



Universidad Veracruzana

UNIVERSIDAD VERACRUZANA  
FACULTAD DE MATEMÁTICAS  
Doctorado en Matemáticas

---

Análisis de Componentes Principales con  
Kernels: Una Propuesta de Mejora del  
Kernel

Que para obtener el grado en  
Doctor en Matemáticas

Presenta

Julia Aurora Montano Rivas

Director de Trabajo Recepcional

Dr. Sergio Francisco Juárez Cerrillo

Xalapa, Equez., Ver. Febrero 2013

## **Agradecimientos**

*Señor Jesús, gracias por acompañarme en el caminar de mi vida*

*Mi más profundo agradecimiento a mi Director de tesis Dr. Sergio Francisco Juárez Cerrillo, por su apoyo y sus invaluable consejos para la culminación de este documento.*

*A los Miembros del Jurado, así como Revisores:*

*Dr. Mario Cantú Sifuentes*

*Dra. Gladis Lynares Fleites*

*Dr. Sergio Hernández González*

*Dr. Pedro Ernesto Menéndez Acuña*

*Dr. Francisco Sergio Salem Silva*

*Dra. Hortensia Reyes Cervantes*

*El recibir de ustedes un regalo tan valioso como su tiempo es motivo de alegría, porque durante la revisión surgieron valiosas aportaciones que ayudaron a detallar notablemente este documento, muchas gracias.*

*Por su apoyo y paciencia al Coordinador del Doctorado Dr. Francisco Gabriel Hernández Zamora. Gracias.*

*A mis maestros, amigos y compañeros que de alguna manera influyeron para que este proyecto sea ahora una realidad.*

## **Dedicatoria**

*Esta tesis es dedicada a las personas que más amo*

*Pedro y Margarita, ¡lo logramos!*

*A LA MEMORIA de mi madre quien me enseñó el amor por el estudio*

*Manuela Rivas Jiménez*

*Esfuézate por obtener el conocimiento porque es el verdadero tesoro del hombre; si lo comprendes tendrás la sabiduría y si mantienes la chispa de entusiasmo al profundizar en el conocimiento científico alcanzaras la plenitud.*

*Aurora*

## Resumen

Una extensión del Análisis de Componentes Principales (ACP) es el Análisis de Componentes Principales con Kernels (ACPK). Esta técnica consiste básicamente en enviar los datos originales, mediante alguna transformación no lineal, a un espacio de Hilbert de alta dimensión y realizar el ACP en este espacio. La clave del éxito del ACPK está en lograr la extracción de direcciones de máxima variabilidad en el espacio de los datos transformados y luego identificar estas direcciones con los patrones de variabilidad de los datos originales. Sin embargo, existen situaciones en las cuales el ACPK no es suficiente para detectar estas direcciones de máxima variabilidad. En este trabajo abordamos este problema con la siguiente propuesta: Construimos un espacio convexo de kernels formado por el conjunto de todas las combinaciones lineales convexas de un conjunto fijo de kernels. En este espacio encontramos el kernel óptimo definido por aquel que proporciona el porcentaje de varianza relativa explicada más alto después de hacer el ACPK. Este problema de optimización se traduce en encontrar al conjunto de coeficientes de la combinación lineal convexa que determinan a este kernel óptimo. El problema de optimización convexa lo resolvemos con un algoritmo genético como método de optimización. La propuesta se ilustra con los datos de Indicadores Socioeconómicos (construidos con la información del II Censo de Población y Vivienda 2005 y la Encuesta Nacional de Ocupación y Empleo), de los 210 Municipios del Estado de Veracruz. Con estos indicadores se construye el índice de marginación, el cual es el primer componente principal de un ACP en la matriz de datos. La proporción de varianza explicada por este componente principal es de 56%. Con la propuesta de investigación, el primer componente principal del ACPK extrae 99% de la variabilidad.

**Palabras clave:** Función kernel, Análisis de componentes principales con kernel, Algoritmos genéticos.

## Abstract

Principal Component Analysis with Kernels (ACPK) is an extension of the Principal Component Analysis (PCA). ACPK is basically an ACP on the original data after they were sent, via a non-linear transformation, to a Hilbert space of high dimension. The key for a successful ACPK is to extract directions of maximum variability in the data transformed space and then identify these directions with patterns of maximum variability of the original data. However, there are situations for which ACPK is not sufficient to detect these directions maximum variability. In this thesis we address this problem with the following proposal: build a convex space of kernels obtained from the set of all convex linear combinations of a fixed set of kernels. In this space we find the optimal kernel defined by that which produces the largest percentage of variance explained by an ACPK. This optimization problem consists of finding the coefficients of the convex linear combination of the optimal kernel. We solve the convex optimization problem with a genetic algorithm as optimization method. The proposal is illustrated with data on socioeconomic variables (built with data from the Second National Survey of Households Population, Occupation and Employment, 2005) of the 210 municipalities in the State of Veracruz. These variables are used to construct the Marginality Index, which is the first principal component of a PCA on the data matrix. The percentage of variance explained by the first component is 56%. With the research proposal, the first principal component of ACPK extracted 99% of the variability.

**Key Words:** Kernel functions, kernel principal component analysis, genetic algorithms.

# Contenido

Introducción .....	1
Capítulo 1. Análisis de Componentes Principales con Kernels.....	8
1.1 Patrones de Máxima Variabilidad en Espacios Euclidianos: ACP.....	8
1.1.1 Interpretación Algebraica del ACP.....	11
1.1.2 Interpretación Geométrica del ACP.....	12
1.1.3 Distribuciones Elípticas y ACP .....	16
1.2 Patrones de Máxima Variabilidad en Espacios de Hilbert: ACPK.....	18
1.2.1 Fases del ACPK.....	21
1.3 Construcción de $\mathcal{H}$ a partir de $\Phi$ .....	22
1.3.1 Funciones Kernel .....	26
Capítulo 2. Mejora del Kernel en ACPK.....	28
2.1 Algoritmos Genéticos .....	28
2.1.1 Constitución de un Cromosoma.....	29
2.2 Propuesta de Mejora del Kernel para ACPK.....	33
Capítulo 3. Ejemplo y Aplicación.....	39
Capítulo 4. Conclusión e Investigación Futura.....	52
Referencias.....	54
Apéndice .....	55

## Notación

$n$	Tamaño de muestra de entrenamiento
$x_i$	Variable medida al sujeto de la muestra
$\mathbf{x}$	Vector de variables medidas al sujeto de la muestra
$p$	Número total de variables
$\mathcal{V}$	Espacio de observaciones
$\mathbf{X}$	Matriz de datos originales
$\lambda_i$	Eigenvalor $i$
$v_i$	Eigenvector $i$
$z_i$	Componente principal $i$
$\Phi$	Función no lineal o mapeo
$k$	Función kernel
$\mathcal{H}$	Espacio de Hilbert
$\mathbf{G}$	Matriz Gram
$\mathbf{K}$	Matriz Kernel
$\mathcal{K}$	Conjunto convexo
$\mathcal{G}$	Conjunto de matrices Gram
$\theta$	Vector de ponderación de la función kernel
$\delta$	Número de funciones kernel



## Introducción

Las técnicas tradicionales de análisis multivariado funcionan adecuadamente bajo ciertas configuraciones geométricas de los datos. Por ejemplo, en el problema de discriminación es ideal que los grupos se puedan separar mediante hiperplanos. En el análisis de componentes principales (ACP) se busca representar a los datos en un subespacio vectorial de dimensión 1 o 2 mediante la construcción de variables que son combinaciones lineales de las variables originales, llamadas componentes principales, con la restricción de que estas nuevas variables tengan varianza máxima y estén incorrelacionadas. El ACP funciona adecuadamente si los datos forman una nube de puntos con forma de hiperelipsoide.

En la medida que los patrones ideales no se presenten en los datos es que estas técnicas no proporcionarán resultados adecuados. Por ejemplo en el ACP se necesitarán mucho más de 2 componentes principales para lograr una representación adecuada de los datos en el sentido de varianza explicada.

Para las situaciones mencionadas en el párrafo anterior se han propuesto una diversidad de alternativas. Una relativamente reciente es la propuesta de Schölkopf, Smola y Muller (1998). Estos autores introducen una nueva clase de algoritmos para técnicas de análisis multivariado como el análisis discriminante, el análisis de componentes principales y el análisis de correlación canónica. Estos algoritmos se basan en la idea, aparentemente paradójica de transformar a los datos, mediante una función no lineal, hacia un espacio de mayor dimensión al espacio en el que se encuentran los datos y realizar el análisis multivariado en los datos transformados. La apuesta es que la técnica si funcione con los datos transformados.

Lo mencionado anteriormente se formaliza así: Sea  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})' \in \mathcal{V} \subseteq \mathbb{R}^p$  el vector de características o variables observadas en el objeto  $i$ , de modo que  $x_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, p$ ) es la característica (variable)  $j$  observada en el objeto  $i$ . Al espacio  $\mathcal{V}$  se le llamará el espacio original. Sea  $\Phi: \mathbb{R}^p \rightarrow \mathcal{H}$  un mapeo no lineal, donde  $\mathcal{H}$  es un espacio de Hilbert de dimensión mayor que  $p$  y por lo tanto de mayor dimensión que la del espacio original  $\mathcal{V}$ . La técnica de análisis multivariado se aplica a los datos transformados  $\Phi(\mathbf{x}_i)$  con la apuesta de que funcionará exitosamente en el espacio  $\mathcal{H}$  después de que el análisis fracasó en el espacio original.

Para ilustrar la idea anterior, considérese el problema de clasificación de los dos grupos que se muestran en la Figura 1, véase Mika (2002), Schölkopf y Smola (2002). Un grupo está formado por las observaciones en el círculo interno y el otro por las observaciones que rodean al círculo. Ningún clasificador lineal, no lineal, y no paramétrico podrá separar ambos grupos.

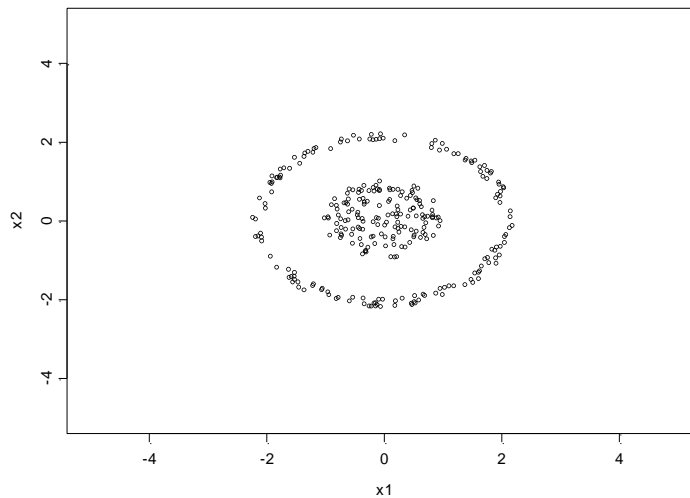


Figura 1. Datos de la “cáscara de naranja”.

Ahora se transforma a las observaciones con la función no lineal  $\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3$  dada por  $(x_1, x_2) \mapsto (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ . El patrón resultante se muestra en la Figura 2.

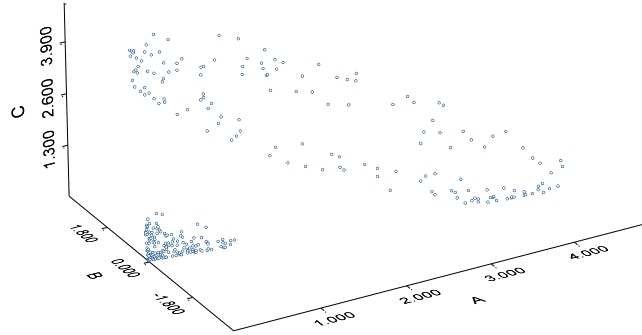


Figura 2. Datos de la “cáscara de naranja” después de haberse transformado con el mapeo no lineal  $(x_1, x_2) \mapsto (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ .

Los datos transformados se pueden separar con la función lineal discriminante de Fisher. Supóngase que los datos originales están en  $\mathbb{R}^3$  y que se consideran monomios de orden 3 con constantes conocidas  $a_i$ 's

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} a_1 x_1^3 \\ a_2 x_2^3 \\ a_3 x_3^3 \\ a_4 x_1^2 x_2 \\ a_5 x_1^2 x_3 \\ a_6 x_1 x_2^2 \\ a_7 x_1 x_3^2 \\ a_8 x_2^2 x_3 \\ a_9 x_2 x_3^2 \\ a_{10} x_1 x_2 x_3 \end{bmatrix}$$

Ahora se va de dimensión 3 a dimensión 10. En general, supóngase que los datos están en  $\mathbb{R}^d$  y que se usan los monomios de orden  $p$  como mapeo no lineal, se puede probar que este mapeo va a  $\mathbb{R}^N$  con

$$N = \binom{d + p - 1}{d}.$$

Por ejemplo, si  $d = 16$  y  $p = 4$ , entonces

$$N = \binom{16 + 4 - 1}{16} = 34,884.$$

La dimensión del espacio de las características es demasiado grande para cuestiones de implementación computacional. Así, la idea simplemente no es factible y por lo tanto la idea no es buena. Efectivamente, en general no lo es. Sin embargo, hay situaciones en las cuales si puede funcionar. Regresando al ejemplo de los datos de la cáscara de naranja, considérense dos puntos  $\mathbf{x} = (x_1, x_2)$  y  $\mathbf{y} = (y_1, y_2)$  y sus imágenes bajo  $\Phi$

$$\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2),$$

$$\Phi(\mathbf{y}) = (y_1^2, \sqrt{2}y_1y_2, y_2^2).$$

Ahora se calcula el producto punto usual en  $\mathbb{R}^3$  entre las imágenes

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 = (x_1y_1 + x_2y_2)^2 = (\langle \mathbf{x}, \mathbf{y} \rangle)^2.$$

De modo que la función  $k$  dada por  $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle)^2$ , donde  $\langle \cdot, \cdot \rangle$  denota el producto punto usual de  $\mathbb{R}^2$ , calcula los productos punto en  $\mathbb{R}^3$  de las imágenes  $\Phi(\mathbf{x})$  y  $\Phi(\mathbf{y})$  de

$\mathbf{x}$  y  $\mathbf{y}$ , respectivamente. Nótese que la función  $k$  evalúa estos productos implícitamente en el sentido de que no requiere conocer al mapeo no lineal  $\Phi$ .

En general, se puede mostrar que la función  $k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle)^d$  calcula el producto punto en el espacio de los monomios de orden  $d$  entre las imágenes  $\Phi(\mathbf{x})$  y  $\Phi(\mathbf{y})$  de  $\mathbf{x}$  y  $\mathbf{y}$ , respectivamente. Es decir  $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$ . La función  $k$  es un ejemplo de lo que se llama *kernel*.

Para generalizar aún más la idea anterior, supóngase que se dispone de una función  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_+$  que satisface  $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j)$  ( $i, j = 1, \dots, n$ ), donde  $\langle \cdot, \cdot \rangle$  es el producto interno de  $\mathcal{H}$ . Es decir, la función  $k$  calcula los productos internos en  $\mathcal{H}$  de los datos transformados. Supóngase también que la técnica de análisis multivariado se puede formular mediante un algoritmo computacional que sólo necesita al producto punto usual de  $\mathbb{R}^p$  evaluado en todos los pares de datos  $\mathbf{x}_i$  y  $\mathbf{x}_j$ . Si en el algoritmo computacional se sustituyen los productos puntos  $\mathbf{x}_i' \mathbf{x}_j = \sum_{s=1}^p x_{is} x_{js}$  por los productos internos  $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$  entonces se estará realizando la aplicación de la técnica de análisis multivariado a los datos mapeados en  $\mathcal{H}$ .

Nótese que en el procedimiento descrito, la función  $k$  permite implementar el algoritmo sin necesidad de especificar explícitamente al mapeo  $\Phi$ . Este procedimiento se conoce como el *truco kernel*.

El truco kernel permite realizar las operaciones algebraicas en el espacio de los datos transformados de manera eficiente y sin conocer a la transformación  $\Phi$ . Así, en principio, cualquier técnica de análisis multivariado para datos en  $\mathcal{V} \subseteq \mathbb{R}^p$  que se pueda formular en un algoritmo computacional en términos de productos punto se puede generalizar a los datos transformados  $\Phi(\mathbf{x}_i) \in \mathcal{H}$  utilizando el truco kernel.

Dos consideraciones que se deben hacer son: i) si es posible construir un espacio de producto interno  $\mathcal{H}$  y un mapeo no lineal  $\Phi$  hacia este espacio y si existe una función  $k$  con la propiedad  $\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = k(\mathbf{x}, \mathbf{y})$  y ii) cómo construir un espacio  $\mathcal{H}$  dado un kernel  $k$ . Se volverá a estos aspectos más adelante.

La aplicación del truco kernel al análisis discriminante de Fisher, al análisis de componentes principales y a la correlación canónica inició a finales de los 90's del siglo pasado. Estos procedimientos se desarrollaron como generalizaciones no lineales de las técnicas clásicas de análisis multivariado. A partir de estos desarrollos el truco kernel ha tenido una influencia considerable en el aprendizaje automático, el reconocimiento de patrones y el aprendizaje estadístico.

Se han propuesto varios kernels en la literatura. Esto ha motivado investigación para determinar de manera adecuada (en un sentido bien definido matemáticamente) el kernel para el problema bajo estudio, sea éste problema de clasificación o de reducción de dimensión. Kim, Magnani y Boyd (2006) proponen un método para seleccionar el kernel óptimo en análisis discriminante de Fisher con kernel. Estos autores consideran el problema de encontrar el kernel sobre un conjunto convexo de kernels y formulan el problema de selección de kernel como un problema de optimización convexa el cual resuelven numéricamente con Matlab y lo prueban empíricamente con datos simulados.

En esta tesis se presenta una propuesta para construir un kernel en el ACP con kernels. La propuesta consiste en determinar un kernel óptimo dentro de un conjunto convexo de kernels. El criterio de optimalidad es la proporción de varianza explicada proporcionada por el kernel. El problema se formula como un problema de optimización convexa el cual, dado su complejidad numérica, se resuelve con un algoritmo genético.

Este trabajo está estructurado de la siguiente forma. En el Capítulo 1 se presenta una revisión del ACP con kernels. En el Capítulo 2 se presenta la propuesta de mejora de kernel para el ACPK. En este capítulo también se incluye una breve presentación de los algoritmos genéticos como una herramienta de optimización. En el Capítulo 3 se presenta un ejemplo de la propuesta, donde se usan los datos de indicadores socioeconómicos de los municipios del Estado de Veracruz. En el Capítulo 4 se presentan las conclusiones y comentarios; así como posibles investigaciones futuras. En el apéndice se incluyen los algoritmos en S-Plus, compatibles con R, que implementan la propuesta, así como los datos de los 210 municipios del estado de Veracruz.

# Capítulo 1. Análisis de Componentes Principales con Kernels

El Análisis de Componentes Principales (ACP) es una técnica de análisis de datos que permite identificar patrones de máxima variabilidad en las observaciones y representar estos patrones en espacios de baja dimensión (una o dos dimensiones). En la estadística el ACP tradicionalmente se ha utilizado para (Johnson y Wichern, 1999):

- Reducir la dimensión del los datos.
- Descubrir e interpretar dependencias entre variables.
- Descubrir patrones de agrupación entre individuos.
- Evaluar normalidad multivariada.
- Detectar valores atípicos (outliers).

En esta tesis se hará énfasis en el ACP como una técnica de reconocimiento de patrones de máxima variabilidad en el espacio original donde viven los objetos bajo estudio.

## 1.1 Patrones de Máxima Variabilidad en Espacios Euclidianos: ACP

Considere un conjunto de  $n$  individuos a los que se les observan  $p$  variables  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip}) \in \mathcal{V} \subseteq \mathbb{R}^p$ ,  $i = 1, \dots, n$ . Estos datos se arreglan en la matriz

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix}.$$



Considérese, sin pérdida de generalidad, que las variables  $\mathbf{X}_j = (x_{1j}, x_{2j}, \dots, x_{nj})'$ , ( $j = 1, \dots, p$ ) tienen media cero y varianza igual a 1, es decir, la matriz  $\mathbf{X}$  está estandarizada. La matriz de varianzas y covarianzas, la cual coincide con la matriz de correlaciones de las variables originales, está dada por

$$\mathbf{S} = \frac{1}{n} \mathbf{X}'\mathbf{X} = \begin{bmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_p^2 \end{bmatrix}.$$

Si las  $p$  variables son linealmente independientes, entonces la matriz de varianzas y covarianzas  $\mathbf{S}$  es simétrica y positiva definida.

La varianza total de los datos originales se define por  $\text{tr}(\mathbf{S}) = \sum_j s_j^2$ . El ACP construye combinaciones lineales de las variables originales que expliquen un alto porcentaje de la varianza total en el siguiente sentido.

El primer componente principal, denotado como  $z_1$ , es una combinación lineal de las variables dada por

$$z_1 = \begin{bmatrix} z_{11} \\ z_{21} \\ \vdots \\ z_{n1} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{p1} \end{bmatrix} = \mathbf{X}v_1 = v_{11}\mathbf{X}_1 + v_{12}\mathbf{X}_2 \cdots + v_{1p}\mathbf{X}_p.$$

Los  $v_{1i}$  son los pesos de cada variable en la primera componente. La componente  $z_1$  es una variable con media cero y se busca que tenga máxima varianza, es decir,

$$\text{Var}(z_1) = \frac{(\sum z_{i1}^2)}{n} = \frac{(z_1'z_1)}{n} = \frac{\text{Var}(v_1'\mathbf{X}'\mathbf{X}v_1)}{n} = v_1'[\mathbf{X}'\mathbf{X}/n]v_1 = v_1'\mathbf{S}v_1$$

debe ser máxima. Para maximizar esta varianza se utilizan multiplicadores de Lagrange con el vector  $v_1$  sujeto a la restricción de ortogonalidad  $v_1'v_1 = 1$ . La varianza de  $z_1$  se determina en base a la función objetivo

$$g(v_1) = v_1'Sv_1 - \lambda(v_1'v_1 - 1),$$

donde  $v_1$  es el vector desconocido que proporciona la combinación lineal óptima. Para determinar el máximo se resuelve la siguiente ecuación

$$\frac{\partial g(v_1)}{\partial v_1} = \frac{\partial(v_1'Sv_1 - \lambda(v_1'v_1 - 1))}{\partial v_1} = 2Sv_1 - 2\lambda v_1 = 0,$$

que resulta en el sistema de ecuaciones lineales  $Sv_1 = \lambda v_1$ . Como  $S$  es una matriz simétrica positiva definida, tiene eigenvalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  reales con sus respectivos eigenvectores  $v_1, v_2, \dots, v_p$ , tales que satisfacen  $v_j'v_j = 1$  y  $v_i'v_j = 0$  para  $i \neq j$ . Si se multiplica a  $Sv_1 = \lambda v_1$  por la izquierda por  $v_1$ , se obtiene  $v_1'Sv_1 = \lambda$ . Y por lo tanto  $\lambda$  es el primer eigenvalor de  $S$ , es decir es  $\lambda_1$ . De modo que la varianza de  $z_1$  es  $\lambda_1$  y está es la máxima varianza que puede alcanzar una combinación lineal de las variables originales y por lo tanto la máxima proporción de varianza explicada que se puede lograr es la que explica  $z_1$  dada por  $\lambda_1/\text{tr}(S)$ . El segundo componentes principal  $z_2$  es la combinación lineal de máxima varianza

$$z_2 = v_{21}X_1 + v_{22}X_2 \dots + v_{2p}X_p = Xv_2$$

que no está correlacionada con el primer componente principal  $z_1$ , es decir  $\text{Cov}(z_2, z_1) = 0$ . Ahora, se tiene que maximizar la varianza de  $z_2$ . Es decir, al igual que para  $z_1$ , se tiene que maximizar la función  $v_2'Sv_2$  sujeta a las restricciones  $v_2'v_2 = 1$  y  $v_2'v_1 = 0$ . La nueva función objetivo es

$$g(v_1, v_2) = v_1' \mathbf{S} v_1 + v_2' \mathbf{S} v_2 - \lambda_1 (v_1' v_1 - 1) - \lambda_2 (v_2' v_2 - 1). \quad (1)$$

Derivando e igualando a cero, se tiene

$$\frac{\partial g(v_1, v_2)}{\partial v_1} = 2\mathbf{S}v_1 - 2\lambda_1 v_1 = 0,$$

$$\frac{\partial g(v_1, v_2)}{\partial v_2} = 2\mathbf{S}v_2 - 2\lambda_2 v_2 = 0,$$

y se obtienen los sistemas de ecuaciones lineales  $\mathbf{S}v_1 = \lambda_1 v_1$  y  $\mathbf{S}v_2 = \lambda_2 v_2$ , donde  $v_1$  y  $v_2$  son eigenvectores de  $\mathbf{S}$ . De modo que la varianza de  $z_2$  es  $\lambda_2$  y el máximo de la función objetivo es  $g(v_1, v_2) = \lambda_1 + \lambda_2$ . Por lo tanto la proporción de varianza explicada por los dos primeros componentes principales es  $(\lambda_1 + \lambda_2)/\text{tr}(\mathbf{S})$ .

En general, si se sigue el mismo procedimiento, se tiene que el ACP extrae  $p$  componentes (variables)  $z_1, \dots, z_q, \dots, z_p$  no correlacionados y cuyas varianzas son igual a la varianza total, es decir  $\sum_j \lambda_j = \text{tr}(\mathbf{S})$ .

De estos componentes se espera que los primeros  $q$  (usualmente  $q = 1, 2, 3$ ) expliquen una alta proporción de la varianza total, es decir se espera que la proporción de varianza explicada  $(\lambda_1 + \lambda_2 + \dots + \lambda_q)/\text{tr}(\mathbf{S})$  sea alta.

### 1.1.1 Interpretación Algebraica del ACP

Desde un punto de vista algebraico, los componentes principales son un conjunto de variables que se obtienen de una transformación lineal de las variables originales. Esto es, los componentes principales son un conjunto de  $p$  nuevas variables que son combinaciones lineales de las variables originales. Como ya se vio en la derivación analítica de arriba, por construcción, estas nuevas variables tienen las propiedades de ser linealmente independientes, estar no correlacionadas y tener varianza máxima.

La interpretación estadística del ACP tradicionalmente se ha hecho algebraicamente en el sentido de que los componentes principales se intentan interpretar en términos de los valores y signos de los pesos  $v_{ij}$ . Es decir, en la interpretación se hace énfasis en los conceptos de combinación y transformación lineal.

Si la proporción de varianza explicada por, digamos, los dos primeros componentes principales es alta, se dice que el ACP reduce la dimensión del problema en el sentido de que se va de la dimensión original de los datos a la dimensión 2; perdiéndose en este proceso una pequeña parte de la variabilidad de los datos en su espacio original. No se necesita de  $p$  variables para representar la variabilidad total de los datos, basta con 2 variables para *encapsular* una parte importante de la variabilidad total de los datos.

### 1.1.2 Interpretación Geométrica del ACP

Para entender al ACP desde un punto de vista geométrico, considérese a la nube de puntos que forman los  $n$  vectores de observaciones  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  en  $\mathbb{R}^p$ . Considérese, sin pérdida de generalidad que esta nube de puntos se traslada al origen de  $\mathbb{R}^p$  centrando las columnas de la matriz de datos  $\mathbf{X}$  con su vector de medias, Figura 3.

$\mathbb{R}^p$

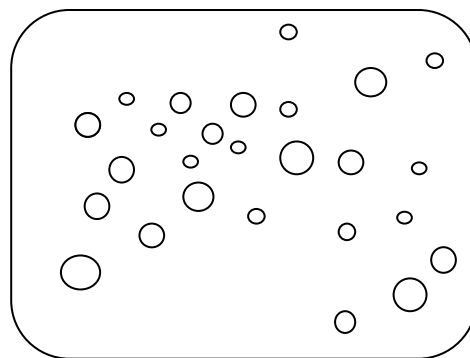


Figura 3. Nube de puntos en  $\mathbb{R}^p$ .

Supóngase que se desea determinar un subespacio vectorial propio de  $\mathbb{R}^p$  tal que cuando los datos se proyecten sobre éste se preserve lo máximo posible el patrón de variabilidad de la nube de puntos en  $\mathbb{R}^p$ . Supóngase que la dimensión se restringe a 1, es decir, se busca determinar la línea recta que pasa por el origen y que da la dirección de máxima variabilidad cuando uno se mueve a través de la nube de puntos en  $\mathbb{R}^p$ . Si el criterio de patrón de máxima variabilidad se da en el sentido de máxima varianza, entonces se busca maximizar la varianza del vector de proyecciones

$$z_1 = \begin{bmatrix} z_{11} \\ z_{21} \\ \vdots \\ z_{n1} \end{bmatrix} = \begin{bmatrix} \langle \mathbf{e}, \mathbf{x}_1 \rangle \mathbf{e} \\ \langle \mathbf{e}, \mathbf{x}_2 \rangle \mathbf{e} \\ \vdots \\ \langle \mathbf{e}, \mathbf{x}_n \rangle \mathbf{e} \end{bmatrix},$$

donde  $\mathbf{e}$  es un vector unitario base de este subespacio, es decir, se busca maximizar

$$\text{Var}(z_1) = \frac{1}{n} \sum_{i=1}^n z_{i1}^2 = \frac{1}{n} \sum_{i=1}^n \langle \mathbf{e}, \mathbf{x}_i \rangle^2 \mathbf{e}.$$

De la expresión anterior, se ve que el problema se reduce a determinar al vector  $\mathbf{e}$  que da la dirección de esta proyección. Del álgebra lineal elemental, se sabe que

$$\|\mathbf{x}_i\|^2 = \|\mathbf{x}_i - \langle \mathbf{e}, \mathbf{x}_i \rangle \mathbf{e}\|^2 + \|\langle \mathbf{e}, \mathbf{x}_i \rangle \mathbf{e}\|^2$$

y además el primer elemento del lado derecho de la ecuación se minimiza cuando  $\mathbf{x}_i$  se proyecta ortogonalmente a lo largo de la dirección del vector  $\mathbf{e}$  y por lo tanto la varianza de  $z_1$  es máxima cuando se hace esta proyección. Una vez más, de la derivación analítica de la Sección 1.1, se ve que esta dirección la da el primer eigenvector  $v_1$  de la matriz  $\mathbf{S}$ . Es decir, los elementos del primer componente principal se obtienen proyectando ortogonalmente a las observaciones  $\mathbf{x}_i$  a lo largo de  $v_1$ .

Después de haber determinado la primera dirección de máxima variabilidad, se busca determinar la segunda dirección de máxima variabilidad que sea ortogonal a  $v_1$ . De las  $p - 1$  direcciones ortogonales a  $v_1$  dadas por  $v_2, \dots, v_q$ , esta dirección está dada por  $v_2$  y las proyecciones ortogonales de los datos  $\mathbf{x}_i$  a lo largo de  $v_2$  proporcionan al segundo componente principal  $z_2$ . Si se sigue esto sucesivamente, se ve que los componentes principales apuntan en direcciones ortogonales de máxima variabilidad. Si se grafican las proyecciones en las dos primeras direcciones de máxima variabilidad se obtiene una gráfica del plano (subespacio de dimensión dos). Si la proporción de varianza explicada por los dos primeros componentes es alta, el diagrama de dispersión del primer componente contra el segundo mostrará un patrón de variabilidad *similar* al que tienen los datos en el espacio original. En el sentido anterior es que el ACP se puede entender como una técnica de búsqueda de patrones lineales y ortogonales de máxima variabilidad en el espacio Euclidiano  $\mathbb{R}^p$ .

Supóngase que los datos están en  $\mathbb{R}^2$  de modo que cada observación  $i$  se puede graficar en un diagrama de dispersión como vector  $(V_{i1}, V_{i2})$ , véase la Figura 4. En los cuatro casos que se ejemplifican ¿cuáles son las direcciones de máxima variabilidad? Para los datos en los dos diagramas de dispersión de la parte superior, visualmente se aprecia que esta dirección pasa por la parte central de los datos moviéndose en una dirección norte-este. Para los datos del diagrama de dispersión de la izquierda en la parte de abajo se aprecia algo similar. Sin embargo, para los datos en el diagrama de dispersión derecho de abajo no es claro que haya direcciones de máxima de variabilidad, o más bien que hay una infinidad de direcciones de máxima variabilidad, o mejor aún, que la variabilidad es igual en cualquier dirección. Al realizar un ACP en los datos donde se identifica dirección de máxima de variabilidad, lo que se está haciendo es cambiar la representación coordenada de los datos en la base canónica estándar

$$\begin{bmatrix} V_{i1} \\ V_{i2} \end{bmatrix} = V_{i1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + V_{i2} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

por la representación coordinada dada por la base que determinan los eigenvectores de la matriz  $\mathbf{S}$

$$\begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix} = V_{i1} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} + V_{i2} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}.$$

Los vectores  $z_1 = (z_{11}, z_{21}, \dots, z_{n1})$  y  $z_2 = (z_{12}, z_{22}, \dots, z_{n2})$  son los componentes principales. Así, ACP consiste básicamente en una rotación ortogonal de los ejes de la base canónica estándar hacia los ejes de la base ortonormal dada por los eigenvectores de la matriz  $\mathbf{S}$ . Geométricamente la configuración de los puntos en  $\mathbb{R}^2$  no se altera con el ACP.

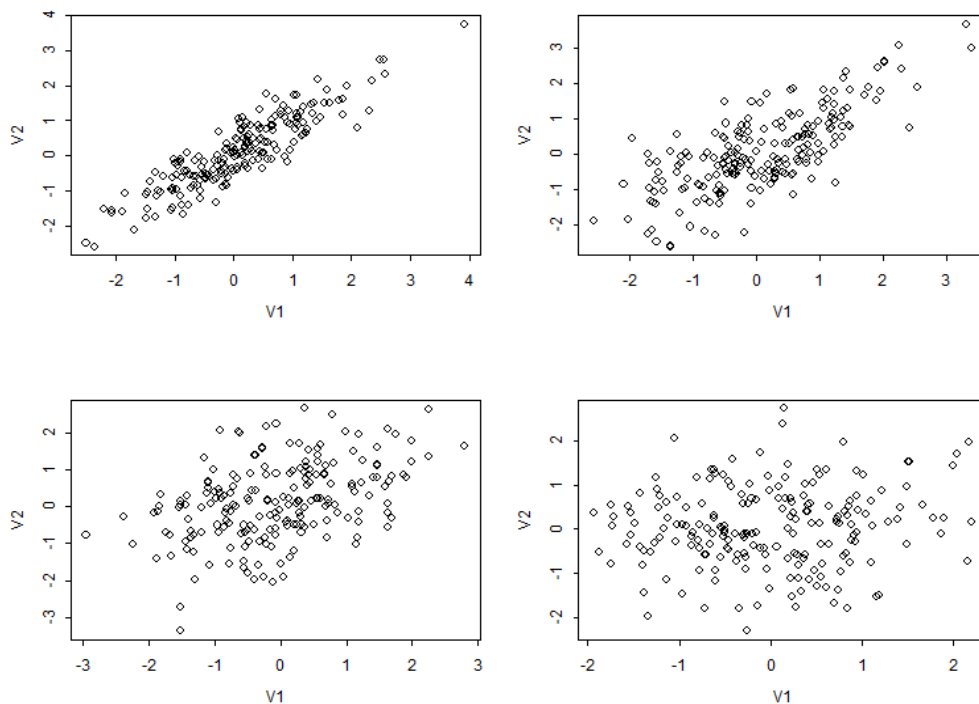


Figura 4. Diagramas de dispersión con configuraciones elipsoidales.

Para el diagrama de dispersión de la derecha en la parte inferior, se tiene que las variables están casi incorrelacionadas por lo que

$$\mathbf{S} \approx \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{bmatrix}.$$

De tal forma que con cualquier rotación ortogonal de los ejes de la base canónica estándar se obtiene otro par de direcciones ortogonales de variabilidad similar. Es decir, como se mencionó anteriormente, cualquier dirección tiene prácticamente la misma variabilidad.

### 1.1.3 Distribuciones Elípticas y ACP

Sea  $\mathbf{X}$  un vector aleatorio  $p \times 1$  cuya función característica  $\psi_{\mathbf{X}}$  satisface  $\psi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}'\mathbf{t})$  para alguna función real  $\phi$ . La función  $\phi$  se llama el generador característico de  $\mathbf{X}$  y se dice que  $\mathbf{X}$  tiene una distribución esférica, lo que se denota por  $\mathbf{X} \sim S_p(\phi)$ . El caso más conocido de distribución esférica es el de la distribución normal multivariada con vector de medias  $\mathbf{0}$  y matriz de varianzas y covarianzas  $\mathbf{I}_p$ . Ahora, sea  $\boldsymbol{\mu}$  un vector  $p \times 1$  y  $\boldsymbol{\Sigma}$  y una matriz  $p \times p$  positiva semi-definida de rango  $d$ . Considérese al vector aleatorio  $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{X}$ , donde  $\mathbf{A}$  es una matriz  $d \times p$  tal que  $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$  y  $\mathbf{X} \sim S_d(\phi)$ . Se dice que  $\mathbf{Y}$  tiene una distribución elíptica y esto se denota por  $\mathbf{Y} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ . Las distribuciones elípticas son generalizaciones de la distribución normal multivariada con vector de medias  $\boldsymbol{\mu}$  y matriz de varianzas y covarianzas  $\boldsymbol{\Sigma}$ .

Las distribuciones elípticas son importantes en diversas áreas de la estadística, en particular en aquellos problemas en los que las distribuciones de interés se pueden caracterizar sólo con su vector de medias y su matriz de varianzas y covarianzas. Lo anterior se debe a la interesante propiedad de que la dispersión de una distribución elíptica está determinada únicamente por su matriz de varianzas y covarianzas.



Cuando la densidad de la distribución elíptica  $E_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  existe y  $\boldsymbol{\Sigma}$  es de rango completo  $p$  y por lo tanto positiva definida, la densidad tiene la forma  $f(\mathbf{y}) = cg((\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}))$ , donde  $c$  es una constante normalizadora y  $g: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  con  $\int_0^\infty g(z) dz < \infty$ . Por ejemplo para la distribución normal multivariada  $g(z) = e^{-z/2}$ .

Una propiedad de las distribuciones elípticas en el caso de dimensión 2 es que las curvas de nivel de la función de densidad son elipses cuyos ejes principales están dados por los eigenvectores de la matriz  $\boldsymbol{\Sigma}$ . Véase la Figura 5.

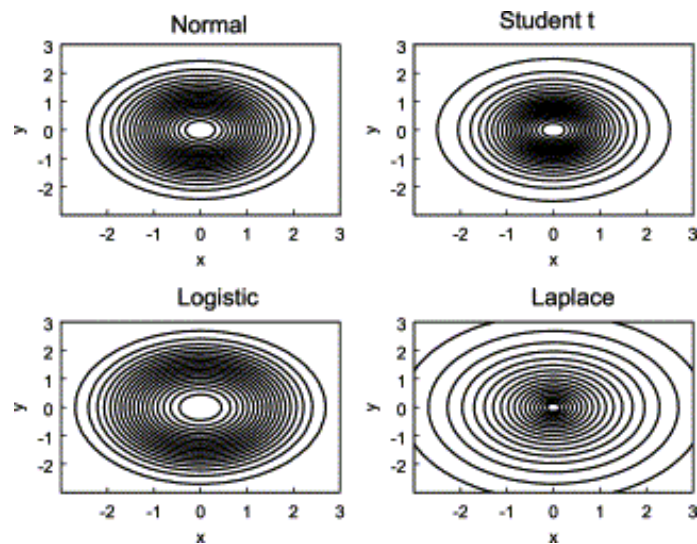


Figura 5. Contornos de diferentes distribuciones elípticas. Imagen tomada de <http://www.sciencedirect.com/science/article/pii/S0167668703001690> de Valdez y Chernih (2003).

Además, como se ve en la estructura de  $f$ , las densidades elípticas son simétricas con respecto a su vector de medias. Esta propiedad se conoce como simetría elíptica.

De manera que si las observaciones son generadas por alguna distribución elíptica  $E_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$  se esperaría obtener una nube de puntos con un patrón de elipse tal y como se ve en los diagramas de dispersión de la Figura 4. En dimensión 3, las superficies de nivel son hiperelipses y, en dimensión  $p$  son hiperelipsoides cuyas hipersuperficies de nivel con nivel  $C$  están determinadas por la forma cuadrática  $(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) = C$ .

Si las  $p$  variables en  $\mathbf{X}$  están correlacionadas, el hiperelipsoide de la nube de puntos no está orientado paralelamente a los ejes de la base canónica estándar. El ACP rota ortogonalmente estos últimos ejes para que coincidan con los ejes del hiperelipsoide  $(\mathbf{y} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\mathbf{y} - \boldsymbol{\mu}) = C$ . Encontrar los ejes principales de este hiperelipsoide se reduce a determinar el eigensistema de la matriz de varianzas y covarianzas  $\mathbf{S}$ , es decir, por la diagonalización  $\mathbf{V}' \mathbf{S} \mathbf{V} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ .

## 1.2 Patrones de Máxima Variabilidad en Espacios de Hilbert: ACPK

Como se vio en la sección anterior, las configuraciones geométricas de la nube de puntos en  $\mathbb{R}^p$  en forma de hiperelipsoides son las que idealmente se identifican con el ACP. El ACP con kernels (ACPK) tiene como objetivo extraer los patrones de máxima variabilidad no lineales de un conjunto de datos transformándolos mediante una función no lineal  $\Phi$  y haciendo el ACP, mediante el truco kernel, en el espacio de los datos transformados (Schölkopf, Mika, y Burges, 1999). Los detalles del ACPK son los siguientes.

Supóngase que se tiene una muestra de entrenamiento de  $n$  observaciones  $\mathbf{x}_i$ , centradas, esto es,  $\sum_{i=1}^n \mathbf{x}_i = 0$ , donde  $\mathbf{x}_i \in \mathcal{V} \subseteq \mathbb{R}^p$ ,  $i = 1, 2, \dots, n$ . Sea  $\Phi: \mathbb{R}^p \rightarrow \mathcal{H}$  un mapeo no lineal, donde  $\mathcal{H}$  es un espacio de Hilbert de dimensión mayor que  $p$ . El ACP se hace con los  $n$  datos transformados  $\{\Phi(\mathbf{x}_i)\}$ . Supóngase que los datos

transformados están centrados, es decir  $\sum_{i=1}^n \Phi(\mathbf{x}_i) = 0$ . En el ACPK se debe diagonalizar la matriz de varianzas y covarianzas de los datos transformados

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_i)' \quad (2)$$

Si el espacio  $\mathcal{H}$  es de dimensión infinita,  $\Phi(\mathbf{x}_i) \Phi(\mathbf{x}_i)'$  se considera como un operador lineal que mapea a  $\mathbf{u} \in \mathcal{H}$ , con  $\mathbf{u} = \Phi(\mathbf{x})$  a  $\Phi(\mathbf{x}_i)(\Phi(\mathbf{x}_i) \cdot \mathbf{u})$ . Ahora se procede a encontrar los eigenvalores  $\lambda \geq 0$  y los eigenvectores  $\mathbf{V} \in \mathcal{H} \setminus \{0\}$  de  $\mathbf{C}$ . Es decir, se resuelve la ecuación  $\lambda \mathbf{V} = \mathbf{C} \mathbf{V}$ .

Nótese que las soluciones de  $\mathbf{V}$  caen en el espacio generado por  $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)$ , lo cual tiene dos consecuencias útiles. Primero se puede considerar la ecuación equivalente

$$\lambda(\Phi(\mathbf{x}_i) \cdot \mathbf{V}) = (\Phi(\mathbf{x}_i) \cdot \mathbf{C} \mathbf{V}) \quad \text{para todo } i = 1, \dots, n, \quad (3)$$

y segundo, como los eigenvectores se pueden expresar como una combinación lineal de los  $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)$ , tenemos que existen coeficientes  $\alpha = (\alpha_1, \dots, \alpha_n)'$  tales que

$$\mathbf{V} = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i) \quad (4)$$

Sea  $\mathbf{K}$  una matriz  $n \times n$  cuyos elementos están dados por  $\mathbf{K}_{ij} := (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$ , la matriz  $\mathbf{K}$  se llama matriz Gram. Sustituyendo en (3) a (2) y (4) se tiene  $\lambda \mathbf{K} \alpha = \mathbf{K}^2 \alpha$ . Como  $\mathbf{K}$  es una matriz simétrica, sus eigenvectores generan todo el espacio y por lo tanto proporcionan todas las soluciones  $\alpha$  de la ecuación  $n \lambda \alpha = \mathbf{K} \alpha$ . Nótese que  $\mathbf{K}$  es una matriz semidefinida positiva, que se puede escribir como

$(\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n))' \cdot (\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n))$ , lo que que implica que para todo  $\mathbf{X} \in \mathcal{H}$ ,  $\mathbf{X} \cdot \mathbf{KX} = \|(\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n))\mathbf{X}\|^2 \geq 0$ . Por lo tanto los eigenvalores de  $\mathbf{K}$  serán no negativos y darán las soluciones de  $n\lambda$  en la ecuación  $n\lambda\alpha = \mathbf{K}\alpha$ . En este caso, solo se necesita diagonalizar  $\mathbf{K}$ . Sean  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  los eigenvalores de  $\mathbf{K}$  y  $\alpha^1, \alpha^2, \dots, \alpha^n$  los respectivos eigenvectores normalizados en el espacio  $\mathcal{H}$  ya que se requiere que los correspondientes vectores en  $\mathcal{H}$  estén normalizados, es decir  $(\mathbf{V}^q \cdot \mathbf{V}^q) = 1$  para todo  $q = 1, \dots, n$ . Usando las ecuaciones (4) y  $n\lambda\alpha = \mathbf{K}\alpha$ , se obtiene

$$\begin{aligned}
 1 &= \sum_{i,j=1}^n \alpha_i^q \alpha_j^q (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \\
 &= \sum_{i,j=1}^n \alpha_i^q \alpha_j^q \mathbf{K}_{ij} \\
 &= (\alpha^q \cdot \mathbf{K}\alpha^q) \\
 &= \lambda_q (\alpha^q \cdot \alpha^q).
 \end{aligned}$$

Después de extraer los componentes principales en ACPK, se procede a calcular las proyecciones sobre los eigenvectores  $\mathbf{V}^q$  en  $\mathcal{H}$ . Si tenemos un punto de prueba  $\mathbf{x}$ , con imagen  $\Phi(\mathbf{x})$  en  $\mathcal{H}$ , entonces

$$(\mathbf{V}^q \cdot \Phi(\mathbf{x})) = \sum_{i=1}^n \alpha_i^q (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) = \sum_{i=1}^n \alpha_i^q k(\mathbf{x}_i, \mathbf{x})$$

es la proyección ortogonal de  $\Phi(\mathbf{x})$  sobre  $\mathbf{V}^q$ . En resumen, para realizar el ACPK se siguen los siguientes pasos:

1. Seleccionar la función kernel que representa el producto punto de los datos transformados  $(\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}))$ .
2. Calcular la matriz  $\mathbf{K}$ .

3. Diagonalizar  $\mathbf{K}$  (encontrar  $\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ , donde  $\mathbf{U}$  es la matriz formada por los eigenvectores de  $\mathbf{K}$  y  $\mathbf{\Lambda}$  es la matriz con los correspondientes eigenvalores en la diagonal principal).
4. Normalizar los vectores  $\mathbf{V}$ .
5. Calcular las proyecciones de un punto de prueba en los eigenvectores, para ello, se deben extraer  $t$  componentes principales.

### 1.2.1 Fases del ACPK

Para llevar a cabo la aplicación del método kernel, a continuación se describen los pasos a seguir:

- Fase 1. Contar con los datos de la muestra de entrenamiento, de la cual interesa descubrir algún patrón.
- Fase 2. Después de estar familiarizados con la estructura de los datos de entrada, elegir la función kernel y calcular la matriz Gram para determinar el patrón de los datos.
- Fase 3. En el espacio de las características, aplicar un algoritmo lineal para análisis de patrones.

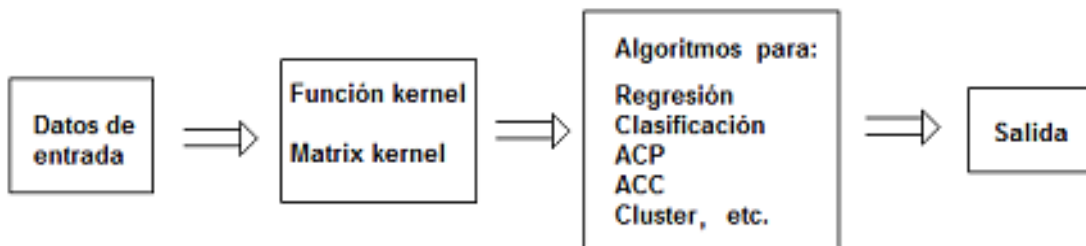


Figura 6. Arquitectura modular del método kernel. Shawe-Taylor y Cristianini (2004).

En la fase 2 se hace el cálculo de los productos punto. Esto es, los datos están en el espacio de las características (de alta dimensión) donde puede ser analizada su estructura.

El método kernel se puede aplicar en todos los algoritmos de análisis de patrones cuyas entradas se pueden expresar en términos de productos punto. Si los datos en el espacio original no se pueden analizar satisfactoriamente con la técnica de análisis multivariado, se podrá optar por el método kernel, con el truco kernel, para intentar estudiar el comportamiento de los patrones de los datos. Para este caso, la Tabla 1 presenta el algoritmo del ACPK.

Tabla 1. Algoritmo de ACPK

---

- 1) Dado los datos  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{V} \subseteq \mathbb{R}^p$  y dado un kernel  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$
  - 2) Se construye la matriz Gram  $\mathbf{K}$  con elementos  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$
  - 3) Se resuelve el problema del eigenvalor  $K\alpha^j = \frac{\lambda_j}{n}\alpha^j$
  - 4) ACPK extrae  $n$  características  $v_j = \frac{1}{\sqrt{\lambda_j}} \sum_{i=1}^n \alpha_i^j \Phi(\mathbf{x}_i)$   $j = 1, \dots, n$
- 

### 1.3 Construcción de $\mathcal{H}$ a partir de $\Phi$

En esta sección se muestra el proceso para construir un espacio de características  $\mathcal{H}$  asociado al mapeo no lineal  $\Phi$ . Este proceso considera los siguientes pasos:

- Convertir la imagen de  $\Phi$  en un espacio vectorial.
- Definir un producto interno  $\langle \cdot, \cdot \rangle$  en  $\mathcal{H}$  que satisfaga  $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ .

Se construye un espacio de producto interno que contenga las imágenes  $\Phi(\mathbf{x}_i)$ . Para esto se define un espacio vectorial tomando combinaciones lineales de la forma

$$f(\cdot) = \sum_{i=1}^m \alpha_i k(\cdot, x_i) \quad \text{y} \quad g(\cdot) = \sum_{j=1}^{m'} \beta_j k(\cdot, x'_j)$$

donde  $m, m' \in \mathbb{N}$ ,  $\alpha_i, \beta_j \in \mathbb{R}$  y  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}'_1, \dots, \mathbf{x}'_n \in \mathcal{V}$ . Se define el producto punto entre  $f$  y  $g$  por

$$\langle f, g \rangle := \sum_{i=1}^m \sum_{j=1}^{m'} \alpha_i \beta_j k(\mathbf{x}_i, \mathbf{x}'_j).$$

Para ver que la expresión de arriba está bien definida se usa  $k(\mathbf{x}'_j, \mathbf{x}_i) = k(\mathbf{x}_i, \mathbf{x}'_j)$  para así obtener

$$\langle f, g \rangle := \sum_{j=1}^{m'} \beta_j f(x'_j) = \sum_{i=1}^m \alpha_i g(x_i)$$

Esta última ecuación muestra que el producto interno  $\langle \cdot, \cdot \rangle$  es bilinear. También muestra que es simétrico ya que  $\langle f, g \rangle = \langle g, f \rangle$ . Además es positivo definido, ya que como el kernel  $k$  es positivo definido, para cualquier función  $f$  dada por  $f(\cdot) = \sum_{i=1}^m \alpha_i k(\cdot, x_i)$  se tiene que

$$\langle f, f \rangle = \sum_{i=1}^m \alpha_i \alpha_j k(x_i, x_j) \geq 0.$$

Por lo tanto  $\langle \cdot, \cdot \rangle$  es un kernel positivo definido sobre el espacio de funciones. Para ver esto, nótese que dadas las funciones  $f_1, \dots, f_n$  y coeficientes  $\gamma_1, \dots, \gamma_n \in \mathbb{R}$ , se tiene

$$\sum_{i,j=1}^n \gamma_i \gamma_j \langle f_i, f_j \rangle = \langle \sum_{i=1}^n \gamma_i f_i, \sum_{j=1}^n \gamma_j f_j \rangle \geq 0.$$

La igualdad de la izquierda resulta de la bilinealidad del  $\langle \cdot, \cdot \rangle$ , mientras que la desigualdad a la derecha resulta de

$$\langle f, f \rangle = \sum_{i,j=1}^m \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

El último paso es probar que reúne las condiciones necesarias de un producto interno. De la definición

$$f(\cdot) = \sum_{i=1}^m \alpha_i k(\cdot, \mathbf{x}_i)$$

se tiene que  $\langle k(\cdot, \mathbf{x}), f \rangle = f(\mathbf{x})$ . En particular  $\langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x}') \rangle = k(\mathbf{x}, \mathbf{x}')$ . Ahora bien, se puede mostrar que cualquier kernel positivo definido satisface la desigualdad de Cauchy-Schwarz

$$|k(\mathbf{x}_1, \mathbf{x}_2)|^2 \leq k(\mathbf{x}_1, \mathbf{x}_1) k(\mathbf{x}_2, \mathbf{x}_2)$$

Por lo tanto

$$|f(\mathbf{x})|^2 = |\langle k(\cdot, \mathbf{x}), f \rangle|^2 \leq k(\mathbf{x}, \mathbf{x}) \langle f, f \rangle.$$

Por lo tanto,  $\langle f, f \rangle = 0$  implica que  $f = 0$ , con lo que se completa la demostración de que  $\langle \cdot, \cdot \rangle$  es un producto interno.

La construcción que se acaba de desarrollar muestra que cualquier kernel positivo definido se puede considerar como una función que evalúa los productos internos en



otro espacio, es decir  $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ . Y, por lo tanto, el espacio de producto interno  $\mathcal{H}$  que se acaba de construir es una instancia para el espacio de las características asociado con el kernel  $k$ .

Por lo tanto, el espacio de producto punto  $\mathcal{H}$  construido de esta manera es una instancia posible del espacio de características asociado con un kernel.

Supóngase que se parte de forma inversa. Es decir a partir de un mapeo  $\Phi$  del espacio original  $\mathcal{V}$  a un espacio de producto interno. En este caso se obtiene un kernel positivo definido positivo vía  $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$  ya que para todos  $c_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathcal{V}$ , ( $i = 1, \dots, m$ ), se tiene

$$\sum_{i,j} c_i c_j k(x_i, x_j) = \left\langle \sum_i c_i \Phi(x_i), \sum_j c_j \Phi(x_j) \right\rangle = \left\| \sum_i c_i \Phi(x_i) \right\|^2 \geq 0.$$

Esto tiene dos consecuencias. La primera es que permite definir a un kernel positivo definido como función  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  con la propiedad de que existe una función  $\Phi$  tal que  $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ . La segunda consecuencia es que permite la construcción de kernels. La identidad  $k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$  es la base para el truco kernel.

**Truco Kernel.** Dado un algoritmo que está basado en productos puntos de datos vectoriales  $\mathbf{x}_1, \dots, \mathbf{x}_n$  de tal manera que este algoritmo se puede formular en términos de un kernel  $k$  positivo definido, se puede construir un algoritmo alternativo reemplazando al kernel  $k$  por otro kernel  $\tilde{k}$  positivo definido que este opera en  $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)$ .

### 1.3.1 Funciones Kernel

Como vimos en la sección anterior, la evaluación de la función  $\Phi$  implícitamente implica un alto costo computacional, sin embargo la función kernel  $k$  permite evaluar el producto interno entre vectores del espacio de características sin necesidad de conocer al mapeo  $\Phi: \mathcal{V} \rightarrow \mathcal{H}$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$$

para todo  $x_i$  y  $x_j$  en el espacio de entrada. Los kernels que han demostrado ser útiles en los métodos de estadística multivariada son:

Gaussiano  $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$

Polinomial  $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + c)^d$ ,  $c \geq 0$

Sigmoidal  $k(\mathbf{x}, \mathbf{y}) = \sigma(\tau \cdot (\mathbf{x} \cdot \mathbf{y}) + \theta)$ ,  $\tau, \theta \in \mathbb{R}$

Otros kernels que también se han usado ampliamente se muestran en la siguiente tabla.

Tabla 2. Otras funciones kernels.

---

Kernel lineal	$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{y} + c$
Kernel exponencial	$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\ \mathbf{x} - \mathbf{y}\ }{2\sigma^2}\right)$
Kernel Laplaciano	$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\ \mathbf{x} - \mathbf{y}\ }{\sigma}\right)$
Kernel ANOVA	$k(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^n \exp(-\sigma(\mathbf{x}^j - \mathbf{y}^j)^2)^d$
Kernel cuadrático racional	$k(\mathbf{x}, \mathbf{y}) = 1 - \frac{\ \mathbf{x} - \mathbf{y}\ ^2}{\ \mathbf{x} - \mathbf{y}\ ^2 + c}$
Kernel multicuadrático	$k(\mathbf{x}, \mathbf{y}) = \sqrt{\ \mathbf{x} - \mathbf{y}\ ^2 + c^2}$
Kernel inverso multicuadrático	$k(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{\ \mathbf{x} - \mathbf{y}\ ^2 + c^2}}$
Kernel circular	$k(\mathbf{x}, \mathbf{y}) = \ \mathbf{x} - \mathbf{y}\  < \sigma$
Kernel esférico	$k(\mathbf{x}, \mathbf{y}) = 1 - \frac{3\ \mathbf{x} - \mathbf{y}\ }{2\sigma} + \frac{1}{2}\left(\frac{\ \mathbf{x} - \mathbf{y}\ }{\sigma}\right)^3$
Kernel Ola	$k(\mathbf{x}, \mathbf{y}) = \frac{\theta}{\ \mathbf{x} - \mathbf{y}\ } \sin \frac{\ \mathbf{x} - \mathbf{y}\ }{\theta}$
Kernel potencia	$k(\mathbf{x}, \mathbf{y}) = -\ \mathbf{x} - \mathbf{y}\ ^d$
Kernel log	$k(\mathbf{x}, \mathbf{y}) = -\log(\ \mathbf{x} - \mathbf{y}\ ^d + 1)$
Kernel Cauchy	$k(\mathbf{x}, \mathbf{y}) = \frac{1}{1 + \frac{\ \mathbf{x} - \mathbf{y}\ ^2}{\sigma}}$
Kernel T-Student generalizado	$k(\mathbf{x}, \mathbf{y}) = \frac{1}{1 + \ \mathbf{x} - \mathbf{y}\ ^d}$

---

## **Capítulo 2. Mejora del Kernel en ACPK**

En este capítulo se presenta la propuesta de esta tesis, la cual consiste en mejorar el kernel en ACPK. El criterio de mejora se define formalmente más adelante. La propuesta es análoga a la de Kim et al. (2006) en el sentido de que se hace la búsqueda de un kernel óptimo dentro de un espacio convexo de kernels. Sin embargo el procedimiento de optimización difiere del de estos autores quienes usan técnicas estándares de programación convexa mientras que en esta tesis la búsqueda se realiza con heurísticas evolutivas.

### **2.1 Algoritmos Genéticos**

Existen problemas de optimización en los cuales la función objetivo es tal que las soluciones óptimas no se pueden establecer analíticamente. Esto ha motivado una diversidad de técnicas alternas de optimización dentro de las que se encuentran los Algoritmos Genéticos (AG), propuestos por Holland (1975). Los AG son algoritmos de optimización que realizan una búsqueda aleatoria en el espacio de todas las soluciones factibles. Los AG están inspirados en la evolución de la naturaleza. Los AG utilizan como paradigma de computación la simulación del mecanismo de genética y selección natural. En la actualidad los AG se han aplicado a una gran variedad de problemas de optimización discretos y continuos.

A continuación se da una breve descripción de los conceptos fundamentales de los AG. Una descripción detallada se puede ver en Montano y Cantú (2011).

### **2.1.1 Constitución de un Cromosoma**

Todos los organismos vivos están constituidos por células. En cada célula existe el mismo juego de cromosomas. Los cromosomas son cadenas de ADN y generan un modelo para el organismo entero. Un cromosoma está formado de genes, bloques de ADN, los cuales tienen su propia posición en el cromosoma. Básicamente se puede decir, que cada gene indica en código un rasgo y todos los posibles rasgos se llaman alelos.

El grupo completo de material genético (todos los cromosomas) se llama genoma y el grupo particular de genes en genoma se llama genotipo.

**Codificación del Cromosoma.** La manera más usual y simple de codificación es mediante una cadena binaria. Cada cromosoma tiene una cadena binaria y cada bit (0,1) de la cadena puede representar alguna característica de la solución. Existen muchas otras maneras de codificación, la cual depende de la solución del problema; ya que se pueden escribir en código números enteros o reales, usando una de las tres formas siguientes: codificación de la permutación y codificación del valor.

**Codificación de la Permutación.** Se puede usar codificando cada cromosoma, la cual es una cadena de números enteros y representa un número en una secuencia. Esta codificación es útil para problemas de clasificación.

**Codificación de Valor.** Es útil en problemas donde los valores son números reales y es necesario desarrollar algunos cruces y mutaciones específicas al problema. En este caso, las codificaciones nos llevan a denotar un cromosoma como una función o modelo con números enteros o reales; por lo tanto, después de haber identificado la codificación del cromosoma es importante conocer los operadores genéticos que se pueden usar.

**Operadores Genéticos.** Los operadores genéticos son mecanismos de búsqueda, los cuales se usan para crear nuevas soluciones, basadas en el conjunto de los mejores individuos escogidos previamente por algún método de selección.

Básicamente, son tres operadores que inducen a un buen resultado en un algoritmo genético, estos son: reproducción, cruce y mutación; pero de acuerdo al tipo de codificación del cromosoma, se generan ocho operadores genéticos, de los cuales son: mutación binaria, mutación uniforme, mutación límite, mutación no uniforme, mutación por intercambio, cruce simple, cruce aritmético y cruce heurístico. En este caso solo se describen los operadores correspondientes a valores reales.

#### 1. Reproducción

El cromosoma con alto valor de ajuste tiene mayor oportunidad de participar en la próxima generación.

#### 2. Mutación por intercambio

Se aplica a codificaciones basadas en el orden (permutaciones). Consiste en elegir dos coordenadas del cromosoma (del modelo o función) al azar e intercambiar sus valores.

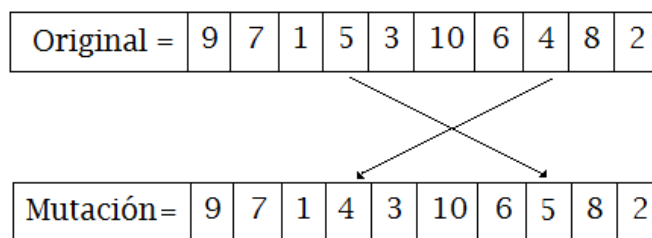


Figura 7. Mutación por intercambio.

### 3. Cruce Aritmético

Toma dos individuos y produce dos nuevos individuos. Este operador se define como una combinación lineal convexa de dos vectores. Si se tienen  $\mathbf{x}_1$  y  $\mathbf{x}_2$  vectores que serán cruzados, los resultados de la descendencia son

$$\mathbf{x}'_1 = a\mathbf{x}_1 + (1 - a)\mathbf{x}_2 \quad \text{y} \quad \mathbf{x}'_2 = a\mathbf{x}_2 + (1 - a)\mathbf{x}_1$$

Éste operador usa un valor aleatorio  $a$  que se encuentra entre 0 y 1.

**Métodos de Selección.** La selección de individuos (cromosomas o modelos) para producir generaciones sucesivas juega un papel sumamente importante en un AG. Una selección probabilística se realiza basada en el valor de mejor ajuste del individuo, de tal manera que los individuos mejores tienen mayor oportunidad de ser seleccionados. El problema es cómo seleccionar a estos individuos; ya que de acuerdo a la teoría de Darwin los mejores deben sobrevivir y crear nuevas descendencias. Para este caso, existen varios esquemas, tales como: el método de la ruleta, técnicas de escala, selección de Boltzman, selección por concurso, selección por clasificación y selección de naturaleza uniforme (Montano y Cantú, 2011).

En este trabajo, se empleo el método de selección de naturaleza uniforme. La idea en esta selección es que una proporción grande de individuos debe sobrevivir para formar la próxima generación. Por lo tanto, en cada generación se seleccionan a los mejores individuos para crear una nueva descendencia, esto es, los cromosomas (modelos) malos tienden a desaparecer, ya que serán sustituidos por la nueva descendencia junto con el resto de la población.

Para aplicar un AG se debe tener bien claro que es lo que se busca. Primero se debe plantear la función objetivo y en base a ésta, se buscan funciones que proporcionen probabilidades mayores o menores que la proporcionada por la función objetivo; por

lo tanto, los modelos que cumplan con la restricción son candidatos a ser funciones óptimas.

Después de plantear la función objetivo, se procede a generar la población inicial; la cual está formada por funciones (progenitoras) con su correspondiente probabilidad de ajuste. El tamaño de la población inicial es el número de funciones progenitoras que forman la generación. Si son pocos individuos la probabilidad de realizar el cruce es poca e implica que solo una parte del espacio de búsqueda es explorado.

Al usar el método de selección, la probabilidad de las funciones es monitoreada, de tal forma que los modelos que expliquen un mejor ajuste tienen mayor probabilidad de ser seleccionadas y producir más de una combinación (más hijos). La Figura 8 ilustra el funcionamiento de un AG.

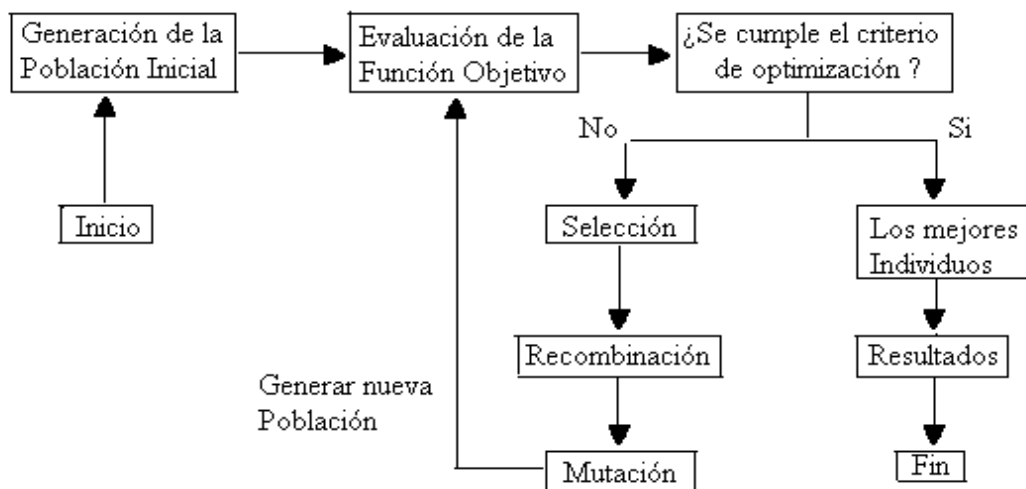


Figura 8. Estructura de la evolución de un Algoritmo Genético.



Montano y Cantú (2011) recomiendan que la población inicial debe contener al menos 20 individuos y a lo más 100. Si se desea procesar con menos de 20 individuos, posiblemente no se encuentre la respuesta óptima, y en caso de que sean más de 100, el algoritmo presentará problemas de ejecución cuando se realicen los cruces aritméticos.

## 2.2 Propuesta de Mejora del Kernel para ACPK

En la propuesta que se hace, se parte de una matriz de datos  $\mathbf{X}$   $n \times p$  la cual se considera sin pérdida de generalidad centrada por su vector de medias. También se tiene un conjunto fijo y arbitrario de  $\delta$  kernels  $k_i$  con los cuales se define el espacio convexo de kernels

$$\mathcal{K} = \left\{ k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}: k = \sum_{i=1}^{\delta} \theta_i k_i, \theta = (\theta_1, \dots, \theta_{\delta}), \sum_{i=1}^{\delta} \theta_i = 1, \theta_i \geq 0 \right\}$$

Como se vio en la Sección 1.3, los elementos de  $\mathcal{K}$  también son kernels. La mejora de kernel se entiende en el sentido de encontrar al kernel  $k^*$  en  $\mathcal{K}$  que maximice la proporción relativa de varianza explicada por el ACPK para una dimensión  $q$  fija. Es decir, se busca maximizar

$$F(k) = \sum_{i=1}^q \lambda_i(k) / \sum_{i=1}^n \lambda_i(k)$$

donde los  $\lambda_i(k)$  son los eigenvalores del ACPK usando el kernel  $k \in \mathcal{K}$ . Ahora bien, sea  $G_i$  la matriz Gram que resulta de evaluar el kernel  $k_i$  en los datos  $\mathbf{X}$  y consideremos al espacio convexo de matrices Gram dado por

$$\mathcal{G} = \left\{ G | G = \sum_{i=1}^{\delta} \theta_i G_i, \theta = (\theta_1, \dots, \theta_{\delta}), \sum_{i=1}^{\delta} \theta_i = 1, \theta_i \geq 0 \right\}$$

Es claro que la selección del kernel óptimo  $k^*$  es equivalente a la selección de la matriz Gram,  $G^*$ , obtenida de evaluar este kernel en los datos. Esto a la vez es equivalente a elegir al vector de ponderaciones  $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_\delta^*)$  que construyen a  $G^*$  a partir de los kernels base  $k_i$ , es decir  $G^* = \theta_1^* G_1 + \theta_2^* G_2 + \dots + \theta_\delta^* G_\delta$  con las restricciones de que  $\sum_{i=1}^{\delta} \theta_i^* = 1$  y  $\theta_i^* \geq 0$ . Por el algoritmo para extraer los componentes principales en ACPK, se tiene que la función objetivo queda dada por

$$F(G) = \frac{\sum_{i=1}^q \lambda_i(G)}{\sum_{i=1}^n \lambda_i(G)}$$

donde ahora los  $\lambda_i(G)$  son los eigenvalores del ACPK extraídos de la matriz Gram  $G$  en  $\mathcal{G}$ .

De esta manera se plantea un problema de optimización convexa con restricciones lineales, más específicamente, un problema de optimización de valores propios en un espacio convexo de matrices simétricas y positivas definidas. Existen muchos algoritmos numéricos de programación convexa semidefinida para resolver este problema de optimización. Dado que este problema de optimización es numéricamente complejo, en este trabajo se propone como alternativa el paradigma de algoritmos evolutivos para resolver la optimización.

A continuación se presentan los detalles con el kernel Gaussiano

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$$

con parámetro  $\sigma^2$ .

Se fijan  $\delta = 10$  kernels Gaussianos y se generan de forma aleatoria  $\delta$  valores de los  $\theta$ 's con las restricciones de que  $\sum_{i=1}^{\delta} \theta_i = 1$  y  $\theta_i \geq 0$ . De esta manera se construyeron las matrices Gram

$$G = \theta_1 \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_1^2}\right) + \dots + \theta_{\delta} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_{\delta}^2}\right).$$

De estas matrices se obtienen sus eigenvalores. El AG requiere de una población cero o población inicial, la cual se obtiene generando los valores de los sigmas y así construir el espacio de matrices de Gram  $\mathcal{G}$ . Dentro del proceso del AG estas matrices permanecerán fijas. El vector de ponderaciones  $\theta$  tomará el papel de coeficientes. Entonces se genera de forma aleatoria al vector  $\theta$  con las restricciones  $\sum_{i=1}^{\delta} \theta_i = 1$  y  $\theta_i \geq 0$ . Estos pasos se repiten  $M$  veces. En los casos de prueba se usó  $M = 30$ . De esta forma la población inicial tiene la siguiente estructura

Tabla 3. Población inicial de  $\theta$ 's .

	$c_1$	$c_2$		$c_{\delta}$
1	$\theta_{11}$	$\theta_{12}$	$\dots$	$\theta_{1\delta}$
2	$\theta_{21}$	$\theta_{22}$	$\dots$	$\theta_{2\delta}$
3	$\theta_{31}$	$\theta_{32}$	$\dots$	$\theta_{3\delta}$
:	:	:		:
:	:	:		:
30	$\theta_{30,1}$	$\theta_{30,2}$	$\dots$	$\theta_{30,\delta}$

El siguiente paso es elegir al renglón que explique mayor proporción de varianza. Cada hilera se evalúa con las matrices Gram

$$G_N = \theta_1 G_1 + \theta_2 G_2 + \dots + \theta_{\delta} G_{\delta}$$

de tal manera que se obtienen 30 nuevas matrices Gram, de las cuales se calculan sus eigenvalores para así conocer la proporción de varianza explicada por cada kernel.

Hasta aquí se sabe cómo explica la varianza la primera generación. Ahora se procede con un método de selección, en este caso se emplea el método de selección de naturaleza uniforme, el cual selecciona los modelos con alto proporción de varianza explicada. La idea es crear nuevos kernels (modelos) con al menos igual proporción de varianza explicada.

En este caso se eligen los 5 modelos con la mayor proporción de varianza explicada y se procede a aplicar el operador cruce aritmético.

Para cada par de modelos seleccionados se genera un valor aleatorio de la distribución uniforme  $r$  y se multiplica por los valores del vector  $\theta$  del primer modelo; el valor  $1 - r$  se multiplica por los valores del vector  $\theta$  del segundo modelo, obteniéndose así un nuevo modelo.

Modelo 1	$r \times$	$\theta_{21}$	$\theta_{22}$	$\dots$	$\theta_{2\delta}$
		+	+		+
Modelo 2	$(1 - r) \times$	$\theta_{31}$	$\theta_{32}$	$\dots$	$\theta_{3\delta}$
Nuevo modelo		$T_1$	$T_2$	$\dots$	$T_\delta$

Después, el Modelo 1 se multiplica por  $1 - r$  y el Modelo 2 se multiplica por  $r$ . Se suman y se obtiene otro nuevo modelo. Este operador toma dos modelos y genera dos nuevos modelos, entonces se realizan los cruces considerando todas las combinaciones entre los 5 modelos seleccionados. Se obtiene así una matriz de 20 nuevos modelos (nueva generación). Cabe señalar, que los modelos son alterados al multiplicarlos por los valores aleatorios  $r$  y  $1 - r$ , sin embargo, se mantiene la matriz Gram que se obtiene continua estando en el espacio factible  $\mathcal{G}$ .

Estos nuevos modelos son evaluados con las matrices de Gram, porque ahora el interés es saber cuáles de estos 20 modelos proporcionan la mayor proporción de varianza explicada.

Posteriormente se procede a la aplicación del operador mutación. Existen varios operadores por mutación, pero se corre el riesgo de que al aplicarlo se alteren los valores del vector  $\theta$ , de tal forma que la matriz Gram quede fuera de  $\mathcal{G}$ . Por lo que se opta por aplicar la mutación por intercambio. Para la mutación por intercambio de un modelo se seleccionan aleatoriamente 5 valores del vector  $\theta$  y se intercambian de posición. Esto se hace en cada uno de los 20 modelos, lo que lleva a evaluar los nuevos modelos resultantes de la mutación para conocer sus proporciones de varianzas explicadas.

Al aplicar este operador se nota que la proporción de varianzas explicadas tendían a disminuir mientras que el operador que reporta un incremento en la proporción de varianzas explicadas es el cruce aritmético. Por lo tanto se descarta el operador mutación por intercambio.

Tabla 4. Algoritmo para la mejora del kernel a través de algoritmos genéticos.

---

```

generar  $\delta\sigma$ 's entre  $10^{-1}$  y  $10^2$  .
  evaluar los  $\delta$  kernels a la base de datos
  repetir
  generar  $\delta\theta$ 's , con restricción  $\sum_{i=1}^{\delta}\theta_i = 1$   $\theta_i \geq 0$ 
  hasta n
  fin de la repetición
  repetir
    para i hasta  $\delta$ 
       $G_N = \text{suma}(\theta_i * G_i)$ 
    fin del para
  hasta n
  fin de la repetición
  repetir
    obtener los valores característicos de  $G_N$ 
     $varex = (v_1 + v_2) / \sum_{i=1}^n v_i$ 
  hasta n
  fin de la repetición
repetir
  ordenar modelos de  $\theta$ 's por  $varex$ 
     $c = \frac{5!}{(5-2)!}$ 
  selección de 5 primeros modelos
  repetir
  generar  $r$  (0,1)
  para i=1 hasta 4
  para j=(i+1) hasta 5
     $nm_z = r * \text{modelo}_i + (1-r) * \text{modelo}_j$ 
     $nm_{z+1} = r * \text{modelo}_j + (1-r) * \text{modelo}_i$ 
  hasta c nuevos modelos
  fin de la repetición
  repetir
  para i hasta  $\delta$ 
     $G_N = \text{suma}(\theta_i * G_i)$ 
  fin del para
  hasta c
  fin de la repetición
  repetir
    obtener los valores característicos de  $G_N$ 
     $varex = (v_1 + v_2) / \sum_{i=1}^n v_i$ 
  hasta c
  fin de la repetición
hasta 5 iteraciones
fin de la repetición

```

---

### Capítulo 3. Ejemplo y Aplicación

Para ejemplificar la técnica propuesta de mejora de kernel en ACPK, se hace uso del conjunto de datos de Indicadores Socioeconómicos de los 210 Municipios del Estado de Veracruz. Estos indicadores se construyeron con la información del II Censo de Población y Vivienda 2005 y la Encuesta Nacional de Ocupación y Empleo (ENOE) 2005. Con estos datos se calcularon ocho de los nueve indicadores relacionados con educación, vivienda y distribución de la población. Los datos de la ENOE se usaron para estimar el décimo indicador concerniente a los ingresos monetarios. La fuente de estos indicadores es INEGI y se utilizan para calcular el índice de marginación en el año 2005 mediante un ACP de la matriz de datos de orden  $210 \times 10$ .

Las variables que se utilizaron para la construcción de los indicadores socioeconómicos fueron:

*Condición de alfabetismo.* Situación que distingue a la población de 15 años o más según declare saber leer y escribir un recado. La condición de alfabetismo se clasifica en alfabetos, analfabetas y no especificados. Así, la población analfabeta será aquella que con 15 años o más de edad no sabe leer ni escribir un recado

*Nivel educativo.* Cada una de las etapas que conforman el Sistema Educativo Nacional. Los niveles son: preescolar, primaria, secundaria, estudios técnicos o comerciales con primaria terminada, normal básica, estudios técnicos o comerciales con secundaria terminada, preparatoria o bachillerato, estudios técnicos o comerciales con preparatoria terminada, profesional, maestría y doctorado.

*Vivienda.* Lugar delimitado por paredes y cubierto por techos con entrada independiente, donde generalmente las personas comen, preparan alimentos, duermen y se protegen del ambiente.

*Drenaje.* Sistema de tuberías que permiten desalojar fuera de la vivienda las aguas utilizadas

*Excusado o sanitario.* La disponibilidad de excusado o sanitario.

*Disponibilidad de energía eléctrica.* Clasificación de las viviendas particulares habitadas según la existencia de energía eléctrica para alumbrarla vivienda, independientemente de la fuente de donde provenga.

*Disponibilidad de agua.* Clasificación de las viviendas particulares habitadas de acuerdo con la forma en la que sus ocupantes se abastecen de agua para la realización de sus actividades cotidianas.

*Dormitorio.* Cuarto de la vivienda que se utiliza para dormir, independientemente de que allí se realicen otras actividades.

*Material en pisos.* Material predominante en los pisos de la vivienda.

*Localidad.* Todo lugar ocupado con una o más viviendas, las cuales pueden estar habitadas o no; este lugar es reconocido por la ley o la costumbre. De acuerdo con sus características y con fines estadísticos, se clasifican en urbanas y rurales.

*Tamaño de localidad.* Clasificación de las localidades de acuerdo con el número de personas que las habitan.

*Población ocupada.* Personas que durante la semana de referencia realizaron algún tipo de actividad económica



*Ingreso por trabajo.* Percepción monetaria y/o en especie (ingreso neto) que recibió u obtuvo la población ocupada por el desempeño de su ocupación.

Los indicadores socioeconómicos  $I_{ij}$ , donde  $i = 1, 2, \dots, 210$  municipios y  $j = 1, 2, \dots, 9$  indicadores, se obtienen haciendo ciertos cálculos, tales como, por ejemplo:

El Porcentaje de población de 15 años o más analfabetas

$$I_{11} = \frac{T_{pa}}{T_p - T_{pm}} \times 100$$

$T_{pa}$ = total de población de 15 años o más analfabetas

$T_p$  = total de población de 15 años o más

$T_{pm}$ = total que no especificaron su condición de analfabetas

Porcentaje de ocupantes en viviendas particulares sin disponibilidad de energía eléctrica

$$I_{i4} = \frac{O_{ve}}{O_v - O_{vne}} \times 100$$

$O_{ve}$ = Ocupantes de viviendas particulares sin disponibilidad de energía eléctrica

$O_v$  = Total de ocupantes en viviendas particulares

$O_{vne}$ = Total de ocupantes en viviendas particulares en las que se desconoce si disponen o no de energía eléctrica

En este caso, la nube de puntos estará formada por los 210 municipios observados  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{210}$  en  $\mathbb{R}^{10}$ , los cuales son estandarizados debido a la alta variabilidad existente entre indicadores.

Tabla 5. Indicadores (variables) socioeconómicos usados en el cálculo de índice de marginación.

---

Población total
Población analfabeta de 15 años o más
Población sin primaria completa de 15 años o más
Ocupantes en viviendas sin drenaje ni servicio sanitario exclusivo
Ocupantes en viviendas sin energía eléctrica
Ocupantes en viviendas sin agua entubada
Viviendas con algún nivel de hacinamiento
Ocupantes en viviendas con piso de tierra
Población en localidades con menos de 5 000 habitantes
Población ocupada con ingreso de hasta 2 salarios mínimos

---

Como ya se mencionó, se pretende mejorar el Kernel en ACPK, para ello primero se aplicó ACP y posteriormente se aplicó la técnica propuesta.

Al aplicar ACP se obtuvieron los eigenvalores ( $\lambda_i$ ) y los porcentajes de varianza explicada, donde se puede observar que los 6 primeros componentes principales explican el 93% de la varianza total, (ver Tabla 6).

Tabla 6. Eigenvalores y porcentaje de varianza explicada.

	1	2	3	4	5	6	7	8	9	10
Eigenvalores	2.382	1.105	0.900	0.837	0.715	0.639	0.534	0.467	0.370	0.192
Porcentaje de varianza	0.567	0.122	0.081	0.070	0.051	0.041	0.028	0.022	0.012	0.004
Varianza acumulada	0.567	0.689	0.770	0.840	0.891	<b>0.932</b>	0.961	0.983	0.996	1.000

Estos porcentajes se aprecian mejor en el siguiente gráfico.

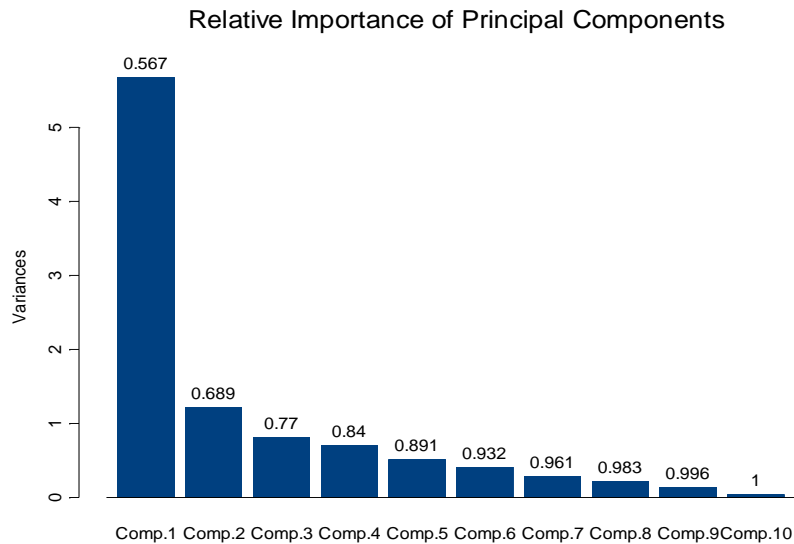


Figura 9. Porcentaje de varianza explicado por ACP.

Como una primera exploración del ACPK empleando el Kernel Gaussiano

$$\exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_1^2}\right)$$

se propone generar valores del parámetro sigma. Es decir,  $\sigma$  tomó 19 valores diferentes generados entre 1 y 10, con incrementos de 0.5, lo cual originó 19 kernels, de los cuales se obtuvieron los eigenvalores para calcular el porcentaje de varianza explicado. Con esta aplicación del ACPK, se observa que a medida que aumenta el valor del parámetro el porcentaje de la varianza explicado por el primer componente principal tiende a incrementarse. Esto se puede apreciar en el siguiente gráfico.

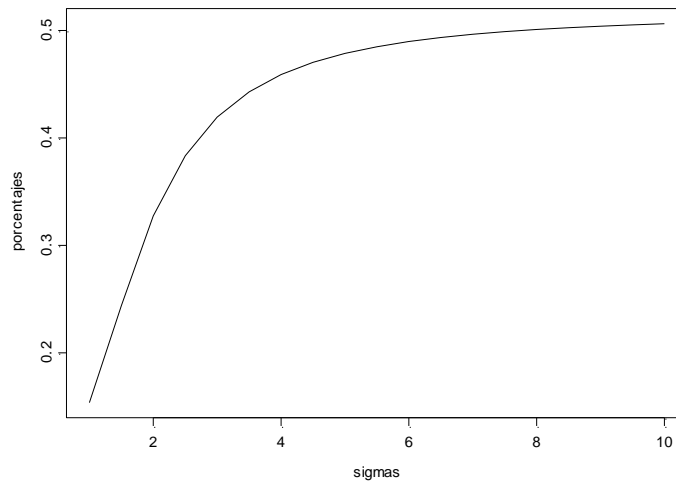


Figura 10. Comportamiento de la varianza al incrementar el valor del parámetro

Nótese que después de cierto valor de sigma el porcentaje de varianza tiende a ser constante, ya que el incremento se presenta en milésimas. En este caso el comportamiento de la varianza explicada solo alcanza un poco más del 50% , es decir ACPK no mejora el porcentaje de varianza explicada (56% ) proporcionado por ACP.

Ahora, al aplicar ACPK y AG con un operador, se obtuvieron 30 porcentajes de varianza derivados de los nuevos kernels de la población inicial. Al usar el método de selección, se eligieron 5 nuevos kernels con mayor porcentaje de varianza explicado por los dos primeros eigenvalores; de tal forma que al aplicar el operador cruce aritmético se generaron 20 nuevos kernels. Después de obtener el porcentaje de varianza, nuevamente se eligieron los 5 mejores nuevos kernels; esto se repitió en 5 iteraciones.

Tabla 7. Porcentajes de varianza explicados por los dos primeros eigenvalores.

Población inicial	Selección 1	Selección 2	Selección 3	Selección 4	Selección 5
0.9994432	0.9995163	0.9995145	0.9995149	0.9995157	0.9995158
0.9993790	0.9995147	0.9995141	0.9995154	0.9995157	0.9995159
0.9993962	0.9995072	0.9995144	0.9995155	0.9995157	0.9995159
0.9994720	0.9995090	0.9995148	0.9995155	0.9995158	0.9995159
0.9993886	0.9995065	0.9995109	0.9995154	0.9995156	0.9995159
0.9994584	0.9995052	0.9995087	0.9995152	0.9995155	0.9995158
0.9994077	0.9995055	0.9995087	0.9995152	0.9995156	0.9995158
0.9994853	0.9995062	0.9995085	0.9995151	0.9995155	0.9995158
0.9994273	0.9995047	0.9995091	0.9995151	0.9995155	0.9995158
0.9994199	0.9995033	0.9995097	0.9995150	0.9995155	0.9995158
0.9993823	0.9995016	0.9995090	0.9995149	0.9995155	0.9995158
0.9993318	0.9994999	0.9995091	0.9995148	0.9995155	0.9995158
0.9994019	0.9994983	0.9995082	0.9995139	0.9995155	0.9995158
0.9994188	0.9994964	0.9995082	0.9995135	0.9995155	0.9995158
0.9994799	0.9994954	0.9995075	0.9995135	0.9995154	0.9995157
0.9993969	0.9994946	0.9995078	0.9995136	0.9995154	0.9995157
0.9993685	0.9994936	0.9995075	0.9995133	0.9995154	0.9995157
0.9992924	0.9994926	0.9995072	0.9995128	0.9995154	0.9995157
0.9994280	0.9994917	0.9995064	0.9995124	0.9995154	0.9995157
0.9992783	0.9994908	0.9995058	0.9995121	0.9995154	0.9995157
0.9993665					
0.9994294					
0.9994143					
0.9993652					
0.9993014					
0.9994934					
0.9994591					
0.9995214					
0.9994045					
0.9993899					

Nótese que los porcentajes de varianza explicados por los dos primeros componentes son muy altos con respecto a los resultados proporcionados por el ACP clásico.

Al observar los valores de la quinta iteración se notó que son cuatro kernels los que proporcionan la máxima varianza explicada. En otras corridas de prueba se requerían entre seis y siete iteraciones para que la proporción de varianza explicada no presentara cambios. En la cuarta iteración se observa que el kernel 4 aporta la máxima varianza

explicada 99.95158% con los dos primeros eigenvectores. En este caso, los kernels que forman la combinación lineal son kernels Gaussianos con valores del parámetro sigma: 30.45975, 30.82689, 33.73920, 35.01175, 39.76460, 45.08298, 60.74184, 62.41404, 68.62161 y 93.63465. Las ponderaciones que multiplican los kernels de la combinación del cuarto kernel son: 0.001493624, 0.01783509, 0.03348784, 0.05917433, 0.06304301, 0.07571440, 0.1272002, 0.1568798, 0.2071066, 0.2580652.

En esta aplicación se encontró un kernel que desde la segunda iteración se identifica como el kernel óptimo. En la Figura 11 se puede observar el proceso iterativo de la propuesta y como la convergencia es rápida a una proporción de varianza explicada de más del 99% con los dos primeros componentes principales.

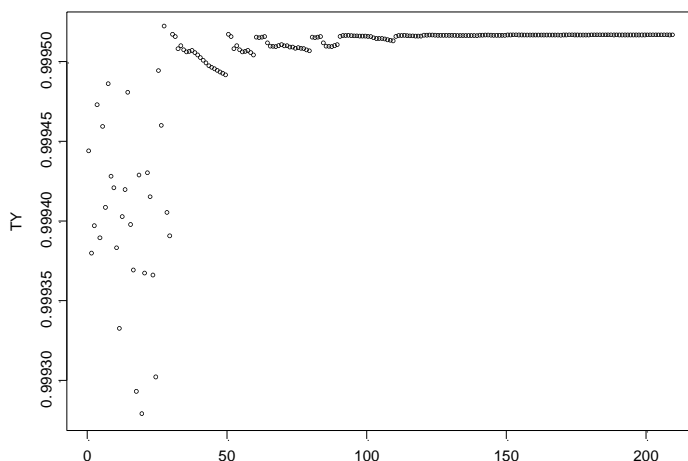


Figura 11. Porcentajes de varianza explicada por los dos primeros componentes principales que produce el algoritmo genético.

Con la finalidad de conocer cuánto explica el primer componente principal, se encontró que el cuarto kernel con el primer componente principal explica el 99.93% de la varianza total, mientras que con el ACP clásico se requieren 9 componentes para explicar el mismo porcentaje, véase la Tabla 8. La Figura 12 muestra la convergencia considerando sólo al primer eigenvector.

Tabla 8. Porcentajes de varianza explicados por los dos primeros eigenvalores.

Población inicial	Selección 1	Selección 2	Selección 3	Selección 4	Selección 5
0.9992460	0.9993450	0.9993425	0.9993431	0.9993442	0.9993444
0.9991591	0.9993429	0.9993421	0.9993438	0.9993442	0.9993444
0.9991824	0.9993327	0.9993424	0.9993439	0.9993442	0.9993444
0.9992851	0.9993352	0.9993430	0.9993440	0.9993443	0.9993444
0.9991722	0.9993318	0.9993377	0.9993438	0.9993441	0.9993444
0.9992666	0.9993300	0.9993348	0.9993435	0.9993440	0.9993444
0.9991980	0.9993304	0.9993347	0.9993435	0.9993440	0.9993444
0.9993030	0.9993313	0.9993345	0.9993434	0.9993440	0.9993443
0.9992245	0.9993293	0.9993353	0.9993434	0.9993440	0.9993443
0.9992146	0.9993274	0.9993360	0.9993433	0.9993440	0.9993443
0.9991637	0.9993252	0.9993351	0.9993432	0.9993440	0.9993443
0.9990952	0.9993229	0.9993353	0.9993430	0.9993440	0.9993443
0.9991901	0.9993207	0.9993341	0.9993418	0.9993439	0.9993443
0.9992130	0.9993181	0.9993340	0.9993412	0.9993439	0.9993443
0.9992958	0.9993168	0.9993331	0.9993413	0.9993439	0.9993443
0.9991834	0.9993156	0.9993335	0.9993414	0.9993438	0.9993443
0.9991449	0.9993143	0.9993331	0.9993410	0.9993438	0.9993443
0.9990419	0.9993130	0.9993327	0.9993403	0.9993438	0.9993442
0.9992254	0.9993117	0.9993317	0.9993398	0.9993438	0.9993442
0.9990229					
0.9991422					
0.9992273					
0.9992069					
0.9991405					
0.9990540					
0.9993140					
0.9992677					
0.9993519					
0.9991937					
0.9991739					

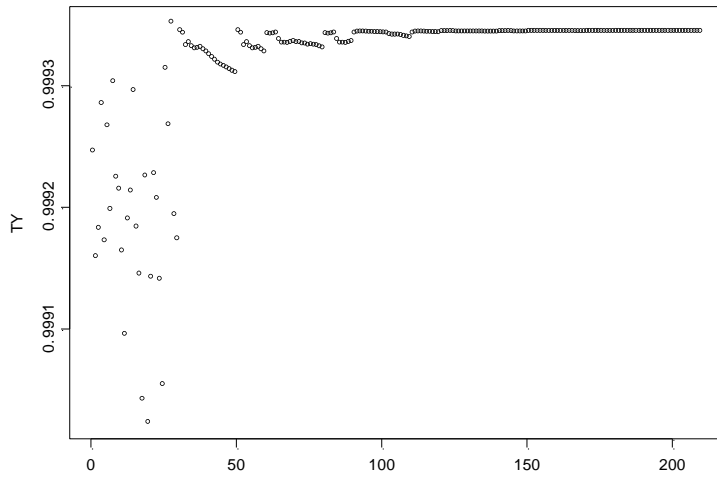


Figura 12. Figura 11. Porcentajes de varianza explicada por el primer componente principal que produce por el algoritmo genético.

Hasta aquí se pudo observar que la metodología propuesta proporciona mejores resultados que el ACP clásico ya que con el primer componente principal prácticamente se explica más del 99% de la variabilidad total de las variables socioeconómicas.

ACPK permite también extraer patrones de máxima variabilidad mediante las proyección ortogonales sobre los primeros eigenvectores principales de las imágenes  $\Phi(\mathbf{x})$  de observaciones  $\mathbf{x}$  en el espacio original. Al proyectar las imágenes de los datos de cada municipio sobre el primer eigenvector del ACPK se puede construir un nuevo índice de marginación. Por lo que se procede a calcular las proyecciones sobre el primer eigenvector  $\alpha^1$  en  $\mathcal{H}$  para cada una de las imágenes bajo  $\Phi$  de las variables de cada municipio. Los diez valores de los  $\sigma$ 's que generaron los kernels y los diez valores óptimos de los  $\theta$ 's que se obtuvieron al aplicar el algoritmo genético se muestran en la tabla 9.



Tabla 9. Valores de los  $\sigma$ 's y de los  $\theta$ 's para los diez kernels.

$j$	1	2	3	4	5	6	7	8	9	10
$\sigma$	18.839	21.683	37.8074	38.604	46.068	48.899	54.235	84.125	91.4185	92.649
$\theta$	0.0048019	0.012347	0.0241879	0.068415	0.088031	0.10764	0.1132976	0.16463	0.203042	0.213611

El kernel óptimo está dado por

$$k(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{10} \theta_j \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_j^2}\right)$$

Las proyecciones ortogonales de cada municipio (cada vector en  $\mathbb{R}^{10}$  con las variables socioeconómicas de los municipios) en el espacio de las características están dadas por

$$v_l = \sum_{i=1}^{210} \alpha_i^1 k(\mathbf{x}_i, \mathbf{x}_l) = \sum_{i=1}^{210} \alpha_i^1 \left[ \sum_{j=1}^{10} \theta_j \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_l\|^2}{2\sigma_j^2}\right) \right], \quad l = 1, \dots, 210,$$

y  $\boldsymbol{\alpha}^1 = (\alpha_1^1, \alpha_2^1, \dots, \alpha_{210}^1)'$  es el primer eigenvector obtenido del ACPK usando el kernel óptimo. Se puede considerar al valor  $v_l$  como un índice de marginación extraído mediante un ACPK.

Los municipios ordenados ascendentemente de acuerdo a sus valores  $v_l$  se muestran en la Tabla 10. Se tiene que entre los municipios con los valores más pequeños están Veracruz, Xalapa, Coatzacoalcos, Boca del Río, Orizaba, Poza Rica y Córdoba, aunque también aparecen municipios que no se esperaría ver de acuerdo al índice de marginación calculado como el ACP. Por otro lado, los municipios con los valores mayores son Yecuatla, Acultzingo, Misantla, Tlaxicoyan, Juan Rodríguez Clara, Sayula de Alemán y Hueyapan de Ocampo. La posibilidad de que estos índices

basados en el ACPK estén descubriendo algún de patrón de variabilidad que no descubre el ACP y el cual además tenga sentido en un contexto socioeconómico es un tema de discusión e investigación futura.

Tabla 10. Índices de Marginación obtenidos con ACPK.

Municipio	Índice	Municipio	Índice	Municipio	Índice
Veracruz	14.4695	Comapa	14.48153	Tempoal	14.48314
Xalapa	14.47174	Chiconamel	14.48155	Espinal	14.48316
Mixtla de Altamirano	14.47476	Texhuacán	14.48174	Landro y Coss	14.48316
Tehuipango	14.47485	Pueblo Viejo	14.48176	Huayacocotla	14.48317
Coatzacoalcos	14.47628	Minas, Las	14.48178	Actopan	14.4832
Filomeno Mata	14.4768	Carlos A. Carrillo	14.48181	Tonayán	14.48322
Soteapan	14.4774	Papantla	14.48184	Ixhuatlán del Sur	14.48323
Boca Del Río	14.47743	Chumatlán	14.4819	Pánuco	14.48323
Orizaba	14.4779	Naranjos Amatlán	14.48196	Cazones	14.48326
Poza Rica	14.47832	Fortín	14.48196	Zentla	14.48327
Nanchital	14.47855	Magdalena	14.48198	Coyutla	14.48328
Texcatepec	14.47914	Calchahuaco	14.48199	Chiconquiaco	14.48328
Mecayapan	14.47928	Cosoleacaque	14.48201	Perote	14.48329
Zozocolco	14.4793	Coatepec	14.48203	Rafael Delgado	14.48329
Río Blanco	14.47939	Aguila	14.48214	San Andrés Tux.	14.48329
Carrillo Puerto	14.47941	Tenampa	14.48225	Temapache	14.48329
Mecatlán	14.47946	Nogales	14.48226	Apazapan	14.4833
Córdoba	14.47954	Coahuatlán	14.48234	Emiliano Zapata	14.48331
Tantoyuca	14.47958	Camarón d Tejada	14.48242	Cuitláhuac	14.48332
Ilamatlán	14.48013	Tatatila	14.48244	Tlaltetela	14.48333
Soledad Atzompa	14.48036	Tlachichilco	14.48245	Chicontepec	14.48333
Zontecomatlán	14.48037	Cosamaloapan	14.4825	Tantima	14.48333
Astacinga	14.48038	Martínez d la Torre	14.48252	Alto Lucero	14.48334
Tatahuicapan	14.48041	Zaragoza	14.48255	Tenochtitlán	14.48334
Atlahuilco	14.48057	Ixhuatlán deMadero	14.48265	Acula	14.48337
Lerdo de Tejada	14.4807	Alpatláhuac	14.48266	Higo, El	14.48337
Reyes, Los	14.48073	Coetzala	14.48266	Jalcomulco	14.4834
Minatitlán	14.48085	Amatitlán	14.48268	Chontla	14.48344
Zacualpan	14.48086	Acatlán	14.48272	Jilotepec	14.48344
Ursulo Galván	14.48087	Huiloapan	14.48272	Ixtaczoquitlán	14.48345
Pajapan	14.4809	Tuxtilla	14.48277	Teocelo	14.48345
Banderilla	14.48091	Ixhuacán d los Reyes	14.4828	Tezonapa	14.48345
Cerro Azul	14.48094	Coatzintla	14.48282	Miahuatlán	14.48348
Coxquihui	14.48102	Choapas, Las	14.48282	Naranjal	14.48348
Ayahualulco	14.48112	Jáltipan	14.48282	Tecolutla	14.4835
Túxpam	14.48114	Tlacolulan	14.48285	Tihuatlán	14.4835
Camerino Z. Mendoz	14.48115	Alvarado	14.48287	Tampico Alto	14.48351
Antigua, La	14.48117	Puente Nacional	14.4829	Yanga	14.48351
Perla, La	14.48117	Ixcatepec	14.48296	Xico	14.48354
Tequila	14.4812	Rafael Lucio	14.483	Tlacotepec	14.48354
Xoxocotla	14.48125	Oluta	14.48304	Atzalan	14.48355
Oteapan	14.48126	Chalma	14.48307	Tepetlán	14.48358
Zongolica	14.48128	Cotaxtla	14.48309	Vega d Alatorre	14.48358
San Andrés Te.	14.48129	Hidalgotitlán	14.4831	Naolinco	14.4836
Tlaquilpa	14.48135	Tepatlaxco	14.48313	Tamiahua	14.48362
Agua Dulce	14.48136	Ozuluama	14.48314	Villa Aldama	14.48364

Continuación de la Tabla 10.

<b>Municipio</b>	<b>Índice</b>	<b>Municipio</b>	<b>Índice</b>	<b>Municipio</b>	<b>Índice</b>
Cosautlán de Carv.	14.48366	Tlapacoyan	14.484	Tlilapan	14.48424
Chinampa	14.48368	Atzacan	14.48402	Ixhuatlán Del Café	14.48426
Paso de Ovejas	14.48368	Nautla	14.48402	Tlacojalpan	14.48426
Coscomatepec	14.4837	Tepetzintla	14.48402	Huatusco	14.48427
Saltabarranca	14.4837	Las Vigas d Ramírez	14.48403	Juchique De Ferrer	14.48427
Benito Juárez	14.48371	Manlio Fabio Altam.	14.48405	Chinameca	14.48429
Medellín	14.48372	Chacaltianguis	14.48406	San Juan Evangelist	14.48433
Tierra Blanca	14.48373	Maltrata	14.48406	Santiago Tuxtla	14.48433
Tomatlán	14.48373	Coacoatzintla	14.48408	Ixhuatlancillo	14.48434
Otatitlán	14.48375	Acajete	14.48411	José Azueta	14.48436
Colipa	14.48379	Amatlán de los Reyes	14.48412	Angel R. Cabada	14.48437
Castillo de Teayo	14.4838	Cuichapa	14.48413	Catemaco	14.48439
Citlaltépetl	14.48381	Tres Valles	14.48413	Tamalín	14.4844
Sochiapa	14.48382	Playa Vicente	14.48414	Isla	14.48441
Uxpanapa	14.48383	Paso Del Macho	14.48415	Soledad De Doblado	14.48442
Ignacio de la Llave	14.48385	Altotonga	14.48417	Jesús Carranza	14.48443
Tlalnelhuayocan	14.48387	Chocamán	14.48418	Soconusco	14.48448
Atoyac	14.48389	Jalacingo	14.48418	Yecuatla	14.48449
Acayucan	14.48395	Platón Sánchez	14.48418	Acultzingo	14.48455
Mariano Escobedo	14.48396	Moloacán	14.48419	Misantla	14.48456
Gutiérrez Zamora	14.48397	Tancoco	14.48419	Tlalixcoyan	14.48458
Ixmatlahuacan	14.48398	Texistepec	14.48421	Juan Rodríguez Clara	14.48467
Tlacotalpan	14.48398	Totutla	14.48421	Sayula De Alemán	14.48468
Omealca	14.484	Jamapa	14.48423	Hueyapan de Ocampo	14.48469

## Capítulo 4. Conclusión e Investigación Futura

En este trabajo se ha presentado un nuevo algoritmo para construir un kernel en el ACPK que mejora los resultados del ACPK cuando se aplica en base a un solo kernel. La propuesta se basa en la selección del kernel que mayor proporción de varianza relativa explica dentro de un conjunto convexo de kernels. El problema de maximización se resuelve con un algoritmo genético como herramienta de optimización. La construcción del kernel se ilustró con un problema para el cual se requiere de la construcción de un índice. Para el ACPK se usó el kernel Gaussiano. Se observó que al aumentar el valor del parámetro  $\sigma$  del kernel, el porcentaje de varianza explicado por el primer componente principal tiende a aumentar para estabilizarse en el porcentaje de varianza explicada por el primer componente principal del ACP en los datos. Es decir, los porcentajes de varianza obtenidos por el ACPK están por debajo de los porcentajes de varianza proporcionados por el primer componente del ACP.

El ACP requiere los 6 primeros componentes para explicar el 93% de la varianza total. El algoritmo propuesto, da un porcentaje de varianza explicado del 99% con los 2 primeros componentes. El algoritmo genético proporciona 4 kernels óptimos (subóptimos), uno de los cuales se identifica desde la segunda iteración. Cabe señalar que los kernel identificados como óptimos explican con el primer componente el 99.93% de la varianza total. Este hecho se debe considerar ya que para obtener los índices de marginación de los municipios con el ACP se emplea el primer componente principal.

La implementación del truco kernel en las técnicas del análisis multivariado ha sido un área de intensa investigación y aplicación en los últimos diez años. Sin embargo aún falta mucho que hacer. Algunos problemas que se identifican para continuar sobre esta línea de investigación son:

1) Abordar el problema de la reconstrucción de las imágenes en el espacio original de las proyecciones ortogonales de los  $\Phi(\mathbf{x}_i)$  en los eigenvectores en el espacio de las características. Este problema pone un reto matemático porque es un problema mal condicionado en el sentido de que el mapeo  $\Phi$  no es uno a uno y por lo tanto la imagen inversa de un punto en el espacio de las características puede estar formada por varios puntos en el espacio original o incluso puede no existir. Se han hecho propuestas para resolver este problema. Una continuación de este trabajo es implementarlas para el kernel que se construye en esta tesis.

2) La selección de kernel se puede extender a espacios de kernels más complejos que el que se usa en esta tesis. Lo cual sería necesario ante conjuntos de datos para los cuales la propuesta que se hace en este trabajo no funcione.

3) Ya se ha investigado la relación entre la geometría del espacio  $\mathcal{H}$  con el kernel  $k$ . Es posible que los resultados de estas investigaciones den elementos para la selección adecuada del kernel. Por ejemplo, ¿será posible seleccionar o construir una función  $\Phi$  que transforme a los datos en datos que sigan una distribución elíptica? En principio, ¿existirá dicho mapeo?, si es que existe, ¿se podrá construir un kernel que evalúe los productos internos entre las imágenes bajo este mapeo?

4) El problema de selección de kernel en análisis de correlación canónica está abierto. Una posibilidad para abordarlo es seguir una estrategia análoga a la que se propone en esta tesis.

## Referencias

- 1) Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan.
- 2) Johnson, R., and Wichern, D. (1999). *Applied Multivariate Statistical Analysis*. Prentice Hall, Fourth Edition. New Jersey.
- 3) Kim, S., Magnani, A. and Boyd, S. (2006) *Optimal Kernel Selection in Kernel Fisher Discriminant Analysis*. Department of Electrical Engineering. Stanford University. Stanford, C.A. USA.
- 4) Mika, S., Rätsch G., Weston J., Schölkopf, B., and Müller, K.(2000). *Fisher Discriminant Análisis with Kernels*. Berlin, Germany.
- 5) Mika, S. (2002). *Kernel Fisher Discriminants*. Von der Fkultät IV – Elektrotechnik und Informatik der Technischen Universität Berlin. Berlin, Germany.
- 6) Montano, J.A. y Cantú, M. (2011). *Algoritmos Genéticos en la Discriminación. Una Solución Óptima para Clasificación*. Editorial Académica Española. España.
- 7) Schölkopf, B., and Smola, A. (2002). *Learning with Kernels*. Cambridge. The MIT Press Cambridge.
- 8) Schölkopf, B., Smola, A. and Muller, K.(1998). *Nonlinear component analysis as a kernel eigenvalue problem*. Neural Computation. 10:1299-1319.
- 9) Schölkopf, B., Mika, S., Burges, C., Knirsch, P., Müller, K., Rätsch G., and Smola, A. (1999). *Input Space vs. Feature Space in Kernel-Based Methods*. IEEE Transactions on Neural Networks.
- 10) Taylor, J., and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. New York. Cambridge.
- 11) Valdez, E.A., and Chernih, A. (2003). Wang's capital allocation formula for elliptically contoured distributions. *Insurance: Mathematics and Economics*, Vol. 33, Issue 3, pp 517-532.

## Apéndice

Algoritmo del ACPK

- 1) Partimos de una matriz  $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^d$ , donde cada columna representa el dato de un individuo.
- 2) Se obtiene el valor promedio de cada una de las columnas.
- 3) Se elige la función kernel  $k(\cdot, \cdot)$  que será programada.
- 4) Se obtiene la matriz Gram  $[\mathbf{K}]_{ij} = [k(x_i, x_j)]$  de orden  $n \times n$ .
- 5) Se obtiene la matriz centrada a partir de la matriz Gram

$$\tilde{\mathbf{K}} = \left( I - \frac{\mathbf{1}_{n \times n}}{n} \right)' \mathbf{K} \left( I - \frac{\mathbf{1}_{n \times n}}{n} \right)$$

donde  $\mathbf{1}$  es una matriz de orden  $n \times n$  con todas las entradas iguales a 1 .

- 6) La matriz  $\tilde{\mathbf{K}}$  es diagonalizada para obtener las parejas de  $\lambda_i$  y  $\alpha_i$ ,  $i = 1, \dots, n$  valores y vectores característicos, respectivamente.
- 7) Se normalizan los vectores  $\alpha_i \leftarrow \frac{\alpha_i}{\sqrt{\lambda_i}}$
- 8) Se retienen los  $c$  primeros vectores característicos correspondientes a los  $c$  valores característicos más grandes, tal que

$$\frac{\sum_{j=1}^c \lambda_j}{\sum_{j=1}^n \lambda_j}$$

- 9) Se proyectan los puntos en los vectores característicos.

$$y = \alpha' \left( I - \frac{\mathbf{1}_{n \times n}}{n} \right) \left( \begin{bmatrix} k(x_1, x) \\ \vdots \\ k(x_n, x) \end{bmatrix} - \mathbf{K} \frac{\mathbf{1}_{n \times 1}}{n} \right)$$

donde es una matriz con todas las entradas igual a 1 y usa las proyecciones en lugar de los puntos de los datos.

## Programa en S- Plus.

```
kernel <- function(x, y, sigma)
{
  k <- exp( - vecnorm(x - y, p = 2)/(2 * sigma^2))
  return(k)
}
convex <- function(X, N, w, semilla, minsig, maxsig, L)
{
  n <- length(X[, 1])
  p <- length(X[1, ])
  q <- w + 1
  set.seed(semilla)
  Qk <- array(0, dim = c(n, n, w))
  Gk <- array(0, dim = c(n, n, w))
  medias <- matrix(, 1, p)
  for(i in 1:p) {
    medias[, i] <- sum(X[, i])/n
  }
  for(j in 1:p) {
    X[, j] <- X[, j] - medias[, j]
  }
  print("matriz centrada")
  I <- diag(n)
  unos <- matrix(rep(1, n^2), nrow = n, ncol = n)
  B <- unos/n
  A <- I - B
  sig <- runif(w, minsig, maxsig)
  sig <- sort(sig)
  for(l in 1:w) {
    for(i in 1:n) {
      for(j in 1:n) {
        Gk[i, j, l] <- kernel(X[i, ], X[j, ], sig[l])
      }
    }
  }
  a <- 1
  H <- matrix(, N, q)
  TY <- matrix(, N, q)
  z <- 1
  repeat {
    pr <- 0
    r <- matrix(0, nrow = n, ncol = n)
    theta <- runif(w, min = 0, max = 1)
    theta <- (theta)/sum(theta)
    theta <- sort(theta)
    for(l in 1:w) {
      Qk[, , l] <- theta[l] * Gk[, , l]
    }
    for(l in 1:w) {
      r <- r + Qk[, , l]
    }
    er <- eigen(r)
    pr <- (er$values[1] + er$values[2])/sum(er$values)
    for(l in 1:w) {
      H[a, l] <- theta[l]
    }
  }
}
```



```

    }
    H[a, q] <- pr
    a <- a + 1
    if(a > N)
        break
}
# print(H)
#print("MATRIZ ORDENADA")
TY[, z] <- H[, q]
b <- 1
repeat {
    d <- H[rev(order(H[, q])), 1:q]
    # print(d)
    print("SELECCION DE RENGLONES DE TETAS")
    S <- select.rows(d, 1:5, )
    #print(S)
    t <- choose(5, 2, order.matters = F)
    zz <- (2 * t)
    #print(zz)
    tec <- matrix(, zz, w)
    # Cruce Aritmético
    f <- 0
    for(i in 1:5) {
        for(j in 1:5) {
            if(i < j) {
                r <- runif(1, min = 0, max = 1)
                #print(r)
                vy <- r * S[i, ] + (1 - r) * S[j,
                ]
                f <- f + 1
                tec[f, ] <- vy
                r <- runif(1, min = 0, max = 1)
                vx <- (1 - r) * S[i, ] + r * S[j,
                ]
                #print(r)
                f <- f + 1
                tec[f, ] <- vx
            }
        }
    }
    #print(tec)
    # FIN DEL CRUCE
    H <- matrix(, zz, q)
    for(i in 1:zz) {
        for(j in 1:w) {
            H[i, j] <- tec[i, j]
        }
    }
    pr <- 0
    r <- matrix(0, nrow = n, ncol = n)
    for(i in 1:zz) {
        for(l in 1:w) {
            Qk[, , l] <- tec[i, l] * Gk[, , l]
        }
        for(l in 1:w) {
            r <- r + Qk[, , l]
        }
    }
    er <- eigen(r)

```

```
        pr <- (er$values[1] + er$values[2])/sum(er$values)
        H[i, q] <- pr
    }
    z <- z + 1
    TY[, z] <- H[, q]
    #print(H)
    b <- b + 1
    if(b > L)
        break
}
# print(TY)
plot(TY)
}
```

## Datos

Datos estandarizados de los indicadores socioeconómicos de los Municipios del Estado de Veracruz.

Población total	% Población analfabeta de 15 años o más	% Población sin primaria completa de 15 años o más	% Ocupantes en viviendas sin drenaje ni servicio sanitario exclusivo	% Ocupantes en viviendas sin energía eléctrica	% Ocupantes en viviendas sin agua entubada	% Viviendas con algún nivel de hacinamiento	% Ocupantes en viviendas con piso de tierra	% Población en localidades con menos de 5 000 habitantes	% Población ocupada con ingreso de hasta 2 salarios mínimos
-0.48	0.02	0.20	0.29	-0.29	-1.13	0.51	-0.51	0.81	0.73
-0.58	-0.49	-0.10	-1.30	-1.13	-1.31	-0.20	-1.55	0.81	0.03
0.86	-0.41	-0.74	-0.21	-0.37	0.08	-0.40	-0.75	-1.06	-0.66
0.12	-0.34	-0.06	-0.46	-0.99	-1.02	-1.51	-1.12	0.81	-0.27
-0.53	0.21	0.31	1.03	-0.16	1.68	-0.58	-0.40	0.81	0.10
-0.29	-0.19	0.14	0.39	0.20	0.27	0.95	0.51	-0.17	0.29
-0.52	-0.56	0.51	2.24	-0.71	-1.13	0.01	-0.49	0.81	0.70
-0.46	1.20	1.34	-0.84	-0.40	0.03	0.99	1.30	0.81	1.05
-0.11	-0.01	0.34	-0.69	-0.99	-1.14	-1.02	-1.10	0.81	0.34
0.39	0.28	0.51	-0.45	1.00	-0.31	0.79	-0.25	-0.08	0.32
0.32	-0.69	-0.95	-0.03	-0.65	-0.15	-1.23	-1.19	-0.98	-1.63
-0.49	-0.02	0.04	0.77	-0.81	0.87	-1.49	-0.68	0.81	-1.49
-0.12	-1.11	-1.51	-1.03	-0.32	-0.53	-1.56	-0.85	-1.43	-1.53
0.07	-0.85	-0.83	0.18	-0.75	0.85	-0.11	-0.48	0.19	-0.45
-0.01	0.20	0.27	0.22	-0.06	-0.57	-1.05	-0.75	-0.22	-0.20
-0.18	-1.24	-1.37	-0.93	-1.04	-1.24	-1.52	-1.27	-1.50	-1.77
-0.56	-0.70	-0.19	-0.19	-0.96	-1.03	-1.32	-0.09	0.81	0.88
-0.59	1.45	1.54	0.75	-0.52	0.58	1.40	0.98	0.81	1.27
-0.52	2.43	1.87	-0.14	0.61	1.68	1.73	1.52	0.81	0.14
-0.47	2.27	1.91	0.33	0.11	0.36	1.65	2.20	0.81	0.90
-0.20	-0.85	-0.79	-0.26	-0.56	-0.39	-0.56	-0.75	-1.07	-0.87
-0.30	0.37	0.14	0.29	-0.23	-0.78	1.01	0.64	-0.58	0.72
0.29	0.39	0.82	0.04	1.33	-0.14	0.86	0.46	0.81	0.96
-0.37	0.29	1.03	1.16	-0.51	-0.84	0.73	0.48	0.81	1.11
-0.24	2.23	1.98	-0.37	-0.67	-1.15	1.45	0.12	0.81	1.03
-0.31	-1.16	-1.68	-1.20	-1.07	-1.11	-0.67	-1.46	-2.06	-1.52
-0.32	0.90	0.35	-0.48	0.13	0.62	0.25	1.30	0.81	0.83
1.96	-1.48	-2.31	-1.23	-1.15	-1.23	-1.89	-1.51	-2.23	-3.08
-0.42	1.71	1.74	-0.25	-0.12	-0.27	1.47	1.23	0.81	1.07
0.12	-0.86	-1.74	-0.95	-1.03	-0.96	-1.11	-0.90	-1.91	-1.96
-0.35	0.43	1.02	4.02	1.43	1.00	0.88	0.24	0.81	0.89
0.24	0.05	-0.11	-0.67	-0.15	-0.84	-0.50	-0.29	-0.78	0.15
-0.17	-0.26	-0.11	-0.96	0.10	1.83	0.21	0.84	0.81	0.51
-0.16	-1.16	-1.71	-1.08	-0.69	-1.07	-1.59	-0.96	-1.85	-1.81
-0.41	0.13	0.17	0.84	0.71	0.39	-0.51	1.12	0.81	0.66
-0.49	-0.07	0.20	-0.21	0.49	-0.90	0.84	-0.62	0.81	0.38
-0.50	1.29	1.42	-0.19	-0.17	1.20	1.01	1.56	0.81	0.82
0.77	-1.00	-1.35	-1.08	-1.00	-1.23	-1.02	-1.02	-1.34	-1.09
4.46	-1.32	-2.05	-1.21	-1.00	-0.94	-1.30	-1.47	-2.12	-2.11
0.12	-0.91	-1.49	-0.99	-0.33	-0.31	-1.05	-0.38	-1.22	-1.30
-0.59	0.31	0.63	0.05	0.31	-1.20	1.19	1.60	0.81	1.29
-0.51	0.03	0.31	1.34	0.57	-0.73	-0.10	0.22	0.81	0.74
-0.30	0.22	1.10	2.86	-0.14	0.21	0.94	0.48	0.81	1.31
2.75	-1.25	-1.82	-1.12	-1.05	-0.71	-1.14	-1.14	-1.66	-1.65

0.41	-0.71	-1.06	-0.50	-0.81	-0.36	-1.58	-1.13	-0.80	-1.72
-0.33	-0.10	0.85	-0.13	-0.66	-1.24	0.51	0.07	0.81	1.10
0.17	0.92	0.95	-0.29	-0.51	-0.64	1.03	0.77	-0.11	0.30
1.23	-0.95	-1.52	-0.84	-0.52	0.42	-0.65	-1.20	-1.53	-1.31
-0.27	-0.19	0.27	2.31	-0.17	0.12	-0.09	-0.37	0.81	0.01
-0.35	1.20	0.43	1.43	2.50	1.47	0.58	1.36	0.81	0.80
-0.22	1.13	0.59	0.48	0.58	1.01	0.85	1.20	-0.32	0.80
-0.42	-0.58	-0.48	-0.14	-0.71	-0.91	-0.15	-0.21	0.81	-0.69
-0.18	-0.69	-0.58	-0.51	-0.86	-0.31	-1.04	-1.13	-0.78	-1.40
-0.40	-0.35	0.07	0.32	-0.33	0.39	-1.20	-0.79	0.81	-0.27
-0.38	0.60	0.07	-0.27	-0.03	2.10	-0.55	0.73	0.81	0.47
-0.50	1.23	0.66	0.31	0.99	2.61	0.43	1.23	0.81	0.94
-0.38	0.60	1.22	-0.03	0.92	0.06	1.40	0.39	0.81	0.99
0.49	-0.01	-0.29	-0.78	-0.37	1.31	-0.39	1.25	0.81	0.58
-0.36	-0.64	-0.59	-0.23	-0.40	0.50	-0.57	-0.75	-0.70	-0.53
-0.36	-0.17	-0.09	-0.59	0.65	1.49	0.24	0.55	0.81	0.79
0.76	-0.24	-0.01	0.75	1.63	1.30	-0.12	-0.35	-0.92	-0.46
-0.34	0.08	0.42	-0.55	-0.41	-0.67	0.53	0.26	-1.03	0.24
-0.34	-0.09	0.29	-0.30	1.11	0.15	-0.43	1.53	0.81	0.80
-0.56	1.64	0.66	0.65	1.11	1.45	0.94	1.53	0.81	1.10
0.22	-1.04	-0.74	-0.30	-0.93	-1.14	-1.04	-1.13	0.36	-0.46
-0.17	0.38	0.06	-0.56	0.09	1.97	-0.01	0.98	0.81	0.71
-0.42	2.64	1.94	3.50	0.04	1.78	2.27	1.68	-1.67	0.81
0.25	-1.13	-1.54	-0.86	-0.84	-0.53	-1.05	-0.96	-1.69	-1.55
-0.12	-0.84	-0.94	-0.76	-0.32	0.65	-0.56	-0.22	-0.86	0.00
-0.28	-0.47	0.18	0.56	0.56	1.83	0.93	-0.08	0.81	0.95
0.26	-0.43	-0.28	-0.73	-0.26	-0.91	-0.02	-0.21	-0.96	0.01
-0.28	0.28	0.25	-0.08	2.20	0.35	-0.07	-0.17	0.81	-0.18
0.13	0.32	0.20	0.42	0.31	0.14	-0.76	0.22	0.21	0.13
-0.52	-1.08	-1.41	-0.75	-0.90	-1.17	-0.37	-0.78	0.81	-1.00
-0.29	0.37	0.27	0.67	0.15	1.45	-0.50	-0.36	0.81	0.28
-0.38	2.73	1.77	-0.10	1.27	1.79	0.91	1.65	0.81	0.67
0.11	-0.33	-0.08	-0.33	-0.40	-0.27	-0.29	-0.79	-1.08	-0.33
-0.38	0.19	-0.06	-0.69	0.83	1.03	0.27	1.77	0.81	1.01
-0.44	0.82	1.45	-0.72	-0.45	-1.04	1.00	0.41	0.81	1.20
-0.25	0.03	0.45	-0.24	-0.32	0.44	0.63	0.94	-0.18	0.89
-0.40	0.78	0.21	-0.74	0.06	-0.36	-0.09	0.22	-0.57	-0.32
-0.37	-0.75	-0.74	-0.31	-0.05	0.81	-0.03	-1.06	-1.25	-1.25
0.31	0.75	0.44	-0.27	0.16	1.87	0.27	1.61	0.81	0.87
-0.51	0.05	0.25	0.89	0.61	0.73	-0.67	-0.60	0.81	-0.01
0.46	-0.86	-1.24	-0.56	-0.73	-1.01	-0.66	-0.56	-0.79	-0.73
0.01	0.07	0.50	-0.43	-0.19	-0.64	1.12	-0.10	-0.11	0.86
6.81	-1.42	-2.33	-1.26	-1.10	-1.26	-1.89	-1.43	-2.11	-2.68
-0.54	-0.15	1.11	0.37	-0.88	-1.23	-0.34	0.19	0.81	0.95
0.09	-0.82	-1.18	-0.57	-0.53	0.59	-0.77	-1.22	-1.66	-0.77
-0.44	-0.68	-0.24	0.80	-0.23	0.60	-0.13	-0.39	0.81	-0.41
-0.14	-0.20	0.16	0.10	-0.07	1.11	0.06	-0.46	0.81	0.26
-0.08	-0.20	-0.22	-0.84	-0.72	-1.21	0.15	-0.40	-1.43	-0.37
-0.38	-0.80	-0.85	-0.57	-0.88	-1.13	-0.05	-1.14	0.81	-0.03
0.01	-0.11	0.17	0.31	-0.33	-0.21	-0.54	-0.87	-0.29	0.23
-0.27	0.33	0.57	0.45	-0.40	-0.76	0.21	0.24	0.81	0.88
-0.60	-0.41	0.00	-0.97	-1.05	-0.94	-1.04	-1.05	0.81	0.64
-0.24	-0.71	-1.09	-0.85	-0.92	-0.95	-2.11	-1.51	-2.00	-1.77
-0.58	0.98	0.82	0.35	-0.55	-1.22	1.53	1.84	0.81	1.08
-0.35	0.15	0.09	-0.35	-0.55	-0.50	0.62	-0.23	-1.33	0.20
-0.23	-0.68	-0.01	0.80	-0.76	-0.44	-0.57	-0.78	0.81	-0.14
-0.08	-0.29	-0.73	-0.59	-0.63	-0.81	-0.72	-0.43	-0.50	-1.12
1.64	-0.86	-1.20	-0.72	-0.67	0.51	-0.86	-0.68	-1.03	-0.94
-0.43	2.16	1.46	1.54	2.11	1.59	1.69	1.84	0.81	0.92
-0.34	1.04	1.07	4.01	0.99	-0.51	1.05	1.41	0.81	0.52
0.04	-0.91	-0.88	-0.19	-0.55	0.59	-0.65	-1.04	-0.09	-1.09
-0.55	-0.53	0.30	-0.65	-0.90	-1.00	0.80	-0.36	0.81	1.07
-0.58	0.77	1.56	-0.31	2.22	-0.71	1.37	0.03	0.81	0.67
2.29	-1.12	-1.60	-0.47	-0.14	0.16	-1.01	-1.06	-1.37	-1.78
0.53	-0.72	-0.63	-0.64	-0.19	-0.14	-0.20	-0.12	-0.34	0.01
-0.47	4.22	2.46	-0.01	4.36	0.70	1.62	2.09	0.81	1.14
-0.31	-0.64	-0.35	-0.09	0.52	0.62	-0.18	-0.88	-0.67	-0.54
-0.28	-0.56	-0.54	-0.81	-0.88	-1.18	-0.26	-1.25	-0.55	0.09

-0.55	-0.22	0.38	1.13	-0.69	-0.57	0.88	0.98	0.81	0.73
-0.44	-0.76	-0.54	-0.34	-0.32	0.50	-1.15	-0.35	0.81	0.19
-0.04	-1.05	-1.55	-0.87	-0.98	-1.23	-0.70	-1.00	-1.22	-1.25
-0.37	-0.39	-0.70	-0.86	-0.46	0.56	-0.47	-1.24	-1.78	-0.43
-0.21	-0.16	0.02	1.35	-0.37	-0.85	0.13	0.08	0.81	0.12
1.63	-1.47	-2.48	-1.27	-1.11	-1.27	-1.99	-1.58	-2.25	-2.53
-0.53	-0.71	-0.49	-0.35	-0.59	-0.46	-1.32	-0.69	0.81	-0.50
-0.40	-0.39	-0.23	1.33	-0.59	2.29	0.33	-0.24	-2.24	0.25
-0.16	-0.62	-0.16	-0.52	1.88	0.72	-0.77	0.12	0.81	0.19
-0.36	1.60	1.52	2.56	0.32	-0.49	1.05	0.90	-0.78	1.11
1.10	-0.99	-0.99	-0.87	0.16	-0.09	-0.53	-0.85	-0.65	-1.11
2.61	-0.54	-0.76	-0.85	0.36	1.20	0.01	0.37	-0.52	-0.30
-0.12	-0.55	0.16	1.17	-0.40	-0.36	0.06	-0.70	-0.55	-0.25
-0.04	-0.85	-0.33	-0.12	-0.93	-1.02	-0.75	-1.03	0.13	-0.94
-0.28	2.02	2.00	0.00	-0.13	-0.92	1.70	1.40	0.81	0.70
0.41	-0.66	-0.77	-0.78	-0.90	-1.27	0.05	-1.17	-0.93	-0.13
-0.29	0.37	-0.08	-0.60	0.40	0.45	-0.31	0.63	-0.93	-0.30
0.31	0.25	0.45	-0.10	-0.08	1.51	0.01	-0.19	0.34	0.34
2.28	-1.39	-2.24	-1.27	-0.96	-0.80	-1.80	-1.27	-2.22	-2.34
-0.36	-0.30	-0.08	-0.45	-0.64	-0.75	0.54	-1.01	-0.89	0.02
0.33	-1.25	-1.59	-0.89	-0.53	-0.50	-0.19	-1.07	-1.65	-1.91
-0.26	-0.92	-0.42	0.27	-0.92	-1.01	-1.47	-0.92	0.81	-1.06
-0.35	-0.12	-0.27	-0.12	-0.10	-0.15	1.09	0.34	-1.91	0.25
-0.52	-0.85	-1.01	-0.96	-0.90	-1.18	0.41	-1.14	0.81	0.00
-0.55	1.73	1.40	-0.32	1.83	0.76	1.34	2.24	0.81	0.91
0.12	-1.41	-2.33	-1.24	-1.10	-1.26	-1.45	-1.50	-2.25	-1.98
-0.52	0.20	0.35	0.03	-0.46	-0.59	-1.33	-1.00	0.81	-0.36
-0.58	0.58	1.13	2.12	-0.83	-1.09	1.79	1.12	0.81	0.85
2.08	0.26	0.11	-0.03	-0.26	-0.27	-0.27	0.42	-0.37	-0.01
0.00	-0.14	0.29	0.18	-0.22	1.08	-0.12	-0.51	0.81	0.53
0.41	0.46	0.35	1.12	0.39	0.09	-0.21	0.07	-0.05	0.32
-0.09	0.04	-0.04	-0.17	-0.21	0.79	0.17	-0.47	-0.37	0.47
-0.41	-0.36	-0.52	-0.55	-0.23	-0.31	-0.01	-0.59	0.81	-0.38
-0.57	-0.28	0.68	0.20	-0.37	-1.14	0.72	0.35	0.81	0.94
-0.31	2.33	1.90	0.10	-0.83	0.36	2.04	1.81	0.81	0.94
-0.11	-0.56	0.02	0.76	-0.48	-0.62	-0.18	-0.52	-0.54	-0.33
-0.10	2.28	1.74	4.39	0.77	-0.76	1.31	1.59	0.81	1.01
-0.41	-0.42	-0.32	-0.52	0.35	0.45	0.13	0.40	0.81	0.70
-0.13	-0.55	-0.19	-0.30	-0.04	1.93	-0.38	0.08	0.21	0.64
-0.39	-0.88	-0.43	-0.48	0.31	1.44	-0.53	-0.83	0.81	0.08
-0.51	-0.09	-0.15	-0.61	0.15	-0.83	-0.29	0.62	0.81	0.65
-0.37	-0.32	0.11	-0.70	0.50	1.34	-0.55	1.01	0.81	1.05
1.18	-0.32	-0.37	-0.62	3.83	1.54	0.08	1.36	-0.01	0.39
-0.53	0.68	1.52	0.50	0.80	1.48	1.14	-0.58	0.81	0.99
-0.25	-0.22	-0.38	-0.71	-0.21	1.16	-0.03	0.88	0.81	0.80
-0.14	-0.56	-0.43	-0.56	0.07	1.80	-0.31	0.10	0.81	0.60
-0.29	4.23	2.39	2.15	2.94	1.91	2.37	2.28	0.81	0.02
1.33	-0.60	-0.65	-0.72	-0.12	1.44	-0.10	0.44	0.13	0.14
0.07	0.00	0.07	-0.12	2.20	0.70	-0.38	0.58	-0.22	0.17
-0.51	0.34	0.68	2.43	-0.03	-0.10	0.63	0.92	0.81	1.25
-0.52	-0.26	0.73	-0.24	0.23	-1.11	1.24	0.52	0.81	1.20
-0.34	-0.69	-0.52	-0.88	-1.00	-1.28	-0.53	-0.24	-1.05	-0.49
-0.48	-0.18	0.98	0.34	-0.16	1.60	0.73	0.85	0.81	1.02
-0.47	0.60	1.10	0.15	0.26	-1.27	0.03	-0.64	0.81	0.58
-0.36	-0.33	-0.31	-0.14	-0.08	-0.53	0.12	1.24	0.81	0.76
-0.40	1.51	1.37	1.05	1.70	0.31	1.38	1.89	0.81	0.98
-0.16	-0.07	0.38	0.58	-0.27	1.10	-0.14	-0.65	-0.02	0.25
-0.45	2.29	1.62	-0.16	2.61	1.86	1.74	1.73	0.81	0.63
-0.54	1.32	0.93	-0.35	1.87	-0.81	1.33	1.57	0.81	0.47
-0.26	-0.28	-0.08	0.23	0.60	0.86	0.36	-0.54	-0.57	0.76
0.34	0.08	0.41	0.19	1.26	1.54	0.32	0.45	0.81	0.40
1.07	-0.60	-0.71	0.30	-0.16	0.00	-0.54	-0.59	-0.72	-1.08
0.92	-0.53	-0.81	-0.72	-0.06	1.55	-0.08	0.18	-0.23	-0.13
-0.54	-0.30	-0.26	-0.05	-0.58	-0.41	-0.90	-0.52	0.81	-0.54
-0.46	0.27	0.96	1.01	1.49	0.01	1.29	0.11	0.81	1.11
-0.34	-0.30	-0.51	0.38	-0.21	0.09	-1.05	-0.78	-0.93	-0.77
-0.56	-0.61	-0.07	1.08	0.01	-1.02	0.10	0.51	0.81	1.16
-0.42	1.28	1.01	0.06	1.54	1.29	0.74	1.00	0.81	0.89
0.07	-0.34	-0.05	0.69	-0.08	0.82	-0.41	-0.47	0.12	0.03

-0.41	-0.40	-0.33	-0.37	-0.70	-0.99	1.10	-0.19	0.81	0.09
0.36	-0.46	-0.50	-0.91	-0.60	-0.79	0.08	-0.43	-1.06	-0.16
-0.51	1.93	1.40	-0.62	0.29	-0.53	1.69	1.93	0.81	0.00
-0.55	-0.04	-0.01	0.62	-0.30	-0.49	1.04	0.16	0.81	0.09
-0.51	-0.65	-0.49	-0.86	-0.68	-1.26	-0.47	-0.24	0.81	0.28
-0.53	0.18	0.90	-0.85	-0.11	-0.71	1.48	0.46	0.81	1.05
-0.34	0.03	0.62	0.28	-0.54	-0.72	0.18	0.33	0.81	1.01
1.78	-1.20	-1.71	-0.92	-0.50	0.35	-1.19	-0.67	-1.45	-1.66
-0.58	-0.26	-0.33	-0.66	-0.81	0.55	-2.20	-0.68	0.81	-0.05
-0.10	-1.12	-1.18	-1.12	-1.06	-1.18	-2.04	-1.45	-0.77	-2.04
-0.27	-0.57	-0.27	-0.51	-0.71	-0.93	-0.94	-1.05	-1.20	-0.23
8.08	-1.48	-2.29	-1.18	-1.07	-1.01	-1.88	-1.38	-2.04	-3.03
-0.47	-0.10	0.25	-0.26	-0.61	-1.14	1.01	-0.83	0.81	0.69
-0.54	1.72	1.02	0.76	0.01	-0.27	2.05	2.15	0.81	0.69
-0.31	-0.90	-0.64	0.12	-0.87	-0.58	-1.31	-1.02	-0.16	-0.89
-0.39	0.01	0.44	-0.10	0.17	-0.53	0.35	0.06	0.81	0.92
-0.49	1.29	1.16	0.06	3.14	1.13	0.38	-0.12	0.81	0.82
-0.46	0.38	0.31	0.96	0.12	0.64	1.26	0.74	-1.96	0.39
-0.39	-0.71	0.40	1.81	-0.11	-0.55	-0.09	-0.15	0.81	0.96
0.13	1.03	0.77	-0.31	2.89	0.26	0.92	1.58	0.33	0.61
-0.39	1.84	1.18	0.50	2.30	0.90	1.64	1.84	0.81	1.20
-0.39	1.00	0.66	1.41	3.63	1.61	0.79	1.49	0.81	1.08
0.21	-1.12	-1.42	-0.95	-0.61	0.50	-1.24	-1.34	-1.82	-1.86
-0.28	-0.72	-0.84	-0.83	-0.33	-1.13	-1.10	-0.46	-0.52	-1.16
-0.11	-1.36	-2.11	-1.23	-0.93	-1.02	-1.52	-1.55	-2.10	-3.23
0.22	-0.50	-0.54	0.15	-0.48	0.40	-0.74	-0.76	-0.41	-1.06
-0.19	-0.59	-0.70	-0.38	-0.76	-0.12	-1.95	-1.18	-1.55	-1.87
-0.39	1.03	1.02	3.04	0.08	-0.87	1.21	1.73	-0.84	0.29
-0.18	-0.08	0.56	0.05	1.02	0.67	1.13	0.21	0.81	0.72