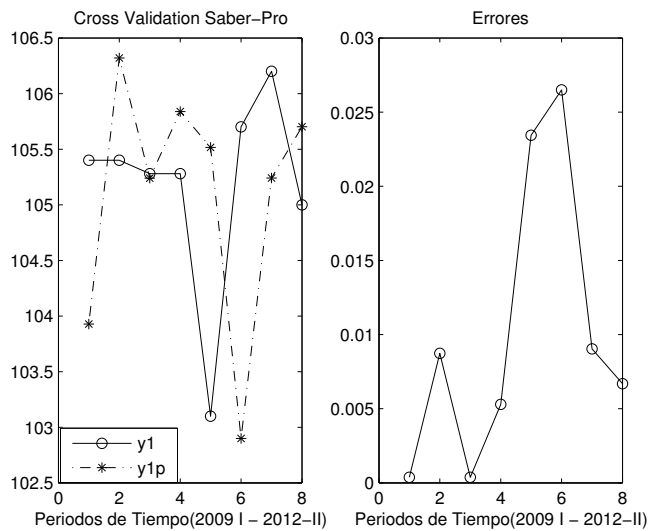
Observemos el comportamiento de Y y \hat{Y} en cada componente:

Gráfica 3.5: Y vs Yfit - Por Componente





Gráfica 3.6: Validación Cruzada de Y - Por Componente



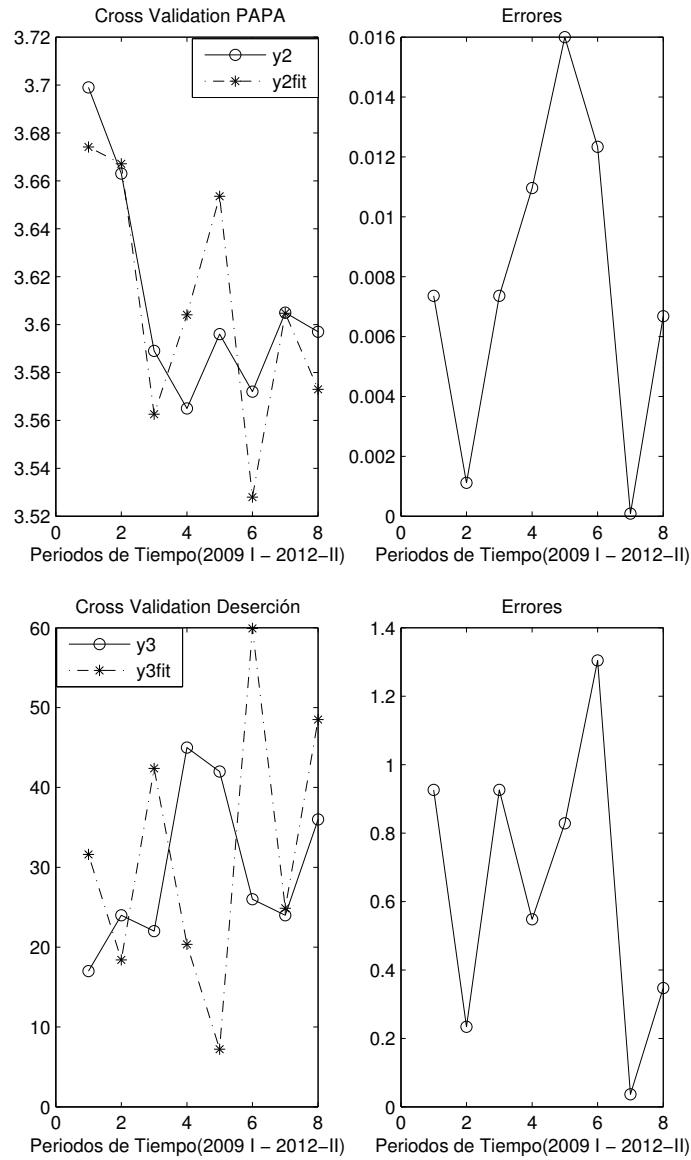


Tabla 3.5: Predicciones y Errores de Predicción en Y

PREDICCIÓN Y ERRORES	OBSERVACIONES							
	1	2	3	4	5	6	7	8
y_1	105,400	105,400	105,280	105,280	103,100	105,700	106,200	105,000
y_{1P}	103,927	106,320	105,240	105,838	105,517	102,899	105,241	105,702
e_1	3,78E-04	8,73E-03	3,78E-04	5,30E-03	2,34E-02	2,65E-02	9,03E-03	6,69E-03
y_2	3,699	3,663	3,589	3,565	3,596	3,572	3,605	3,597
y_{2P}	3,674	3,667	3,563	3,604	3,654	3,528	3,605	3,573
e_2	7,36E-03	1,12E-03	7,36E-03	1,10E-02	1,60E-02	1,23E-02	8,69E-05	6,68E-03
y_3	17,000	24,000	22,000	45,000	42,000	26,000	24,000	36,000
y_{3P}	31,606	18,390	42,378	20,338	7,214	59,917	24,875	48,500
e_3	9,26E-01	2,34E-01	9,26E-01	5,48E-01	8,28E-01	1,30	3,65E-02	3,47E-01

Exploración sobre el comportamiento las variables de salida y_2 y y_3 en forma independiente, usando PLS1R

Ya se observó el comportamiento de y_1 (Saber Pro) usando PLS1, Ahora exploremos el comportamiento de y_2 (Promedio Académico) y y_3 (Deserción) en forma independiente aplicando PLS1, para determinar si hay mejoras en predicción comparado con los comportamientos obtenidos en forma conjunta de PLS2-R.

PLS1 en y_2 (Promedio Académico)

Tabla 3.6: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	3,699	3,6742	6,72E-03
2	3,663	3,6671	1,12E-03
3	3,589	3,5626	7,36E-03
4	3,565	3,6041	1,10E-02
5	3,596	3,6535	1,60E-02
6	3,572	3,5279	1,23E-02
7	3,605	3,6047	8,69E-05
8	3,597	3,5730	6,68E-03

PLS1 en y_3 (Deserción)

Tabla 3.7: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	17	31,60634	8,59E-01
2	24	18,39045	2,34E-01
3	22	42,37799	9,26E-01
4	45	20,33803	5,48E-01
5	42	7,213934	8,28E-01
6	26	59,91666	1,30
7	24	24,87549	3,65E-02
8	36	48,49957	3,47E-01

Explorando métodos de Linealización en Modelos de Regresión PLS2

Se exploran técnicas sencillas de linealización en la variable de salida y_3 (la cual posee problemas en predicción usando el método PLS2R) sobre el nuevo espacio de variables linealizadas (\tilde{Y}). Se toman los modelos de mejor comportamiento predictivo en todas las variables de salida.

Se decide transformar o aplicar linealización sobre la variable de salida y_3 , bajo aplicaciones de la forma: $T(w_i) = e^{\sqrt[k]{w_i}}$, con $k > 0$. Este tipo de transformación lleva los datos iniciales a otro espacio más reducido en escala y por ende con más opciones de ser ajustadas a una línea recta. Lo importante de este método es que además de obtener buena predicción en el nuevo espacio se pueden seguir obteniendo información de influencia toda vez que el comportamiento en cuanto a variabilidad entre estos dos espacios es de tipo directa.

A continuación se presenta el modelo ajustado para el programa Administración de Empresas Nocturna con las tablas de errores y predicciones. El valor elegido para k en cada programa será aquel que garantice un porcentaje de predicción global entre el 96 y el 98 %

Transformando los valores de Y en su tercera componente y aplicando la técnica PLS-R obtenemos (en términos de \tilde{Y}) el modelo:

$$\tilde{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \tilde{y}_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ e^{\sqrt[3]{y_3}} \end{bmatrix} = B^T [x_1 \ x_2 \ x_3 \ \dots \ x_{20}]^T \quad (3.4)$$

Donde B esta dada por la matriz:

$$B = \begin{bmatrix} 118,740321 & 3,026418 & 2,328072 \\ -5,204E-04 & -2,398E-05 & 2,594E-05 \\ 2,116E-02 & 8,862E-04 & -1,800E-03 \\ 2,914E-02 & 1,161E-03 & -1,083E-03 \\ -1,087E-02 & 6,716E-04 & 1,079E-03 \\ -1,238E-02 & -1,565E-03 & 3,023E-03 \\ -1,077E-02 & 2,994E-05 & -1,330E-04 \\ -2,846E-02 & 4,305E-04 & -4,375E-05 \\ -3,020E-02 & -7,472E-04 & 2,558E-03 \\ -2,814E-02 & -5,922E-04 & 1,550E-03 \\ -3,731E-02 & -1,484E-04 & 6,766E-04 \\ 5,826E-02 & 8,864E-04 & -2,501E-03 \\ -5,426E-03 & -5,124E-05 & 9,896E-05 \\ -7,738E-03 & -3,307E-04 & 7,206E-04 \\ -6,187E-04 & 1,681E-05 & 3,362E-06 \\ -2,608E-02 & -6,095E-04 & 1,285E-03 \\ 3,115E-03 & 4,163E-04 & -4,515E-04 \\ -4,258E-02 & -1,296E-03 & 2,375E-03 \\ -2,094E-03 & 2,809E-04 & -1,823E-04 \\ -9,450E-04 & 1,180E-04 & -7,516E-05 \\ -1,665E-02 & -2,899E-04 & 9,081E-04 \end{bmatrix}$$

Haciendo Validación Cruzada sobre \tilde{Y} , se obtienen los resultados:

Tabla 3.8: Predicciones y Errores de Predicción en \tilde{Y}

PREDICCIÓN	OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8
y_1	105,400	105,400	105,280	105,280	103,100	105,700	106,200	105,000
y_{1p}	103,927	106,320	105,240	105,838	105,517	102,899	105,241	105,702
e_1	3,781E-04	8,728E-03	3,781E-04	5,298E-03	2,344E-02	2,650E-02	9,031E-03	6,690E-03
y_2	3,699	3,663	3,589	3,565	3,596	3,572	3,605	3,597
y_{2p}	3,674	3,667	3,563	3,604	3,654	3,528	3,605	3,573
e_2	7,359E-03	1,116E-03	7,359E-03	1,096E-02	1,600E-02	1,234E-02	8,689E-05	6,679E-03
\tilde{y}_3	3,001	3,040	3,030	3,112	3,104	3,049	3,040	3,086
\tilde{y}_{3p}	3,066	3,009	3,097	3,026	2,970	3,186	3,052	3,119
e_3	2,225E-02	1,006E-02	2,225E-02	2,753E-02	4,314E-02	4,488E-02	4,088E-03	1,080E-02

Para obtener valores en términos de Y (nuevamente), se usa la expresión:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ (\ln \lambda_3)^{30} \end{bmatrix} \quad (3.5)$$

Donde:

$$\lambda_1 = (B_{*1})^T [1 \quad x_1 \quad x_2 \quad x_3 \quad \dots \quad x_{20}]^T$$

$$\lambda_2 = (B_{*2})^T [1 \quad x_1 \quad x_2 \quad x_3 \quad \dots \quad x_{20}]^T$$

$$\lambda_3 = (B_{*3})^T [1 \quad x_1 \quad x_2 \quad x_3 \quad \dots \quad x_{20}]^T$$

Cada B_{*i} con $i = 1 : 3$, son las columnas 1, 2, 3 de B respectivamente dada en (3.4).

Ahora se presentan las tablas de predicciones de los modelos PLS1R, PLS2R y PLS2R linealizado de mejor ajuste, para el resto de programas en estudio:

3.1.2. Administración de Empresas (D)

Tabla 3.9: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	107,500	107,945	4,140E-03
2	107,500	106,843	6,111E-03
3	104,100	104,235	1,297E-03
4	104,100	104,624	5,037E-03
5	106,000	105,709	2,750E-03
6	105,700	106,200	4,733E-03
7	106,200	105,129	1,008E-02
8	105,000	105,538	5,126E-03

Tabla 3.10: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN	OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8
y_1	107,500	107,500	104,100	104,100	106,000	105,700	106,200	105,000
y_2	3,794	3,793	3,724	3,650	3,727	3,730	3,723	3,697
y_3	24,000	36,000	31,000	10,000	14,000	18,000	31,000	18,000
y_{1P}	107,500	106,843	104,235	104,624	105,709	106,200	105,129	105,538
y_{2P}	3,794	3,751	3,678	3,723	3,683	3,782	3,680	3,747
y_{3P}	24,000	13,654	4,307	48,654	-11,806	46,010	-0,926	55,437
e_1	2,644E-16	6,111E-03	1,297E-03	5,037E-03	2,750E-03	4,733E-03	1,008E-02	5,126E-03
e_2	4,682E-16	1,105E-02	1,234E-02	2,002E-02	1,192E-02	1,394E-02	1,160E-02	1,344E-02
e_3	1,776E-15	6,207E-01	8,611E-01	3,865	1,843	1,556	1,030	2,080

Tabla 3.11: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 40$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	107,500	107,500	104,100	104,100	106,000	105,700	106,200	105,000	
y_2	3,794	3,793	3,724	3,650	3,727	3,730	3,723	3,697	
\tilde{y}_3	2,953	2,985	2,973	2,884	2,910	2,930	2,973	2,930	
y_{1P}	107,945	106,843	104,235	104,624	105,709	106,200	105,129	105,538	
y_{2P}	3,833	3,751	3,678	3,723	3,683	3,782	3,680	3,747	
\tilde{y}_{3P}	3,027	2,913	2,884	3,013	2,826	3,021	2,871	3,051	
e_1	1,30E-03	6,11E-03	1,30E-03	5,04E-03	2,75E-03	4,73E-03	1,01E-02	5,13E-03	
e_2	1,23E-02	1,11E-02	1,23E-02	2,00E-02	1,19E-02	1,39E-02	1,16E-02	1,34E-02	
e_3	3,02E-02	2,42E-02	3,02E-02	4,47E-02	2,90E-02	3,12E-02	3,45E-02	4,13E-02	

3.1.3. Administración en Sistemas Informáticos

Tabla 3.12: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	99,700	98,285	1,419E-02
2	99,700	100,642	9,445E-03
3	98,510	101,509	3,044E-02
4	98,510	97,277	1,252E-02
5	104,600	102,453	2,052E-02
6	104,100	105,417	1,265E-02
7	103,900	105,034	1,092E-02
8	103,300	101,297	1,939E-02

Tabla 3.13: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	99,700	99,700	98,510	98,510	104,600	104,100	103,900	103,300	
y_2	3,644	3,548	3,443	3,435	3,419	3,474	3,418	3,491	
y_3	13,000	28,000	25,000	23,000	32,000	34,000	23,000	22,000	
y_{1P}	99,700	100,642	101,509	97,277	102,453	105,417	105,034	101,297	
y_{2P}	3,644	3,633	3,587	3,317	3,295	3,465	3,539	3,303	
y_{3P}	13,000	17,851	13,373	47,589	46,352	28,411	12,068	38,720	
e_1	4,276E-16	9,445E-03	3,044E-02	1,252E-02	2,052E-02	1,265E-02	1,092E-02	1,939E-02	
e_2	1,219E-16	2,408E-02	4,183E-02	3,428E-02	3,631E-02	2,474E-03	3,553E-02	5,384E-02	
e_3	5,466E-16	3,625E-01	4,651E-01	1,069	4,485E-01	1,644E-01	4,753E-01	7,600E-01	

Tabla 3.14: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 30$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	99,700	99,700	98,510	98,510	104,600	104,100	103,900	103,300	
y_2	3,644	3,548	3,443	3,435	3,419	3,474	3,418	3,491	
\tilde{y}_3	2,972	3,057	3,044	3,035	3,072	3,079	3,035	3,030	
y_{1P}	98,285	100,642	101,509	97,277	102,453	105,417	105,034	101,297	
y_{2P}	3,534	3,633	3,587	3,317	3,295	3,465	3,539	3,303	
\tilde{y}_{3P}	3,045	2,995	2,973	3,180	3,160	3,052	2,965	3,134	
e_1	3,04E-02	9,44E-03	3,04E-02	1,25E-02	2,05E-02	1,27E-02	1,09E-02	1,94E-02	
e_2	4,18E-02	2,41E-02	4,18E-02	3,43E-02	3,63E-02	2,47E-03	3,55E-02	5,38E-02	
e_3	2,34E-02	2,03E-02	2,34E-02	4,78E-02	2,85E-02	8,90E-03	2,29E-02	3,45E-02	

3.1.4. Arquitectura

Tabla 3.15: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	99,400	101,878	2,493E-02
2	99,400	103,144	3,767E-02
3	103,360	98,560	4,644E-02
4	103,360	98,600	4,606E-02
5	104,340	99,606	4,537E-02
6	102,600	97,735	4,742E-02
7	102,700	106,321	3,526E-02
8	103,300	97,257	5,850E-02

Tabla 3.16: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN	OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8
y_1	99,400	99,400	103,360	103,360	104,340	102,600	102,700	103,300
y_2	3,701	3,671	3,681	3,600	3,599	3,600	3,625	3,651
y_3	11,000	10,000	12,000	12,000	16,000	19,000	17,000	20,000
y_{1P}	101,878	103,144	98,560	98,600	99,606	97,735	106,321	97,257
y_{2P}	3,706	3,625	3,692	3,682	3,622	3,661	3,589	3,718
y_{3P}	8,025	15,601	8,779	3,990	15,849	10,420	22,156	9,801
e_1	2,49E-02	3,77E-02	4,64E-02	4,61E-02	4,54E-02	4,74E-02	3,53E-02	5,85E-02
e_2	1,30E-03	1,25E-02	3,10E-03	2,29E-02	6,30E-03	1,69E-02	9,95E-03	1,85E-02
e_3	2,70E-01	5,60E-01	2,68E-01	6,68E-01	9,45E-03	4,52E-01	3,03E-01	5,10E-01

Tabla 3.17: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 20$

PREDICCIÓN	OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8
y_1	99,400	99,400	103,360	103,360	104,340	102,600	102,700	103,300
y_2	3,701	3,671	3,681	3,600	3,599	3,600	3,625	3,651
\tilde{y}_3	3,088	3,071	3,103	3,103	3,154	3,186	3,165	3,195
y_{1P}	101,878	103,144	98,560	98,600	99,606	97,735	106,321	97,257
y_{2P}	3,706	3,625	3,692	3,682	3,622	3,661	3,589	3,718
\tilde{y}_{3P}	3,054	3,141	3,061	3,002	3,147	3,080	3,229	3,070
e_1	4,64E-02	3,77E-02	4,64E-02	4,61E-02	4,54E-02	4,74E-02	3,53E-02	5,85E-02
e_2	3,10E-03	1,25E-02	3,10E-03	2,29E-02	6,30E-03	1,69E-02	9,95E-03	1,85E-02
e_3	1,36E-02	2,29E-02	1,36E-02	3,23E-02	2,28E-03	3,31E-02	2,01E-02	3,91E-02

3.1.5. Ingeniería Civil

Tabla 3.18: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	103,300	97,705	5,416E-02
2	103,300	110,263	6,741E-02
3	99,610	107,944	8,367E-02
4	99,610	95,046	4,582E-02
5	101,200	110,613	9,302E-02
6	108,600	101,743	6,314E-02
7	106,200	112,050	5,509E-02
8	105,800	94,722	1,047E-01

Tabla 3.19: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	103,300	103,300	99,610	99,610	101,200	108,600	106,200	105,800	
y_2	3,651	3,630	3,609	3,603	3,633	3,651	3,674	3,642	
y_3	18,000	13,000	23,000	18,000	21,000	13,000	14,000	11,000	
y_{1P}	97,705	110,263	107,944	95,046	110,613	101,743	112,050	94,722	
y_{2P}	3,637	3,651	3,629	3,590	3,654	3,636	3,684	3,625	
y_{3P}	23,332	8,063	13,487	21,222	12,185	20,056	7,595	23,679	
e_1	5,42E-02	6,74E-02	8,37E-02	4,58E-02	9,30E-02	6,31E-02	5,51E-02	1,05E-01	
e_2	3,79E-03	5,66E-03	5,55E-03	3,57E-03	5,78E-03	4,04E-03	2,69E-03	4,73E-03	
e_3	2,96E-01	3,80E-01	4,14E-01	1,79E-01	4,20E-01	5,43E-01	4,57E-01	1,15	

Tabla 3.20: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 5$ para y_1 , $k = 20$ para y_3 .

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
\tilde{y}_1	12,532	12,532	12,304	12,304	12,402	12,854	12,709	12,684	
\tilde{y}_2	3,651	3,630	3,609	3,603	3,633	3,651	3,674	3,642	
\tilde{y}_3	3,176	3,117	3,221	3,176	3,204	3,117	3,130	3,088	
\tilde{y}_{1P}	12,189	12,957	12,814	12,026	12,978	12,435	13,067	12,006	
\tilde{y}_{2P}	3,637	3,651	3,629	3,590	3,654	3,636	3,684	3,625	
\tilde{y}_{3P}	3,231	3,068	3,120	3,200	3,111	3,192	3,061	3,225	
e_1	4,15E-02	3,40E-02	4,15E-02	2,26E-02	4,64E-02	3,26E-02	2,82E-02	5,35E-02	
e_2	5,55E-03	5,66E-03	5,55E-03	3,57E-03	5,78E-03	4,04E-03	2,69E-03	4,73E-03	
e_3	3,13E-02	1,57E-02	3,13E-02	7,73E-03	2,91E-02	2,42E-02	2,21E-02	4,45E-02	

3.1.6. Ingeniería Eléctrica

Tabla 3.21: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	94,100	90,660	3,656E-02
2	94,100	92,237	1,980E-02
3	87,800	95,001	8,201E-02
4	102,500	98,475	3,927E-02
5	101,000	102,933	1,914E-02
6	107,500	103,040	4,149E-02
7	103,800	104,990	1,147E-02
8	102,600	100,260	2,281E-02

Tabla 3.22: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	94,100	94,100	87,800	102,500	101,000	107,500	103,800	102,600	
y_2	3,689	3,663	3,624	3,535	3,596	3,599	3,640	3,619	
y_3	29,000	20,000	18,000	13,000	22,000	20,000	23,000	23,000	
y_{1P}	90,660	92,237	95,001	98,475	102,933	103,040	104,990	100,260	
y_{2P}	3,681	3,602	3,668	3,685	3,642	3,613	3,613	3,657	
y_{3P}	21,091	20,370	25,197	18,615	18,324	18,032	24,087	24,994	
e_1	3,66E-02	1,98E-02	8,20E-02	3,93E-02	1,91E-02	4,15E-02	1,15E-02	2,28E-02	
e_2	2,21E-03	1,65E-02	1,20E-02	4,25E-02	1,29E-02	3,83E-03	7,52E-03	1,04E-02	
e_3	2,73E-01	1,85E-02	4,00E-01	4,32E-01	1,67E-01	9,84E-02	4,73E-02	8,67E-02	

Tabla 3.23: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 10$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	94,100	94,100	87,800	102,500	101,000	107,500	103,800	102,600	
y_2	3,689	3,663	3,624	3,535	3,596	3,599	3,640	3,619	
\tilde{y}_3	4,057	3,855	3,801	3,641	3,905	3,855	3,929	3,929	
y_{1P}	90,660	92,237	95,001	98,475	102,933	103,040	104,990	100,260	
y_{2P}	3,681	3,602	3,668	3,685	3,642	3,613	3,613	3,657	
\tilde{y}_{3P}	3,881	3,840	3,971	3,826	3,838	3,814	3,945	3,978	
e_1	8,20E-02	1,98E-02	8,20E-02	3,93E-02	1,91E-02	4,15E-02	1,15E-02	2,28E-02	
e_2	1,20E-02	1,65E-02	1,20E-02	4,25E-02	1,29E-02	3,83E-03	7,52E-03	1,04E-02	
e_3	4,49E-02	3,71E-03	4,49E-02	5,07E-02	1,71E-02	1,05E-02	4,29E-03	1,26E-02	

3.1.7. Ingeniería Electrónica

Tabla 3.24: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	102,900	102,720	1,754E-03
2	102,900	102,567	3,235E-03
3	99,850	102,161	2,314E-02
4	99,850	103,073	3,228E-02
5	107,000	109,192	2,049E-02
6	108,000	105,373	2,433E-02
7	107,300	105,477	1,699E-02
8	106,500	109,518	2,834E-02

Tabla 3.25: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,900	102,900	99,850	99,850	107,000	108,000	107,300	106,500	
y_2	3,806	3,795	3,685	3,669	3,645	3,700	3,743	3,687	
y_3	20,000	27,000	27,000	15,000	21,000	26,000	17,000	19,000	
y_{1P}	102,720	102,567	102,161	103,073	109,192	105,373	105,477	109,518	
y_{2P}	3,780	3,798	3,741	3,865	3,774	3,553	3,631	3,874	
y_{3P}	25,694	21,705	25,910	26,572	20,475	27,813	15,361	21,328	
e_1	1,75E-03	3,23E-03	2,31E-02	3,23E-02	2,05E-02	2,43E-02	1,70E-02	2,83E-02	
e_2	6,75E-03	7,81E-04	1,53E-02	5,35E-02	3,55E-02	3,97E-02	2,99E-02	5,07E-02	
e_3	2,85E-01	1,96E-01	4,04E-02	7,71E-01	2,50E-02	6,97E-02	9,64E-02	1,23E-01	

Tabla 3.26: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 20$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,900	102,900	99,850	99,850	107,000	108,000	107,300	106,500	
y_2	3,806	3,795	3,685	3,669	3,645	3,700	3,743	3,687	
\tilde{y}_3	3,195	3,252	3,252	3,142	3,204	3,244	3,165	3,186	
y_{1P}	102,720	102,567	102,161	103,073	109,192	105,373	105,477	109,518	
y_{2P}	3,780	3,798	3,741	3,865	3,774	3,553	3,631	3,874	
\tilde{y}_{3P}	3,241	3,209	3,241	3,239	3,198	3,260	3,154	3,201	
e_1	2,31E-02	3,23E-03	2,31E-02	3,23E-02	2,05E-02	2,43E-02	1,70E-02	2,83E-02	
e_2	1,53E-02	7,81E-04	1,53E-02	5,35E-02	3,55E-02	3,97E-02	2,99E-02	5,07E-02	
e_3	3,41E-03	1,33E-02	3,41E-03	3,09E-02	1,82E-03	4,94E-03	3,56E-03	4,91E-03	

3.1.8. Ingeniería Física

Tabla 3.27: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	10,630	10,514	1,092E-02
2	10,630	10,714	7,866E-03
3	10,560	10,784	2,122E-02
4	10,770	10,567	1,888E-02
5	10,470	10,556	8,198E-03
6	10,730	10,376	3,297E-02
7	10,530	10,584	5,088E-03
8	11,010	10,950	5,443E-03

Tabla 3.28: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	10,630	10,630	10,560	10,770	10,470	10,730	10,530	11,010	
y_2	3,665	3,714	3,583	3,664	3,592	3,478	3,699	3,484	
y_3	17,000	7,000	7,000	13,000	12,000	16,000	13,000	21,000	
y_{1P}	10,514	10,714	10,784	10,567	10,556	10,376	10,584	10,950	
y_{2P}	3,648	3,734	3,555	3,647	3,633	3,527	3,592	3,614	
y_{3P}	4,404	17,596	20,638	-5,399	21,549	-3,489	4,599	37,084	
e_1	1,09E-02	7,87E-03	2,12E-02	1,89E-02	8,20E-03	3,30E-02	5,09E-03	5,44E-03	
e_2	4,74E-03	5,41E-03	7,78E-03	4,59E-03	1,15E-02	1,41E-02	2,91E-02	3,73E-02	
e_3	7,41E-01	1,51	1,95	1,42	7,96E-01	1,22	6,46E-01	7,66E-01	

Tabla 3.29: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 40$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	10,630	10,630	10,560	10,770	10,470	10,730	10,530	11,010	
y_2	3,665	3,714	3,583	3,664	3,592	3,478	3,699	3,484	
\tilde{y}_3	2,925	2,857	2,857	2,904	2,898	2,921	2,904	2,942	
y_{1P}	10,514	10,714	10,784	10,567	10,556	10,376	10,584	10,950	
y_{2P}	3,648	3,734	3,555	3,647	3,633	3,527	3,592	3,614	
\tilde{y}_{3P}	2,834	2,934	2,953	2,772	2,968	2,783	2,838	3,067	
e_1	2,12E-02	7,87E-03	2,12E-02	1,89E-02	8,20E-03	3,30E-02	5,09E-03	5,44E-03	
e_2	7,78E-03	5,41E-03	7,78E-03	4,59E-03	1,15E-02	1,41E-02	2,91E-02	3,73E-02	
e_3	3,35E-02	2,70E-02	3,35E-02	4,56E-02	2,42E-02	4,71E-02	2,27E-02	4,24E-02	

3.1.9. Ingeniería Industrial

Tabla 3.30: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	102,800	101,037	1,715E-02
2	102,800	104,463	1,617E-02
3	101,520	100,193	1,307E-02
4	101,520	104,899	3,328E-02
5	102,000	105,622	3,551E-02
6	108,700	102,543	5,664E-02
7	108,300	105,296	2,773E-02
8	106,500	110,538	3,791E-02

Tabla 3.31: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,800	102,800	101,520	101,520	102,000	108,700	108,300	106,500	
y_2	3,792	3,796	3,740	3,736	3,750	3,763	3,798	3,807	
y_3	10,000	12,000	10,000	8,000	12,000	6,000	12,000	6,000	
y_{1P}	101,037	104,463	100,193	104,899	105,622	102,543	105,296	110,538	
y_{2P}	3,797	3,786	3,759	3,722	3,729	3,775	3,814	3,782	
y_{3P}	13,290	9,226	11,257	7,700	7,235	14,706	9,678	9,314	
e_1	1,71E-02	1,62E-02	1,31E-02	3,33E-02	3,55E-02	5,66E-02	2,77E-02	3,79E-02	
e_2	1,38E-03	2,75E-03	5,19E-03	3,83E-03	5,60E-03	3,18E-03	4,23E-03	6,69E-03	
e_3	3,29E-01	2,31E-01	1,26E-01	3,75E-02	3,97E-01	1,45E+00	1,94E-01	5,52E-01	

Tabla 3.32: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 20$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,800	102,800	101,520	101,520	102,000	108,700	108,300	106,500	
y_2	3,792	3,796	3,740	3,736	3,750	3,763	3,798	3,807	
\tilde{y}_3	3,071	3,103	3,071	3,033	3,103	2,985	3,103	2,985	
y_{1P}	101,037	104,463	100,193	104,899	105,622	102,543	105,296	110,538	
y_{2P}	3,797	3,786	3,759	3,722	3,729	3,775	3,814	3,782	
\tilde{y}_{3P}	3,127	3,057	3,090	3,031	3,021	3,141	3,051	3,060	
e_1	1,31E-02	1,62E-02	1,31E-02	3,33E-02	3,55E-02	5,66E-02	2,77E-02	3,79E-02	
e_2	5,19E-03	2,75E-03	5,19E-03	3,83E-03	5,60E-03	3,18E-03	4,23E-03	6,69E-03	
e_3	6,17E-03	1,49E-02	6,17E-03	7,67E-04	2,62E-02	5,22E-02	1,67E-02	2,49E-02	

3.1.10. Ingeniería Química

Tabla 3.33: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	100,200	99,234	9,64E-03
2	100,200	101,586	1,38E-02
3	97,950	108,920	1,12E-01
4	97,950	112,679	1,50E-01
5	106,100	107,831	1,63E-02
6	107,500	103,512	3,71E-02
7	107,100	100,975	5,72E-02
8	107,000	117,186	9,52E-02

Tabla 3.34: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	100,200	100,200	97,950	97,950	106,100	107,500	107,100	107,000	
y_2	3,686	3,627	3,582	3,566	3,544	3,585	3,602	3,622	
y_3	26,000	24,000	16,000	10,000	14,000	26,000	19,000	31,000	
y_{1P}	99,234	101,586	108,920	112,679	107,831	103,512	100,975	117,186	
y_{2P}	3,604	3,703	3,709	3,533	3,653	3,468	3,521	3,690	
y_{3P}	18,970	30,251	33,954	12,613	25,517	13,741	11,284	33,626	
e_1	9,64E-03	1,38E-02	1,12E-01	1,50E-01	1,63E-02	3,71E-02	5,72E-02	9,52E-02	
e_2	2,22E-02	2,10E-02	3,54E-02	9,33E-03	3,07E-02	3,28E-02	2,24E-02	1,87E-02	
e_3	2,70E-01	2,60E-01	1,12	2,61E-01	8,23E-01	4,72E-01	4,06E-01	8,47E-02	

Tabla 3.35: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 10$ para $y_1, k = 20$ para y_3

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
\tilde{y}_1	4,880	4,880	4,863	4,863	4,925	4,935	4,932	4,932	
y_2	3,686	3,627	3,582	3,566	3,544	3,585	3,602	3,622	
\tilde{y}_3	3,244	3,229	3,154	3,071	3,130	3,244	3,186	3,278	
\tilde{y}_{1P}	4,873	4,891	4,946	4,974	4,938	4,905	4,886	5,008	
y_{2P}	3,604	3,703	3,709	3,533	3,653	3,468	3,521	3,690	
\tilde{y}_{3P}	3,179	3,289	3,330	3,135	3,240	3,124	3,101	3,325	
e_1	1,71E-02	2,12E-03	1,71E-02	2,29E-02	2,61E-03	6,06E-03	9,34E-03	1,56E-02	
e_2	3,54E-02	2,10E-02	3,54E-02	9,33E-03	3,07E-02	3,28E-02	2,24E-02	1,87E-02	
e_3	5,58E-02	1,84E-02	5,58E-02	2,10E-02	3,50E-02	3,71E-02	2,66E-02	1,42E-02	

3.1.11. Matemáticas

Tabla 3.36: Predicciones del modelo y Errores de Predicción con PLS1

OBS.	Real	Predicción	Errores
1	91,800	94,341	2,768E-02
2	91,800	89,299	2,724E-02
3	102,400	106,586	4,088E-02
4	104,900	102,123	2,648E-02
5	101,800	104,287	2,443E-02
6	101,800	103,097	1,274E-02
7	108,900	101,161	7,107E-02
8	108,900	96,965	1,096E-01

Tabla 3.37: Predicciones y Errores de Predicción en Y usando PLS2R

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	91,800	91,800	102,400	104,900	101,800	101,800	108,900	108,900	
y_2	3,338	3,736	3,102	3,567	3,174	3,266	3,508	3,416	
y_3	6,000	6,000	5,000	10,000	5,000	23,000	8,000	26,000	
y_{1P}	94,341	89,299	106,586	102,123	104,287	103,097	101,161	96,965	
y_{2P}	3,546	3,467	3,149	3,540	3,370	3,349	3,046	2,711	
y_{3P}	6,813	8,608	24,285	-1,879	20,131	6,167	21,082	2,725	
e_1	2,77E-02	2,72E-02	4,09E-02	2,65E-02	2,44E-02	1,27E-02	7,11E-02	1,10E-01	
e_2	6,23E-02	7,20E-02	1,51E-02	7,57E-03	6,16E-02	2,53E-02	1,32E-01	2,06E-01	
e_3	1,35E-01	4,35E-01	3,86	1,190	3,030	7,32E-01	1,640	8,95E-01	

Tabla 3.38: Predicciones y Errores de Predicción sobre \tilde{Y} (Linealizada), $k = 10$ para y_1 y $k = 40$ para y_3

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
\tilde{y}_1	4,813	4,813	4,897	4,916	4,893	4,893	4,945	4,945	
\tilde{y}_2	3,090	3,130	3,064	3,113	3,072	3,082	3,107	3,098	
\tilde{y}_3	2,846	2,846	2,832	2,884	2,832	2,949	2,867	2,959	
\tilde{y}_{1P}	4,833	4,794	4,930	4,896	4,911	4,902	4,888	4,856	
\tilde{y}_{2P}	3,110	3,103	3,071	3,108	3,093	3,090	3,060	3,025	
\tilde{y}_{3P}	2,844	2,870	2,967	2,789	2,933	2,846	2,939	2,805	
e_1	4,11E-03	3,97E-03	6,61E-03	4,14E-03	3,72E-03	1,84E-03	1,15E-02	1,80E-02	
e_2	6,55E-03	8,37E-03	2,22E-03	1,71E-03	6,76E-03	2,63E-03	1,52E-02	2,36E-02	
e_3	6,50E-04	8,45E-03	4,75E-02	3,30E-02	3,57E-02	3,50E-02	2,49E-02	5,19E-02	

3.2. Construcción de Modelos de Regresión Multivariado Usando técnicas No lineales

3.2.1. Administración de Empresas (N)

Aplicando la técnica KPLS a un modelo de regresión multiple ($L = 1$) (KPLSR)

Se aplica la técnica KPLSR (caso $L = 1$) para el programa Administración de Empresas Nocturna tomando como variable de salida (Deserción), la cuál presentó más dificultad en predicción usando los modelos lineales.

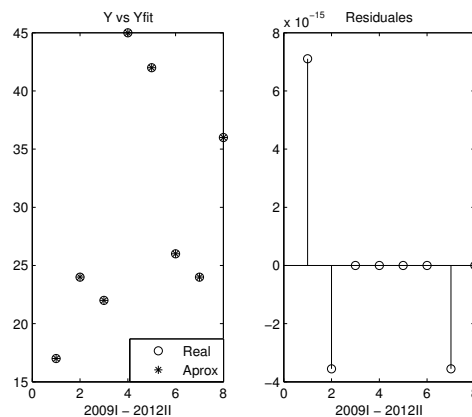
En este trabajo, cada vez que se aplica la técnica KPLSR, tomamos como referencia la función Kernel Gaussiana:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{h}\right)} \quad (3.6)$$

Donde las \mathbf{x}_i 's son los vectores o variables de entrada, con $i, j = 1, \dots, 8$, para nuestro caso.

Los resultados obtenidos se muestran a continuación:

Gráfica 3.7: Y vs Yfit - Aplicando KPLSR, $L = 1$



3.2. CONSTRUCCIÓN DE MODELOS DE REGRESIÓN MULTIVARIADO USANDO TÉCNICAS NO L

El modelo tiene la forma descrita en (1.30): $Y(\mathbf{x}, \mathbf{d}) = \sum_{i=1}^n d_i K(\mathbf{x}, \mathbf{x}_i)$, donde:

$$\mathbf{d} = \begin{bmatrix} -1,1495 \\ -1,2603 \\ 22,1600 \\ 19,0234 \\ 3,0173 \\ 0,8141 \\ 13,1899 \end{bmatrix} \quad (3.7)$$

Tabla 3.39: Predicciones del modelo y Errores de Predicción - con KPLSR, $L = 1$

OBS.	Real	Predicción	Errores
1	17	14,000	1,76E-01
2	24	15,875	3,39E-01
3	22	36,938	6,79E-01
4	45	55,500	2,33E-01
5	42	39,250	6,55E-02
6	26	29,000	1,15E-01
7	24	34,000	4,17E-01
8	36	36,094	2,60E-03

Tabla 3.40: Comparación de errores PLS1R vs KPLSR, $L = 1$

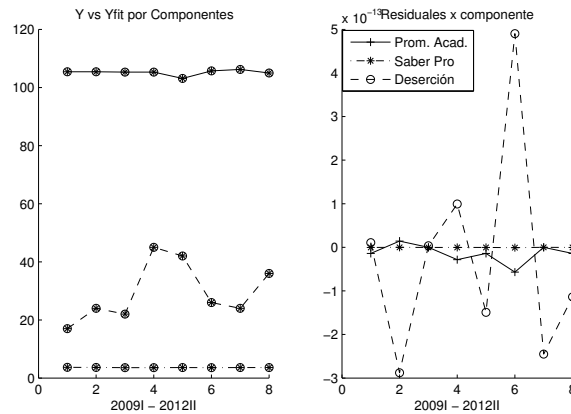
OBS.	Errores PLS1R	Errores KPLSR
1	8,59E-01	1,76E-01
2	2,34E-01	3,39E-01
3	9,26E-01	6,79E-01
4	5,48E-01	2,33E-01
5	8,28E-01	6,55E-02
6	1,30	1,15E-01
7	3,65E-02	4,17E-01
8	3,47E-01	2,60E-03

Aplicación de la técnica KPLS en un modelo de regresión multiobjetivo ó de salida múltiple ($L > 1$)

Ahora se aplica la técnica KPLSR para el caso ($L > 1$), tomando como variables de salida: Resultados SABER-PRO, Promedio Académico y Deserción.

La gráfica presenta el comportamiento y ajuste de los datos aproximados (en asterisco) a los datos reales (en círculos) y en la parte derecha los residuales para cada componente de Y o cada variable de salida en cada periodo de tiempo.

En este procedimiento se obtienen los siguientes resultados:

Gráfica 3.8: Aplicando la técnica KPLSR, $L > 1$ - Programa: Admon Empresas (N)

El modelo para este caso toma la forma: $\hat{Y} = \mathbf{K}U(T^T\mathbf{K}U)^{-1}T^TY$, descrita en (1.26), donde U esta dada por la matriz:

$$\begin{bmatrix} 1,37E-03 & -1,22E-04 & 4,37E-07 & 6,93E-09 & 1,95E-12 & -1,42E-14 & -1,42E-14 \\ 1,39E-03 & -1,23E-04 & 4,29E-07 & 6,73E-09 & -2,01E-12 & -1,42E-14 & -1,42E-14 \\ 2,00E-03 & -1,36E-04 & 1,16E-06 & -5,23E-09 & -4,10E-14 & -1,44E-14 & -1,42E-14 \\ 1,98E-03 & -1,36E-04 & 1,16E-06 & -5,25E-09 & 3,60E-14 & -1,40E-14 & -1,42E-14 \\ 7,39E-03 & 1,31E-04 & -4,01E-07 & 1,75E-10 & -1,44E-14 & -1,46E-14 & -1,46E-14 \\ 3,67E-03 & -1,88E-04 & -2,60E-06 & -2,46E-09 & -3,73E-15 & -1,41E-14 & -1,41E-14 \\ 8,41E-03 & 4,62E-04 & 5,17E-08 & -3,96E-10 & -1,38E-14 & -1,41E-14 & -1,41E-14 \\ -2,62E-02 & 1,12E-04 & -2,40E-07 & -5,00E-10 & -1,42E-14 & -1,43E-14 & -1,43E-14 \end{bmatrix}$$

y T es la matriz:

$$\begin{bmatrix} 9,51E-02 & 5,15E-02 & -2,09E-01 & 1,35E-01 & 5,57E-01 & 7,01E-01 & -1,31E-02 \\ 9,51E-02 & 5,22E-02 & -2,09E-01 & 1,33E-01 & 5,41E-01 & -7,13E-01 & 1,43E-02 \\ 4,49E-02 & 7,40E-02 & -2,34E-01 & 3,68E-01 & -4,21E-01 & -9,48E-03 & -7,07E-01 \\ 4,49E-02 & 7,30E-02 & -2,34E-01 & 3,68E-01 & -4,22E-01 & 1,79E-02 & 7,07E-01 \\ -8,64E-01 & 2,51E-01 & 2,23E-01 & -1,26E-01 & 1,41E-02 & 8,17E-05 & 1,41E-04 \\ 2,19E-01 & 1,33E-01 & -3,15E-01 & -8,19E-01 & -1,98E-01 & 3,72E-03 & -4,06E-04 \\ 4,25E-01 & 2,75E-01 & 7,86E-01 & 1,64E-02 & -3,19E-02 & 1,08E-04 & 1,64E-04 \\ -5,98E-02 & -9,10E-01 & 1,91E-01 & -7,53E-02 & -4,02E-02 & 1,94E-05 & -6,55E-04 \end{bmatrix}$$

ó considerando la forma: $\hat{Y} = \mathbf{K}\mathbf{d} = f(\mathbf{x}, \mathbf{d}) = \sum_{i=1}^n d_{i*}K(\mathbf{x}, \mathbf{x}_i)$. Donde

$$\mathbf{d} = \begin{bmatrix} -81,927 & -3,568E-01 & 523,765 \\ -81,927 & -3,932E-01 & 530,839 \\ -82,046 & -4,655E-01 & 528,429 \\ -82,046 & -4,899E-01 & 551,849 \\ -84,225 & -4,593E-01 & 548,713 \\ -81,624 & -4,833E-01 & 532,707 \\ -81,117 & -4,499E-01 & 530,503 \\ -82,355 & -4,581E-01 & 542,879 \end{bmatrix}$$

Tabla 3.41: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	105,400	105,400	105,280	105,280	103,100	105,700	106,200	105,000	
y_{1P}	105,650	105,494	105,483	103,186	103,475	105,200	105,200	105,020	
e_1	2,37E-03	8,89E-04	1,93E-03	1,99E-02	3,64E-03	4,73E-03	9,42E-03	1,86E-04	
y_2	3,699	3,663	3,589	3,565	3,596	3,572	3,605	3,597	
y_{2P}	3,764	3,650	3,540	3,543	3,555	3,552	3,582	3,597	
e_2	1,75E-02	3,50E-03	1,36E-02	6,16E-03	1,15E-02	5,47E-03	6,50E-03	7,95E-05	
y_3	17,000	24,000	22,000	45,000	42,000	26,000	24,000	36,000	
y_{3P}	9,750	15,875	36,250	59,875	39,000	34,000	39,000	35,934	
e_3	4,26E-01	3,39E-01	6,48E-01	3,31E-01	7,14E-02	3,08E-01	6,25E-01	1,84E-03	

Este procedimiento se replica para el resto de programas en estudio, obteniendo los resultados:

3.2.2. Administración de Empresas (D)

Tabla 3.42: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	107,500	107,500	104,100	104,100	106,000	105,700	106,200	105,000	
y_2	3,794	3,793	3,724	3,650	3,727	3,730	3,723	3,697	
y_3	24,000	36,000	31,000	10,000	14,000	18,000	31,000	18,000	
y_{1P}	109,125	106,250	102,350	104,475	105,945	104,923	105,497	105,289	
y_{2P}	3,853	3,809	3,638	3,646	3,725	3,697	3,690	3,706	
y_{3P}	36,500	46,500	15,000	-0,500	8,094	12,531	27,313	18,842	
e_1	1,51E-02	1,16E-02	1,68E-02	3,60E-03	5,16E-04	7,35E-03	6,62E-03	2,75E-03	
e_2	1,54E-02	4,12E-03	2,31E-02	1,07E-03	5,24E-04	8,77E-03	8,92E-03	2,31E-03	
e_3	5,21E-01	2,92E-01	5,16E-01	1,05	4,22E-01	3,04E-01	1,19E-01	4,68E-02	

3.2.3. Administración en Sistemas Informáticos

Tabla 3.43: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	99,700	99,700	98,510	98,510	104,600	104,100	103,900	103,300	
y_2	3,644	3,548	3,443	3,435	3,419	3,474	3,418	3,491	
y_3	13,000	28,000	25,000	23,000	32,000	34,000	23,000	22,000	
y_{1P}	99,700	96,013	95,635	101,510	109,100	106,600	106,400	103,334	
y_{2P}	3,712	3,549	3,377	3,347	3,403	3,478	3,402	3,486	
y_{3P}	16,500	26,000	23,750	30,000	41,500	33,125	20,000	22,163	
e_1	1,00E-18	3,70E-02	2,92E-02	3,05E-02	4,30E-02	2,40E-02	2,41E-02	3,26E-04	
e_2	1,88E-02	2,75E-04	1,93E-02	2,56E-02	4,57E-03	1,12E-03	4,57E-03	1,39E-03	
e_3	2,69E-01	7,14E-02	5,00E-02	3,04E-01	2,97E-01	2,57E-02	1,30E-01	7,39E-03	

3.2.4. Arquitectura

Tabla 3.44: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN Y ERRORES	OBSERVACIONES							
	1	2	3	4	5	6	7	8
y_1	99,400	99,400	103,360	103,360	104,340	102,600	102,700	103,300
y_2	3,701	3,671	3,681	3,600	3,599	3,600	3,625	3,651
y_3	11,000	10,000	12,000	12,000	16,000	19,000	17,000	20,000
y_{1P}	96,275	100,056	104,735	103,860	105,590	103,663	103,794	103,307
y_{2P}	3,736	3,717	3,646	3,584	3,540	3,603	3,632	3,651
y_{3P}	5,500	7,063	9,750	12,000	21,500	25,078	22,797	19,998
e_1	3,14E-02	6,60E-03	1,33E-02	4,84E-03	1,20E-02	1,04E-02	1,06E-02	7,09E-05
e_2	9,50E-03	1,26E-02	9,55E-03	4,34E-03	1,63E-02	9,49E-04	1,89E-03	8,36E-06
e_3	5,00E-01	2,94E-01	1,88E-01	1,00E-18	3,44E-01	3,20E-01	3,41E-01	1,22E-04

3.2.5. Ingeniería Civil

Tabla 3.45: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN Y ERRORES	OBSERVACIONES							
	1	2	3	4	5	6	7	8
y_1	103,300	103,300	99,610	99,610	101,200	108,600	106,200	105,800
y_2	3,651	3,630	3,609	3,603	3,633	3,651	3,674	3,642
y_3	18,000	13,000	23,000	18,000	21,000	13,000	14,000	11,000
y_{1P}	102,902	99,378	96,610	95,610	106,638	111,100	109,075	105,464
y_{2P}	3,645	3,602	3,578	3,589	3,653	3,651	3,685	3,640
y_{3P}	15,289	19,266	25,500	24,000	17,875	7,250	8,188	11,287
e_1	3,86E-03	3,80E-02	3,01E-02	4,02E-02	5,37E-02	2,30E-02	2,71E-02	3,18E-03
e_2	1,54E-03	7,73E-03	8,66E-03	3,79E-03	5,51E-03	1,00E-18	2,92E-03	5,20E-04
e_3	1,51E-01	4,82E-01	1,09E-01	3,33E-01	1,49E-01	4,42E-01	4,15E-01	2,61E-02

3.2.6. Ingeniería Eléctrica

Tabla 3.46: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN Y ERRORES	OBSERVACIONES							
	1	2	3	4	5	6	7	8
y_1	94,100	94,100	87,800	102,500	101,000	107,500	103,800	102,600
y_2	3,689	3,663	3,624	3,535	3,596	3,599	3,640	3,619
y_3	29,000	20,000	18,000	13,000	22,000	20,000	23,000	23,000
y_{1P}	87,225	84,100	81,800	103,000	109,875	112,000	108,800	102,518
y_{2P}	3,742	3,671	3,624	3,500	3,570	3,595	3,638	3,618
y_{3P}	29,016	17,688	12,000	12,922	21,328	22,063	25,391	22,859
e_1	7,31E-02	1,06E-01	6,83E-02	4,88E-03	8,79E-02	4,19E-02	4,82E-02	8,00E-04
e_2	1,44E-02	2,13E-03	1,00E-18	9,95E-03	7,13E-03	1,09E-03	5,37E-04	2,15E-04
e_3	5,39E-04	1,16E-01	3,33E-01	6,01E-03	3,05E-02	1,03E-01	1,04E-01	6,11E-03

3.2.7. Ingeniería Electrónica

Tabla 3.47: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,900	102,900	99,850	99,850	107,000	108,000	107,300	106,500	
y_2	3,806	3,795	3,685	3,669	3,645	3,700	3,743	3,687	
y_3	20,000	27,000	27,000	15,000	21,000	26,000	17,000	19,000	
y_{1P}	99,275	97,463	94,913	100,600	111,375	111,125	111,300	106,742	
y_{2P}	3,889	3,821	3,633	3,595	3,621	3,668	3,727	3,681	
y_{3P}	27,734	33,033	21,156	12,625	24,871	24,125	14,250	18,659	
e_1	3,52E-02	5,28E-02	4,94E-02	7,51E-03	4,09E-02	2,89E-02	3,73E-02	2,27E-03	
e_2	2,17E-02	6,75E-03	1,40E-02	2,02E-02	6,50E-03	8,64E-03	4,17E-03	1,49E-03	
e_3	3,87E-01	2,23E-01	2,16E-01	1,58E-01	1,84E-01	7,21E-02	1,62E-01	1,80E-02	

3.2.8. Ingeniería Física

Tabla 3.48: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	10,630	10,630	10,560	10,770	10,470	10,730	10,530	11,010	
y_2	3,665	3,714	3,583	3,664	3,592	3,478	3,699	3,484	
y_3	17,000	7,000	7,000	13,000	12,000	16,000	13,000	21,000	
y_{1P}	10,560	10,568	10,626	10,571	10,509	11,066	10,874	10,965	
y_{2P}	3,778	3,706	3,630	3,664	3,459	3,337	3,582	3,489	
y_{3P}	11,250	1,500	5,688	10,625	14,375	24,000	21,250	20,539	
e_1	6,61E-03	5,88E-03	6,29E-03	1,85E-02	3,73E-03	3,13E-02	3,26E-02	4,08E-03	
e_2	3,09E-02	2,10E-03	1,31E-02	1,00E-18	3,70E-02	4,04E-02	3,17E-02	1,47E-03	
e_3	3,38E-01	7,86E-01	1,88E-01	1,83E-01	1,98E-01	5,00E-01	6,35E-01	2,19E-02	

3.2.9. Ingeniería Industrial

Tabla 3.49: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	102,800	102,800	101,520	101,520	102,000	108,700	108,300	106,500	
y_2	3,792	3,796	3,740	3,736	3,750	3,763	3,798	3,807	
y_3	10,000	12,000	10,000	8,000	12,000	6,000	12,000	6,000	
y_{1P}	101,800	92,800	99,020	100,520	106,500	112,200	110,300	109,859	
y_{2P}	3,835	3,741	3,702	3,713	3,734	3,798	3,837	3,820	
y_{3P}	11,750	14,500	8,313	9,750	9,125	1,938	9,500	5,363	
e_1	9,73E-03	9,73E-02	2,46E-02	9,85E-03	4,41E-02	3,22E-02	1,85E-02	3,15E-02	
e_2	1,13E-02	1,44E-02	1,02E-02	6,27E-03	4,17E-03	9,34E-03	1,03E-02	3,33E-03	
e_3	1,75E-01	2,08E-01	1,69E-01	2,19E-01	2,40E-01	6,77E-01	2,08E-01	1,06E-01	

3.2.10. Ingeniería Química

Tabla 3.50: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	100,200	100,200	97,950	97,950	106,100	107,500	107,100	107,000	
y_2	3,686	3,627	3,582	3,566	3,544	3,585	3,602	3,622	
y_3	26,000	24,000	16,000	10,000	14,000	26,000	19,000	31,000	
y_{1P}	95,450	94,450	91,825	99,700	111,163	111,750	111,350	107,164	
y_{2P}	3,720	3,643	3,544	3,506	3,519	3,602	3,623	3,618	
y_{3P}	28,813	22,625	4,875	1,500	18,000	37,125	28,750	30,543	
e_1	4,74E-02	5,74E-02	6,25E-02	1,79E-02	4,77E-02	3,95E-02	3,97E-02	1,53E-03	
e_2	9,24E-03	4,38E-03	1,06E-02	1,68E-02	7,18E-03	4,70E-03	5,96E-03	1,11E-03	
e_3	1,08E-01	5,73E-02	6,95E-01	8,50E-01	2,86E-01	4,28E-01	5,13E-01	1,47E-02	

3.2.11. Matemáticas

Tabla 3.51: Predicciones y Errores de Predicción en Y usando KPLSR, $L > 1$

PREDICCIÓN		OBSERVACIONES							
Y ERRORES	1	2	3	4	5	6	7	8	
y_1	91,800	91,800	102,400	104,900	101,800	101,800	108,900	108,900	
y_2	3,338	3,736	3,102	3,567	3,174	3,266	3,508	3,416	
y_3	6,000	6,000	5,000	10,000	5,000	23,000	8,000	26,000	
y_{1P}	79,050	87,800	105,838	105,400	102,847	109,425	115,900	107,650	
y_{2P}	3,627	3,470	3,241	3,382	3,031	3,276	3,563	3,426	
y_{3P}	-2,125	-1,500	2,844	2,250	15,695	39,375	21,000	25,574	
e_1	1,39E-01	4,36E-02	3,36E-02	4,77E-03	1,03E-02	7,49E-02	6,43E-02	1,15E-02	
e_2	8,66E-02	7,11E-02	4,47E-02	5,17E-02	4,51E-02	2,99E-03	1,56E-02	3,07E-03	
e_3	1,35	1,25	4,31E-01	7,75E-01	2,14	7,12E-01	1,63	1,64E-02	

3.2.12. Una aplicación de la técnica KPLS en una función real no lineal, $N = L = 1$

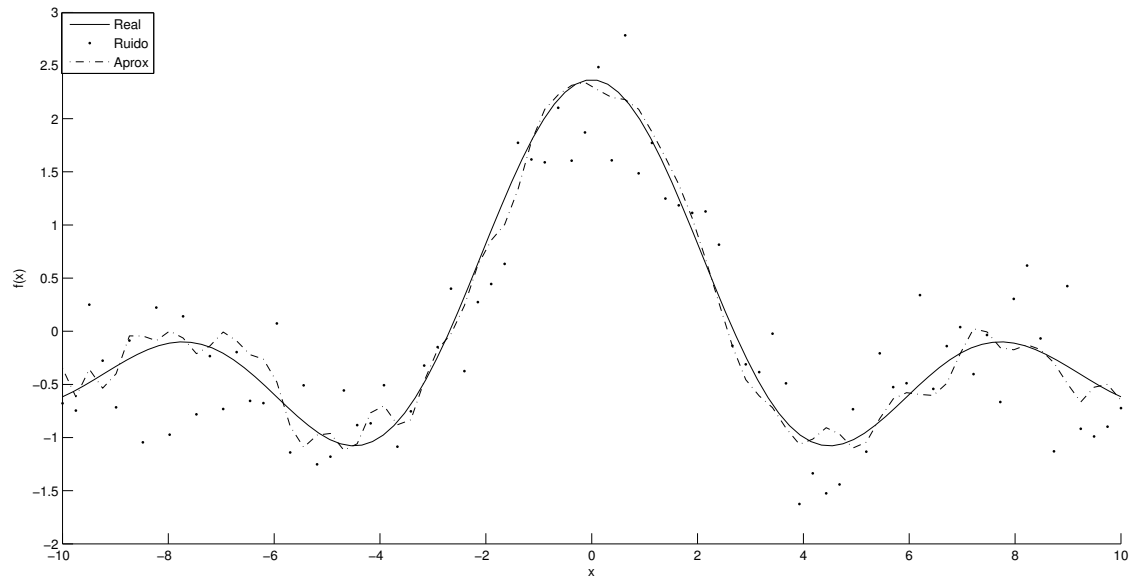
Se presenta un ejemplo en la que exploramos la técnica KPLS y el algoritmo KPLSR, para el caso de la función real no lineal

$$f(x) = \text{sinc}(x) = \frac{\sin|x|}{|x|}$$

Trazamos la gráfica de f tomando 100 puntos (x_i) del intervalo $[-10, 10]$, a estos $f(x_i)$ le aplicamos ruido gaussiano ($\sigma = 0,2$) y calculamos las aproximaciones respectivas usando KPLSR. Luego elegimos 80 puntos de prueba (x_t) en el mismo intervalo (diferentes a los usados en la gráfica de $f(x_i)$) y calculamos los $f(x_t)$ aproximados, obteniendo los resultados de la gráfica (3.7)

3.2. CONSTRUCCIÓN DE MODELOS DE REGRESIÓN MULTIVARIADO USANDO TÉCNICAS NO L

Gráfica 3.9: Aplicando la técnica KPLSR, en una función real no lineal



Capítulo 4

Discusión y Conclusiones

4.1. Sobre los modelos de regresión basados en las técnicas PLS1 y LS

En este punto consideramos el caso Y : Resultados Saber - Pro, como única variable de salida.

Para el programa de Administración de Empresas Nocturna, el modelo obtenido usando la técnica de mínimos cuadrados (LS) presenta en el proceso de validación un error máximo del 4,97 %, lo que implica un porcentaje mínimo de predicción del 95,03 % (de aquí en adelante a esto se le llamará porcentaje de predicción del modelo), es decir, que el intervalo de predicción para esta variable está dado por $0,9503Y_r \leq Y_p \leq 1,0497Y_r$, donde Y_r es el valor real o esperado y Y_p es la predicción de Y_r .

El modelo obtenido con la técnica PLS1 mejora estos resultados predictivos presentando un error máximo de 2,64 %, es decir, el porcentaje de predicción es del 97,36 %. Significa esto que si se espera un resultado de 102 puntos, entonces el valor predicho con la técnica LS está entre 96,9 y 106,9 puntos (aproximadamente), mientras que con la técnica PLS1 oscila entre 99,3 y 104,04 puntos, generando una ganancia en precisión. Ambos modelos son considerados adecuados para efectos predictivos.

Ahora bien, si nos concentramos en el aspecto explicativo las diferencias son notables. Recuerde que la técnica LS, para obtener los coeficientes de peso usa como criterio de optimización el error mínimo entre Y y Y_{fit} ó $Y_{aproximado}$, lo que no permite controlar o establecer por ejemplo qué variables de entrada tienen influencia, explican o están más relacionadas con la variabilidad o el comportamiento de Y (más aún cuando el número de variables excede el número de observaciones). Note que la técnica LS permite obtener un modelo que se expresa en 8 variables originales de las 20 usadas en el estudio, pero ¿Son estas variables las que mejor explican la variabilidad de Y ? ¿Al hacer esta reducción de variables se resuelve el problema de multicolinealidad?. Por su parte, la técnica PLS1 además del aspecto predictivo resuelve en gran medida el aspecto explicativo y de paso el de multicolinealidad. Siguiendo lo expuesto en (1.2.1) y el caso que nos ocupa, la técnica PLS1 reduce la explica-

ción de Y en siete variables latentes ortogonales que son combinación lineal de las originales, es decir, Y queda expresado como

$$Y = c_1 \mathbf{t}_1 + c_2 \mathbf{t}_2 + c_3 \mathbf{t}_3 + \cdots + c_7 \mathbf{t}_7 \quad (4.1)$$

Esto resuelve el problema de multicolinealidad.

$$\mathbf{t}_i = p_{1i} \mathbf{x}_1 + p_{3i} \mathbf{x}_2 + \cdots + p_{20i} \mathbf{x}_{20}, \quad i = 1, \dots, 7 \quad (4.2)$$

donde p_{1i}, \dots, p_{20i} , son los elementos de la primera columna de P^T de la ecuación (1.13). El porcentaje explicativo de cada t_i en Y se resume en la tabla (3.2).

La relación (4.2) nos da una luz sobre como resolver el problema explicativo. Indagamos sobre las variables originales más correlacionadas con las componentes latentes con mayor porcentaje explicativo en Y ($cor(\mathbf{x}_i, \mathbf{t}_j)$), de esta transitividad sabremos en cierta forma qué variables originales aportan más, están influyendo o mejor explican la variabilidad en Y [12]. Las variables originales que cumplan con estas condiciones de ahora en adelante se dirán las **más influyentes** o que **mejor explican** el comportamiento de Y .

En este orden de ideas, se exploró en cada uno de los 11 programas académicos en estudio sobre las variables originales que más influyeron en el comportamiento de los resultados Saber-Pro, durante los periodos 2009-I a 2012-II, obteniendo los siguientes resultados:

Tabla 4.1: Resumen de variables influyentes en Y usando la técnica PLS1R.

N	Programas Académicos	Variables originales más influyentes en Y
1	Administración de Empresas Nocturna	$\mathbf{x}_{*2}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*17}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
2	Administración de Empresas Diurna	$\mathbf{x}_{*7}, \mathbf{x}_{*9}, \mathbf{x}_{*10}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*16}, \mathbf{x}_{*17}, \mathbf{x}_{*20}$.
3	Administración en Sistemas Informáticos	$\mathbf{x}_{*3}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*16}, \mathbf{x}_{*19}, \mathbf{x}_{*20}$.
4	Arquitectura	$\mathbf{x}_{*1}, \mathbf{x}_{*10}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*16}, \mathbf{x}_{*17}, \mathbf{x}_{*18}, \mathbf{x}_{*20}$.
5	Ingeniería Civil	$\mathbf{x}_{*3}, \mathbf{x}_{*4}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*14}, \mathbf{x}_{*16}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
6	Ingeniería Eléctrica	$\mathbf{x}_{*3}, \mathbf{x}_{*5}, \mathbf{x}_{*9}, \mathbf{x}_{*10}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
7	Ingeniería Electrónica	$\mathbf{x}_{*2}, \mathbf{x}_{*3}, \mathbf{x}_{*8}, \mathbf{x}_{*10}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
8	Ingeniería Física	$\mathbf{x}_{*1}, \mathbf{x}_{*4}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*9}, \mathbf{x}_{*15}, \mathbf{x}_{*16}, \mathbf{x}_{*17}$.
9	Ingeniería Industrial	$\mathbf{x}_{*3}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*11}, \mathbf{x}_{*14}, \mathbf{x}_{*16}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
10	Ingeniería Química	$\mathbf{x}_{*1}, \mathbf{x}_{*2}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*16}$.
11	Matemáticas	$\mathbf{x}_{*1}, \mathbf{x}_{*8}, \mathbf{x}_{*9}, \mathbf{x}_{*10}, \mathbf{x}_{*11}, \mathbf{x}_{*16}, \mathbf{x}_{*17}, \mathbf{x}_{*18}$.

Note que los estudios realizados se hacen en forma independiente con cada programa académico, teniendo en cuenta sus particularidades (debilidades y fortalezas). Sin embargo, desde una perspectiva global, vemos que las variables explicativas que con mayor frecuencia se relacionan en los 11 programas, son:

1. Número de grupos de investigación & Número de docentes que lideran grupos de investigación.
2. Productividad Académica.

3. Número de estudiantes que realizan prestamos en biblioteca.
4. Número de estudiantes que acceden a consultas en las bases de datos Sinab.
5. Número de estudiantes vinculados a semilleros de investigación.

Cabe aclarar que el tipo de relación explicativa que aquí se nombra no necesariamente es directa; sin embargo, es una información importante para tener en cuenta en la toma de decisiones de tipo Académica y/o Administrativa. De otra parte, note que la mayoría de las variables explicativas citadas son de tipo escolar interna, es decir, corresponden a procesos de trabajo y apoyo a estudiantes, trabajo docente, procesos académicos e investigativos, lo que refleja el grado de contribución de la Universidad Nacional- Sede Manizales en aporte de **valor agregado** para los estudiantes de estos programas.

Como vemos, la técnica PLS1 genera modelos más eficientes tanto en lo predictivo como en lo explicativo, por lo tanto las exploraciones, estructuración y validación de modelos en los demás programas académicos se hacen usando esta técnica. Los resultados de validación se muestran en las tablas: 3.9, 3.12, 3.15, 3.18, 3.21, 3.24, 3.27, 3.30, 3.33 y 3.36. Note que en todos los programas se obtienen modelos adecuados en lo predictivo, excepto para Ingeniería Civil y Matemáticas, cuyos errores máximos de predicción llegan al 10,4% y 10,9%, respectivamente. Este tipo de comportamiento quizás lo expliquen los fuertes cambios presentados en Y (Ver observaciones 1, 4, 8 en Civil y 1, 2, 8 en Matemáticas).

4.2. Sobre los modelos de regresión usando la técnica PLS2

En este punto consideramos $Y \in \mathbb{R}^3$ ($L = 3$). Las variables de respuesta o salida, son: Resultados Saber - Pro, Promedio Académico y Deserción. Nuevamente se toma como referencia de aplicación al programa **Administración de Empresas Nocturna**.

Como puede apreciarse en la Tabla 3.5, el modelo de regresión obtenido con la técnica PLS2 (PLS2R) muestra buen comportamiento predictivo en las dos primeras variables de salida Saber - Pro (97,36%) y Promedio Académico (99,2%), pero no para la variable Deserción. Este comportamiento quizás se de por dos motivos:

1. Porque la variabilidad de esta última con el paso del tiempo es un poco más fuerte que las dos primeras (Ver Tabla 3.5), lo que posiblemente implique una tendencia no lineal, y en cuyo caso se tendrá que explorar con técnicas no lineales.
2. Por acomodamiento de pesos: Pensamos que el hecho de intentar mejorar predicción en unas variables implicaría desmejoramiento en otras, es decir, que los pesos generados en B para conseguir buen comportamiento predictivo en las variables Saber-Pro y Promedio Académico, afectaron por acomodamiento el buen comportamiento para la variable Deserción. Lo que sugiere un tratamiendo individual.

Una exploración realizada con cada variable de salida, aplicando PLS1R, descarta la segunda posibilidad (compare tablas de predicción 3.3, 3.6 y 3.7 con la tabla 3.5), es decir, un tratamiento en conjunto con PLS2R o de forma individual con PLS1R no resuelve el problema predictivo. Además, puede verificarse que los pesos en B para cada caso se mantienen (compare por ejemplo B de la ecuación (3.3) con B de la ecuación (3.2)). Este comportamiento quizás lo explique la relación $\hat{Y} = TC^T$ descrita en (1.19), que también se expresaría en la forma

$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} T(C_{1*})^T \\ T(C_{2*})^T \\ T(C_{3*})^T \end{bmatrix} \quad (4.3)$$

Donde $(C_{1*})^T$, $(C_{2*})^T$ y $(C_{3*})^T$, son las columnas 1, 2 y 3 de C , respectivamente. Así que, la solución al mejoramiento predictivo en la variable Deserción, estaría de momento por la vía de explorar con técnicas no lineales.

Note que los modelos PLS2R, en todos los programas académicos estudiados, presentan dificultades en el porcentaje de predicción asociado a la variable de salida **deserción** (ver tablas: 3.5, 3.10, 3.13, 3.16, 3.19, 3.22, 3.25, 3.28, 3.31, 3.34, 3.37); además en los programas de Ingeniería Civil y Matemáticas se suma la variable resultados Saber - Pro, tal como se expresó en el caso PLS1R.

Para efectos explicativos, se utiliza la relación (4.3) y procedemos tal como se hizo con los modelos PLS1R. Elegimos las variables originales que más están correlacionadas con las variables latentes más influyentes en la variabilidad de Y . De esta manera se determinan las variables de entrada que mejor explican el comportamiento de los resultados Saber - Pro, Promedio Académico y Deserción, en cada programa durante los periodos 2009-I a 2012-II:

Tabla 4.2: Resumen de variables influyentes en Y usando la técnica PLS2R.

N	Programas Académicos	Variables originales más influyentes en Y
1	Administración de Empresas Nocturna	$\mathbf{x}_{*1}, \mathbf{x}_{*5}, \mathbf{x}_{*11}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*17}$.
2	Administración de Empresas Diurna	$\mathbf{x}_{*2}, \mathbf{x}_{*6}, \mathbf{x}_{*7}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
3	Administración en Sistemas Informáticos	$\mathbf{x}_{*3}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*16}, \mathbf{x}_{*19}, \mathbf{x}_{*20}$.
4	Arquitectura	$\mathbf{x}_{*2}, \mathbf{x}_{*5}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*14}, \mathbf{x}_{*17}, \mathbf{x}_{*19}, \mathbf{x}_{*20}$.
5	Ingeniería Civil	$\mathbf{x}_{*2}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*16}, \mathbf{x}_{*19}$.
6	Ingeniería Eléctrica	$\mathbf{x}_{*2}, \mathbf{x}_{*3}, \mathbf{x}_{*5}, \mathbf{x}_{*10}, \mathbf{x}_{*12}, \mathbf{x}_{*14}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
7	Ingeniería Electrónica	$\mathbf{x}_{*2}, \mathbf{x}_{*5}, \mathbf{x}_{*6}, \mathbf{x}_{*7}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*18}$.
8	Ingeniería Física	$\mathbf{x}_{*6}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*10}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*15}, \mathbf{x}_{*16}$.
9	Ingeniería Industrial	$\mathbf{x}_{*2}, \mathbf{x}_{*7}, \mathbf{x}_{*8}, \mathbf{x}_{*11}, \mathbf{x}_{*16}, \mathbf{x}_{*17}, \mathbf{x}_{*18}, \mathbf{x}_{*19}$.
10	Ingeniería Química	$\mathbf{x}_{*4}, \mathbf{x}_{*5}, \mathbf{x}_{*7}, \mathbf{x}_{*12}, \mathbf{x}_{*14}, \mathbf{x}_{*15}, \mathbf{x}_{*16}, \mathbf{x}_{*19}$.
11	Matemáticas	$\mathbf{x}_{*1}, \mathbf{x}_{*6}, \mathbf{x}_{*10}, \mathbf{x}_{*12}, \mathbf{x}_{*13}, \mathbf{x}_{*14}, \mathbf{x}_{*16}, \mathbf{x}_{*17}$.

De la misma manera que se hizo con PLS1, elegimos las variables explicativas que con mayor frecuencia se relacionan en los 11 programas, obteniendo en su orden:

1. Número de estudiantes que realizan préstamos en biblioteca.

2. Número de grupos de investigación y de docentes que lideran grupos de investigación.
3. Número de docentes que laboran con dedicación exclusiva.
4. Número de estudiantes vinculados a semilleros de investigación.
5. Número de docentes con título de doctorado vinculados al programa.
6. Número de estudiantes que realizan consultas en bases de datos SINAB.

Información aún más interesante para efectos de toma de decisiones y aporte de valor agregado, por cuanto ahora se mira la influencia sobre tres variables de salida, consideradas de gran impacto en calidad educativa.

4.3. Sobre los modelos de regresión usando la técnica PLS2 Linealizado

Tal como se expuso en la sesión 4.2, se aplican métodos de linealización sobre las variables de salida que presentaron dificultades en predicción, con los modelos de regresión lineales PLS2R. El procedimiento utilizado para esta técnica se expuso en la sesión 3.1. Los resultados obtenidos son satisfactorios, esto es, se lograron estructurar modelos con porcentajes predictivos entre el 95 y 99%, sobre el nuevo espacio de variables \tilde{Y} (compare por ejemplo Tabla 3.5 y Tabla 3.8).

En el aspecto explicativo, no se encontraron diferencias notorias sobre las variables explicativas más influyentes encontradas con esta técnica y la técnica PLS2R, es decir, las variables originales de mejor influencia sobre \tilde{Y} , coinciden en la mayoría de los programas con las expuestas en la tabla 4.2. Este comportamiento quizás se deba a la relación directa: $\tilde{y}_i = e^{\sqrt[k]{y_i}}$, utilizada en el proceso de linealización.

Por lo tanto, se resalta el hecho que aunque los modelos PLS2R presentan en algunos casos dificultades en predicción siguen siendo muy importantes en el ámbito explicativo.

4.4. Sobre los modelos de regresión no lineales aplicando la técnica KPLS (KPLSR)

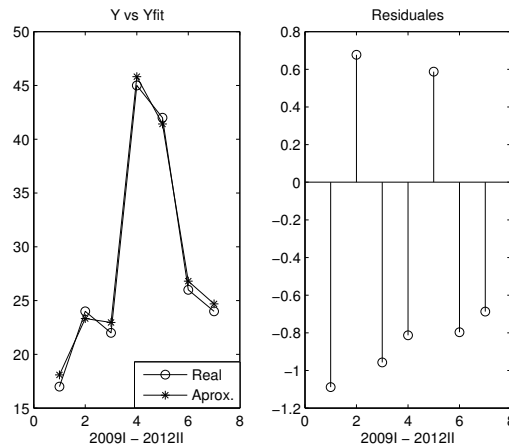
Primero consideremos el caso $L = 1$, Y : Deserción. Tomando como referencia de aplicación al programa **Administración de Empresas Nocturna**.

Aplicamos la técnica KPLSR con el ánimo de construir un modelo de regresión no lineal, con la variable **deserción** como única variable de salida, la cual presentó problemas en predicción con el modelo de regresión lineal PLS1R.

Los resultados en predicción mejoran considerablemente con respecto a lo obtenido con PLS1R (Ver Tabla 3.40). Note que el error máximo con PLS1R es de 1,30 aproximadamente, mientras que con KPLSR es de 0,115; una disminución muy interesante.

Varias exploraciones realizadas con esta técnica, dan pie para pensar que estos resultados predictivos pueden mejorar aún más, si contáramos con más observaciones en el estudio. Fijese que esta técnica aplicada con 8 observaciones genera modelos cuyos residuales son muy pequeños (tal como se puede apreciar en la Gráfica 3.7), pero como es sabido al realizar el proceso de validación cruzada(LOO) debemos prescindir de una de estas observaciones y trabajar solo con 7 de ellas, obteniendo modelos cuyos residuales ya no son tan pequeños (Ver Gráfica 4.1) y por tanto las predicciones con estos parámetros obtenidos no van a ser las deseadas.

Gráfica 4.1: Aplicando KPLSR, $L = 1$ - con 7 observaciones



Aún así, el mejoramiento en lo predictivo comparado con la técnica lineal PLS1 es notorio. Una referencia de la eficiencia de la técnica KPLSR cuando se tienen suficientes observaciones se expuso en **3.2.12**.

En lo que respecta a la aplicación de la técnica KPLSR(salida múltiple ó $L > 1$), los resultados predictivos comparados con PLS2R son interesantes(compare las tablas 3.5 y 3.41). Note que las predicciones mejoran en las variable de salida con más dificultad, característica que se mantiene en todos los programas académicos estudiados. Por ejemplo, el error máximo en la predicción de la variable **deserción** para el programa Administración de empresas nocturna con la técnica PLS2R es de 1,30 y con la técnica KPLSR se reduce a 0,426 aproximadamente, es decir, en más del 50 %.

Observe que cuando usábamos PLS2R e intentamos mejorar los errores de predicción para la variable deserción recurriendo a un tratamiento individual con PLS1R, los resultados se mantuvieron; pero ahora aplicando la técnica KPLSR con $L = 1$ ó $L > 1$ los resultados varían notoriamente(comparar tablas 3.39 y 3.41), quiere decir que con esta técnica un tratamiento individual es más provechoso para objetos predictivos.

4.4. SOBRE LOS MODELOS DE REGRESIÓN NO LINEALES APLICANDO LA TÉCNICA KPLS (KPLS)

Para evidenciar las mejoras en el resto de programas académicos compare los resultados mostrados en las tablas 3.10, 3.13, 3.16, 3.19, 3.22, 3.25, 3.28, 3.31, 3.34, 3.37 con los de las tablas 3.42, 3.43, 3.44, 3.45, 3.46, 3.47, 3.48, 3.49, 3.50, 3.51.

En cuanto a la interpretación de estos modelos, se recurre a la aproximación $Y = TT'Y = TC$, donde $C = T'Y$, descrita en la ecuación (1.26). Para el caso $L = 1$, Y queda expresada en la forma de la ecuación (1.31). En el caso $L > 1$, podemos generalizar la ecuación (1.31), de modo que cada componente de \hat{Y} se puede ver como una combinación lineal de las columnas de T , donde cada t_i es una función no lineal que depende de \mathbf{x} . A diferencia de la técnica PLS2R, aquí solo podremos explorar la relación de Y con las componentes latentes, en el sentido de conocer cuál es el porcentaje explicativo de cada una de ellas en Y , pero no podemos decir mucho sobre la influencia de las variables originales sobre Y , ya que ahora los t_i 's no son combinación lineal de las variables originales.

Bibliografía

- [1] Consejo Nacional de Acreditación; Lineamientos para la acreditación institucional; serie documentos CNA; Bogotá - Colombia; Noviembre de 2006.
- [2] Consejo Nacional de Acreditación; Indicadores para la autoevaluación con fines de acreditación institucional; serie documentos CNA; Segunda Edición; Bogotá - Colombia; Noviembre de 2006.
- [3] Consejo Nacional de Acreditación; Guía para la evaluación externa con fines de acreditación institucional; serie documentos CNA; Segunda Edición; Bogotá - Colombia; Noviembre de 2006.
- [4] Consejo Nacional de Acreditación; Lineamientos para la acreditación de programas de pregrado; serie documentos CNA; Bogotá - Colombia; Enero de 2013.
- [5] Consejo Nacional de Acreditación; Guía de autoevaluación con fines de acreditación de programas de pregrado; serie documentos CNA; Cuarta edición; Bogotá - Colombia; Enero de 2013.
- [6] Alvin C. Rencher; *Methods of Multivariate Analysis*; Second Edition; Wiley- Interscience Publication; United States of America; 2002.
- [7] Alan Julian Izenman; *Modern Multivariate Statistical Techniques -Regression, Classification, and Manifold Learning*; Springer; New York -United States of America; 2013.
- [8] Andrés Álvarez Meza, Genaro Daza, Carlos Acosta Medina y Germán Castellanos D.; *Selección de Parámetros en Mínimos cuadrados- Máquinas de Vectores de Soporte Orientadas a Regresión, Utilizando Validación Cruzada Generalizada*; Dyna Medellin, Febrero de 2012.
- [9] Vincenzo Esposito Vinzi, Wynne W. Chin, Jörg Henseler y Huiwen Wang; *Handbook of Partial Least Squares*; Springer; New York -United States of America; 2010.
- [10] Gabriel Cepeda Carrión y Jose Luis Roldán Salgueiro, *Aplicando en la práctica la técnica PLS en la Administración de empresas*, Facultad de ciencias económicas y empresariales, Sevilla (España), 2004.
- [11] *Partial Least Square(PLS) Methods: Origins, Evolution and Application to Social sciences*; Gregoria Mateos y Aparicio Morales, Universidad de Complutense de Madrid, Facultad de ciencias económicas y empresariales, Campus de Sogamosa, 2011.

- [12] Tenenhaus M., La régression PLS théorie et pratique. Editions Technip, Paris, 1998.
- [13] Hervé Abdi, Partial Least Square(PLS) Regression, Program in Cognition and Neuroscience, University of Texas at Dallas, 2003.
- [14] Roman Rosipal and Nicole Krämer, Overview and Recent Advances in Partial Least Squares, Austrian Research Institute for Artificial Intelligence, Vienna-Austria, Department of Computer Science and Electrical Engineering, Berlin (Germany), 2006.
- [15] A. Höskuldsson. PLS Regression Methods. Journal of Chemometrics,1998.
- [16] Baibing Li , Julian Morris, Elaine B. Martin, Model selection for partial least squares regression, Centre for Process Analytics and Control Technology, University of Newcastle, Newcastle upon Tyne, accepted 24 June 2002(ELSEVIER).
- [17] Roman Rosipal and Leonard J. Trejo, Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space, Applied Computational Intelligence Research Unit University of Paisley Paisley PA1 2BE - Scotland (Slovak Republic) and Computational Sciences Division NASA Ames Research Center Moffett Field, 2001.
- [18] Michinari Momma and Kristin P. Bennett, Sparse Kernel Partial Least Squares Regression, Department of Decision Sciences and Engineering Systems, Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy N Y, USA, 2003.
- [19] K. P. Bennett and M. J. Embrechts, An Optimizacion Perspective on Kernel Partial Least Squares Regression, Department of Decision Sciences and Engineering Systems, Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy N Y, USA, 2003.
- [20] Gilles Blanchard and Nicole Krämer, Kernel Partial Least Squares is Universally Consistent, Weierstrass Institute for Applied Analysis and Stochastics, Berlin Institute of Technology Machine Learning Group, Berlin - Germany, 2010.

ANEXO 1 - TABLA DE DATOS

Tabla 4.3: PROGRAMA: ADMINISTRACIÓN DE EMPRESAS DIURNA

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
<i>2009-I</i>	0,900	216	224	606,2	66	0	280	218	12	11	3
<i>2009-II</i>	0,847	228	231	573,8	84	2	274	221	15	13	1
<i>2010-I</i>	0,984	246	230	566,1	93	4	312	204	16	18	3
<i>2010-II</i>	0,949	242	233	563,4	139	2	334	238	17	12	4
<i>2011-I</i>	0,900	262	208	585,1	85	5	328	241	15	14	3
<i>2011-II</i>	0,873	270	210	554,4	82	7	295	208	11	17	10
<i>2012-I</i>	0,793	275	198	596,7	80	0	332	229	18	23	10
<i>2012-II</i>	0,684	285	191	565,5	70	6	369	250	17	24	5

OBS.	Variable Docentes			Variable Proc. Acad. e Invest.					
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19	X*20
<i>2009-I</i>	26	1	2	84	12	25	13	8	148
<i>2009-II</i>	27	2	2	84	12	25	13	8	148
<i>2010-I</i>	29	2	2	100	6	47	10	7	74,5
<i>2010-II</i>	28	4	2	100	6	47	10	7	74,5
<i>2011-I</i>	27	3	1	76	4	25	4	4	196,03
<i>2011-II</i>	27	3	1	76	4	25	4	4	196,03
<i>2012-I</i>	28	3	1	79	3	17	3	4	53,59
<i>2012-II</i>	28	3	1	79	3	17	3	4	53,59

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
<i>2009-I</i>	107,50	3,794	24
<i>2009-II</i>	107,50	3,793	36
<i>2010-I</i>	104,10	3,724	31
<i>2010-II</i>	104,10	3,650	10
<i>2011-I</i>	106,00	3,727	14
<i>2011-II</i>	105,70	3,730	18
<i>2012-I</i>	106,20	3,723	31
<i>2012-II</i>	105,00	3,697	18

Tabla 4.4: PROGRAMA: ADMINISTRACIÓN DE EMPRESAS NOCTURNA

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
2009-I	0,508	307	138	602,5	25	0	228	141	12	20	12
2009-II	0,547	307	130	598,2	27	2	197	149	15	33	15
2010-I	1,000	304	117	555,6	35	4	200	140	16	29	12
2010-II	0,894	306	126	559,3	48	2	192	165	17	25	17
2011-I	0,955	310	94	575,3	25	5	216	151	15	30	7
2011-II	0,980	311	95	553,5	31	7	168	154	11	37	29
2012-I	0,808	334	98	562,6	34	0	190	202	18	25	15
2012-II	0,768	353	102	545,9	31	6	212	250	17	25	18

OBS.	Variable Docentes			Variable Proc. Acad. e Invest.					
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19	X*20
2009-I	26	1	2	84	12	25	13	8	148
2009-II	27	2	2	84	12	25	13	8	148
2010-I	29	2	2	100	6	47	10	7	74,5
2010-II	28	4	2	100	6	47	10	7	74,5
2011-I	27	3	1	76	4	25	4	4	196,03
2011-II	27	3	1	76	4	25	4	4	196,03
2012-I	28	3	1	79	3	17	3	4	53,59
2012-II	28	3	1	79	3	17	3	4	53,59

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	105,40	3,699	17
2009-II	105,40	3,663	24
2010-I	105,28	3,589	22
2010-II	105,28	3,565	45
2011-I	103,10	3,596	42
2011-II	105,70	3,572	26
2012-I	106,20	3,605	24
2012-II	105,00	3,597	36

Tabla 4.5: PROGRAMA: ADMINISTRACIÓN EN SISTEMAS INFORMÁTICOS

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
<i>2009-I</i>	0,72	227	127	591,3	63	0	219	67	15	8	4
<i>2009-II</i>	0,66	224	121	571,2	58	0	218	84	15	14	5
<i>2010-I</i>	0,86	232	116	582,3	61	0	229	97	15	21	4
<i>2010-II</i>	0,84	216	107	549,8	120	1	193	73	7	17	3
<i>2011-I</i>	0,91	238	82	555,0	53	1	191	157	11	8	4
<i>2011-II</i>	0,84	225	76	559,0	53	0	162	138	10	13	7
<i>2012-I</i>	0,86	267	79	560,5	62	0	189	115	14	9	3
<i>2012-II</i>	0,93	265	84	539,9	49	2	216	92	13	10	7

OBS.	Variable Docentes			Variable Proc. Acad. e Invest.					
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19	X*20
<i>2009-I</i>	13	1	6	41	1	1	2	1	91,50
<i>2009-II</i>	13	2	6	41	1	1	2	1	91,50
<i>2010-I</i>	15	2	7	28	1	17	1	0	72,90
<i>2010-II</i>	15	2	7	28	1	17	1	0	72,90
<i>2011-I</i>	15	3	8	25	2	16	2	3	50,63
<i>2011-II</i>	15	3	8	25	2	16	2	3	50,63
<i>2012-I</i>	16	4	8	22	2	21	2	3	40,00
<i>2012-II</i>	17	5	9	22	2	21	2	3	40,00

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
<i>2009-I</i>	99,70	3,644	23
<i>2009-II</i>	99,70	3,548	52
<i>2010-I</i>	98,51	3,443	56
<i>2010-II</i>	98,85	3,435	39
<i>2011-I</i>	104,60	3,419	29
<i>2011-II</i>	104,10	3,474	23
<i>2012-I</i>	103,90	3,418	57
<i>2012-II</i>	103,30	3,491	34

Tabla 4.6: PROGRAMA: ARQUITECTURA

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
2009-I	0,870	197	211	645,6	76	1	315	19	6	6	2
2009-II	0,915	205	213	594,8	75	1	319	26	7	10	8
2010-I	0,889	213	208	615,4	63	1	309	33	4	17	3
2010-II	0,949	206	202	591,5	93	3	316	23	6	18	7
2011-I	0,891	247	199	633,7	66	11	354	34	10	13	7
2011-II	0,939	233	184	613,8	49	8	307	23	13	12	7
2012-I	0,961	253	177	624,5	44	6	354	66	10	17	2
2012-II	0,765	248	195	614,9	39	5	401	109	8	14	7

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.				
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19	X*20
2009-I	23	4	24	3	3	3	7	4	102,00
2009-II	23	5	24	3	3	3	7	4	102,00
2010-I	26	6	24	3	0	0	0	4	73,50
2010-II	27	6	23	3	0	0	0	4	73,50
2011-I	26	6	22	3	1	2	1	4	154,39
2011-II	26	6	22	3	1	2	1	4	154,39
2012-I	26	6	22	3	1	5	1	5	134,08
2012-II	27	7	19	6	1	5	1	5	134,08

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	99,40	3,701	13
2009-II	99,40	3,671	43
2010-I	103,36	3,681	48
2010-II	103,36	3,600	55
2011-I	104,34	3,599	35
2011-II	102,60	3,600	41
2012-I	102,70	3,625	36
2012-II	103,30	3,651	30

Tabla 4.7: PROGRAMA: INGENIERÍA CIVIL

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
2009-I	0,897	271	232	652,8	107	1	406	72	20	9	5
2009-II	0,909	269	238	622,2	124	0	380	64	21	21	4
2010-I	0,983	275	236	643,9	102	0	406	87	29	14	4
2010-II	0,873	272	228	631,5	165	1	385	73	17	10	7
2011-I	0,863	314	203	666,7	104	3	402	69	17	12	11
2011-II	0,855	325	211	661,3	104	1	376	93	19	18	8
2012-I	0,917	345	218	672,1	102	1	444	133	18	11	5
2012-II	0,871	355	219	652,2	96	2	512	173	17	7	11

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.			
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*20
2009-I	15	5	11	8	1	5	2	175
2009-II	14	5	11	7	1	5	2	175
2010-I	16	6	11	8	2	10	3	108
2010-II	15	6	11	8	2	10	3	108
2011-I	14	5	11	7	0	0	2	63,51
2011-II	15	6	11	7	0	0	2	63,51
2012-I	17	8	11	7	4	4	2	148,33
2012-II	17	8	8	10	4	4	2	148,33

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	103,30	3,651	19
2009-II	103,30	3,630	64
2010-I	99,61	3,609	54
2010-II	99,61	3,603	56
2011-I	101,20	3,633	21
2011-II	108,60	3,651	37
2012-I	106,20	3,674	42
2012-II	105,80	3,642	53

Tabla 4.8: PROGRAMA: INGENIERÍA ELÉCTRICA

OBS.	Variable Estudiantes									
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10
<i>2009-I</i>	0,738	295	133	599,6	114	349	113	20	6	4
<i>2009-II</i>	0,746	293	122	580,9	124	316	50	21	7	9
<i>2010-I</i>	0,901	301	120	590,6	101	329	86	29	6	4
<i>2010-II</i>	0,984	295	127	572,1	180	304	89	17	9	12
<i>2011-I</i>	0,851	316	100	587,8	106	315	61	17	14	9
<i>2011-II</i>	0,742	317	106	581,0	99	272	101	19	15	23
<i>2012-I</i>	0,882	339	108	595,2	113	317	120	18	22	13
<i>2012-II</i>	1,000	361	113	565,2	107	362	139	17	17	12

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.				
	X*11	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19
<i>2009-I</i>	23	8	10	10	6	33	8	35	270,9
<i>2009-II</i>	23	8	10	10	6	33	8	35	270,9
<i>2010-I</i>	24	9	10	10	2	5	2	31	289,5
<i>2010-II</i>	23	8	8	11	2	5	2	31	289,5
<i>2011-I</i>	23	8	8	11	3	14	3	21	572,54
<i>2011-II</i>	23	9	8	11	3	14	3	21	572,54
<i>2012-I</i>	23	10	8	11	4	21	4	23	529,13
<i>2012-II</i>	23	11	6	13	4	21	4	23	529,13

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
<i>2009-I</i>	94,10	3,689	13
<i>2009-II</i>	94,10	3,663	41
<i>2010-I</i>	87,80	3,624	45
<i>2010-II</i>	102,50	3,535	52
<i>2011-I</i>	101,00	3,596	32
<i>2011-II</i>	107,50	3,599	32
<i>2012-I</i>	103,80	3,640	40
<i>2012-II</i>	102,60	3,619	32

Tabla 4.9: PROGRAMA: INGENIERÍA ELECTRÓNICA

OBS.	Variable Estudiantes									
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10
2009-I	0,561	194	169	654,2	81	244	68	6	19	6
2009-II	0,500	192	162	636,3	85	248	61	7	27	4
2010-I	0,928	209	161	605,2	92	261	48	4	14	12
2010-II	0,766	206	156	595,6	157	241	63	6	12	8
2011-I	0,809	223	147	652,7	86	225	57	10	10	8
2011-II	0,810	226	146	593,6	86	228	65	13	10	15
2012-I	0,940	266	139	622,1	117	280	118	10	18	17
2012-II	0,828	271	140	572,2	106	332	171	8	19	18

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.				
	X*11	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19
2009-I	23	8	10	10	6	33	8	35	270,9
2009-II	23	8	10	10	6	33	8	35	270,9
2010-I	24	9	10	10	2	5	2	31	289,5
2010-II	23	8	8	11	2	5	2	31	289,5
2011-I	23	8	8	11	3	14	3	21	572,54
2011-II	23	9	8	11	3	14	3	21	572,54
2012-I	23	10	8	11	4	21	4	23	529,13
2012-II	23	11	6	13	4	21	4	23	529,13

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	102,90	3,806	16
2009-II	102,90	3,795	59
2010-I	99,85	3,685	60
2010-II	99,85	3,669	55
2011-I	107,00	3,645	34
2011-II	108,00	3,700	39
2012-I	107,30	3,743	34
2012-II	106,50	3,687	45

Tabla 4.10: PROGRAMA: INGENIERÍA FÍSICA

OBS.	Variable Estudiantes									
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10
2009-I	0,623	91	51	576,1	27	0	116	47	5	0
2009-II	0,623	76	45	576,1	38	0	98	41	5	3
2010-I	0,737	99	53	564,2	32	0	118	40	4	1
2010-II	0,737	73	46	564,2	45	1	94	42	5	2
2011-I	0,846	111	56	592,2	33	0	131	45	5	4
2011-II	0,754	124	58	577,6	29	0	132	42	6	16
2012-I	0,877	127	67	604,1	47	0	153	48	15	9
2012-II	0,585	130	65	562,8	37	0	174	54	22	6

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.				
	X*11	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19
2009-I	23	10	12	5	12	14	18	12	327,9
2009-II	24	11	12	5	12	14	18	12	327,9
2010-I	26	13	13	5	2	4	4	14	434,8
2010-II	26	14	13	5	2	4	4	14	434,8
2011-I	26	14	9	9	3	7	3	19	524,32
2011-II	26	14	9	9	3	7	3	19	524,32
2012-I	26	14	9	9	10	35	11	17	631,97
2012-II	26	15	7	11	10	35	11	17	631,97

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	10,63	3,665	13
2009-II	10,63	3,714	26
2010-I	10,56	3,583	34
2010-II	10,77	3,664	24
2011-I	10,47	3,592	30
2011-II	10,73	3,478	55
2012-I	10,53	3,699	48
2012-II	11,01	3,484	50

Tabla 4.11: PROGRAMA: INGENIERÍA INDUSTRIAL

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
2009-I	0,909	186	264	655,0	62	0	325	97	12	9	3
2009-II	0,827	200	262	642,3	71	4	314	98	14	17	7
2010-I	0,926	208	256	666,4	75	1	327	119	14	22	3
2010-II	0,875	196	255	645,6	120	4	299	142	12	21	7
2011-I	0,860	221	226	695,7	46	4	326	141	12	15	8
2011-II	0,778	238	232	650,1	46	4	285	167	11	11	13
2012-I	1,018	251	240	661,1	67	0	340	194	13	12	7
2012-II	0,830	249	240	639,7	56	1	395	221	16	15	14

OBS.	Variable Docentes				Var. Proc. Acad. e Invest.				
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19	X*20
2009-I	13	7	10	6	3	13	6	5	263,1
2009-II	14	7	10	6	3	13	6	5	263,1
2010-I	15	7	10	6	5	27	8	6	318,5
2010-II	18	10	7	8	5	27	8	6	318,5
2011-I	17	9	7	7	1	4	1	5	360,38
2011-II	17	9	7	7	1	4	1	5	360,38
2012-I	17	9	7	7	1	5	1	8	253,11
2012-II	17	9	7	7	1	5	1	8	253,11

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2008-I	100,70	3,600	12
2008-II	100,70	3,623	10
2009-I	102,80	3,792	8
2009-II	102,80	3,796	66
2010-I	101,52	3,740	39
2010-II	101,52	3,736	48
2011-I	102,00	3,750	21
2011-II	108,70	3,763	38
2012-I	108,30	3,798	32
2012-II	106,50	3,807	38

Tabla 4.12: PROGRAMA: INGENIERÍA QUÍMICA

OBS.	Variable Estudiantes										
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10	X*11
2009-I	0,579	208	148	664,7	112	0	323	221	23	11	2
2009-II	0,407	203	149	658,2	114	0	312	206	20	21	3
2010-I	0,904	228	149	615,6	82	0	339	238	10	15	1
2010-II	0,787	233	149	598,3	169	2	351	233	10	8	3
2011-I	0,821	248	148	613,1	112	2	354	276	12	11	9
2011-II	0,700	242	138	578,8	110	0	333	246	14	18	13
2012-I	0,864	251	149	653,7	111	0	364	328	16	25	4
2012-II	1,000	267	136	598,2	97	1	395	410	16	25	8

OBS.	Variable Docentes			Var. Proc. Acad. e Invest.				
	X*12	X*13	X*14	X*15	X*16	X*17	X*18	X*19
2009-I	13	7	12	8	16	13	28	321,9
2009-II	13	7	12	8	16	13	28	321,9
2010-I	14	8	13	5	5	8	28	365,5
2010-II	14	8	13	5	5	8	28	365,5
2011-I	14	8	13	4	10	4	16	332,72
2011-II	14	8	13	4	10	4	16	332,72
2012-I	14	8	13	7	18	7	18	359,23
2012-II	14	9	13	7	18	7	18	359,23

OBS.	Variables de Salida		
	W*1	W*2	W*3
2009-I	100,20	3,686	12
2009-II	100,20	3,627	46
2010-I	97,95	3,582	66
2010-II	97,95	3,566	52
2011-I	106,10	3,544	43
2011-II	107,50	3,585	34
2012-I	107,10	3,602	59
2012-II	107,00	3,622	47

Tabla 4.13: PROGRAMA: MATEMÁTICAS

OBS.	Variable Estudiantes									
	X*1	X*2	X*3	X*4	X*5	X*6	X*7	X*8	X*9	X*10
2009-I	0,538	38	14	569,5	14	46	11	23	0	1
2009-II	0,538	24	8	569,5	9	30	9	20	2	1
2010-I	0,692	47	19	557,1	8	52	13	10	4	14
2010-II	0,692	28	10	557,1	16	27	10	10	8	2
2011-I	0,692	53	21	567,1	15	50	8	12	6	1
2011-II	0,657	50	16	572,9	19	49	10	14	2	9
2012-I	0,750	55	15	586,6	20	55	13	16	8	4
2012-II	0,667	65	21	547,9	13	61	16	16	8	6

OBS.	Variable Docentes				Variable P. Acad. e Inv.			
	X*11	X*12	X*13	X*14	X*15	X*16	X*17	X*18
2009-I	13	3	15	2	0	0	1	146,5
2009-II	13	3	15	2	0	0	1	146,5
2010-I	14	3	15	2	2	3	4	324,4
2010-II	14	3	15	2	2	3	4	324,4
2011-I	14	4	9	8	5	7	4	226,59
2011-II	14	5	9	8	5	7	4	226,59
2012-I	14	6	9	8	3	7	10	224
2012-II	14	6	4	13	3	7	10	224

OBS.	Variables de Salida		
	Y*1	Y*2	Y*3
2009-I	91,80	3,338	11
2009-II	91,80	3,736	6
2010-I	102,40	3,102	37
2010-II	104,90	3,567	9
2011-I	101,80	3,174	28
2011-II	101,80	3,266	31
2012-I	108,90	3,508	16
2012-II	108,90	3,416	26