

# UN ALGORITMO GENÉTICO PARA SELECCIÓN DE KERNEL EN ANÁLISIS DE COMPONENTES PRINCIPALES CON KERNELS

J. Aurora Montano Rivas<sup>1</sup> y Sergio F. Juárez Cerrillo

Facultad de Estadística e Informática, Universidad Veracruzana.

## ABSTRACT

Principal Component Analysis with Kernels (KPCA) is an extension of Principal Component Analysis (PCA) which is basically a PCA on the original data after they were sent, via a non-linear transformation, to a space called the feature space. The key for a successful KPCA is to extract directions of maximum variability in the transformed data and then identify these directions with patterns of maximum variability of the original data. However, there are situations for which KPCA is not sufficient to detect these directions of maximum variability. In this work we address this problem: we 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 explained variance by a KPCA. 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. The proposal is illustrated producing a ranking of the 210 municipalities in the State of Veracruz using 10 socioeconomic variables. The proportion of explained variance by the first component of a PCA is 56%. With our proposal, the first principal component of the ACPK extracts 99% of the variability in the feature space.

**KEYWORDS:** Evolutive Algorithms, Learning with Kernels, Marginality Index, Principal Component Analysis (PCA)

**MSC:** 62H25

## RESUMEN

El Análisis de Componentes Principales con Kernels (ACPK) es una extensión del Análisis de Componentes Principales (ACP) que consiste en enviar los datos mediante una transformación no lineal, a otro espacio, llamado el espacio de las características, 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 las características y luego identificar estas direcciones con las direcciones (no lineales) de variabilidad de los datos en el espacio original. Sin embargo, existen situaciones donde el ACPK no es suficiente para detectar estas direcciones no lineales de máxima variabilidad. En este trabajo construimos un espacio convexo de kernels formado por 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 los coeficientes de la combinación lineal convexa que determinan el kernel óptimo. El problema de optimización lo resolvemos con un algoritmo genético. La propuesta se ilustra con el ordenamiento que produce el ACPK con el primer componente principal de los 210 Municipios del Estado de Veracruz medidos en 10 indicadores socioeconómicos. La proporción de varianza explicada por el primer componente principal de un ACP es de 56% mientras que con nuestra propuesta, el primer componente principal del ACPK extrae 99% de la variabilidad en el espacio de las características.

## 1. INTRODUCCIÓN

Las técnicas de análisis multivariado funcionan adecuadamente bajo ciertas configuraciones geométricas de los datos. Por ejemplo, en discriminación es ideal que los grupos se puedan separar mediante hiperplanos. En el análisis de componentes principales (ACP) los datos se proyectan ortogonalmente sobre direcciones de máxima variabilidad que sean a la vez ortogonales entre ellas. En este caso el ACP funciona adecuadamente si los datos forman una nube de puntos con forma de hiperelipsoide y las direcciones de máxima variabilidad son los ejes principales del hiperelipsoide determinado por la forma cuadrática definida por la matriz de varianzas y covarianzas de los datos (Johnson *et al.*, 1999).

---

<sup>1</sup> julmontano@uv.mx, sejuarez@uv.mx

La utilidad de las técnicas del análisis multivariado es limitada sin la presencia de estos patrones. Para abordar situaciones sin la presencia de los patrones ideales se han propuesto diversas alternativas. Una relativamente reciente es la de Schölkopf, Smola y Müller (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 (Mika, 2002). La idea de estos algoritmos, aparentemente paradójica, es mapear a los datos mediante una función no lineal, hacia un espacio de mayor dimensión que la del espacio en el que se encuentran y llevar a cabo el análisis multivariado en los datos transformados, es decir en el espacio de mayor dimensión.

El dispositivo que permite realizar lo anterior de una forma analítica y numéricamente eficiente se conoce como el truco kernel (Schölkopf *et al.*, 2002). Este dispositivo permite realizar los algoritmos de análisis de patrones siempre que éstos se puedan expresar en términos de productos punto de los datos de entrenamiento. Así, si los datos en el espacio original no se pueden analizar satisfactoriamente con la técnica de análisis multivariado, se podrá optar por una versión kernel de la técnica para estudiar los patrones de los datos (Mika *et al.*, 2000).

En este artículo 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 artículo está estructurado de la siguiente forma. En la siguiente sección se presenta una revisión del ACP con kernels. En la sección 3 se presenta la propuesta de mejora del kernel para el ACPK y se describe el algoritmo genético que se usó como una herramienta de optimización. En la sección 4 se ilustra la propuesta para obtener un ordenamiento de los 210 municipios del Estado de Veracruz en base a 10 indicadores socioeconómicos. En la sección 5 se presentan las conclusiones y comentarios así como posibles investigaciones futuras.

## 2. ANALISIS DE COMPONENTES PRINCIPALES CON KERNELS

Supóngase que se tiene una muestra de entrenamiento de  $n$  vectores  $p \times 1$ ,  $\mathbf{x}_i \in \mathcal{V} \subset \mathbb{R}^p$  ( $i = 1, \dots, n$ ), tales que  $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$ , es decir, las observaciones están centradas. Consideremos la Descomposición en Valores Singulares (DVS) de la matriz de datos

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_n \end{bmatrix} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$$

donde  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p]$ ,  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$ ,  $\mathbf{U}'\mathbf{U} = \mathbf{I}$ ,  $\mathbf{V}'\mathbf{V} = \mathbf{I}$  y  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . De modo que los pares  $\lambda_i^2$ ,  $\mathbf{v}_i$  y  $\lambda_i^2$ ,  $\mathbf{u}_i$  son los eigenvalores y eigenvectores de las matrices de varianzas y covarianzas  $n\mathbf{S} = \mathbf{X}'\mathbf{X}$  y  $\mathbf{X}\mathbf{X}'$ , respectivamente. Formemos a las matrices  $\mathbf{U}_r$  y  $\mathbf{V}_r$  con las primeras  $r$  columnas de  $\mathbf{U}$  y  $\mathbf{V}$  respectivamente y sea  $\mathbf{\Lambda}_r^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_r)$ , de modo que  $\mathbf{V}_r = \mathbf{X}'\mathbf{U}_r\mathbf{\Lambda}_r^{-1}$  y

$$\mathbf{v}_j = \frac{1}{\lambda_j} \sum_{i=1}^n (\mathbf{u}_j)_i \mathbf{x}_i = \sum_{i=1}^n \alpha_{ij} \mathbf{x}_i \quad (1)$$

donde  $(\mathbf{u}_j)_i$  es el elemento  $i$  del vector  $\mathbf{u}_j$  y  $\alpha_{ij} = (\mathbf{u}_j)_i / \lambda_j$ . Ahora bien, supongamos que  $\mathbf{x}$  es un punto en el espacio de las observaciones  $\mathcal{V}$  el cual deseamos proyectar ortogonalmente a lo largo de  $\mathbf{v}_j$ . El componente de esta proyección es

$$P_{\mathbf{v}_j}(\mathbf{x}) = \mathbf{v}_j' \mathbf{x} = \left( \sum_{i=1}^n \alpha_{ij} \mathbf{x}_i \right)' \mathbf{x} = \sum_{i=1}^n \alpha_{ij} \mathbf{x}_i' \mathbf{x}. \quad (2)$$

Así vemos de (1) que los eigenvectores de la matriz de varianzas y covarianzas son combinaciones lineales de los datos de entrenamiento  $\mathbf{x}_i$  y los coeficientes  $\alpha_{ij}$  de estas combinaciones lineales se

calculan a partir de los eigenvalores y eigenvectores de la matriz  $\mathbf{X}\mathbf{X}'$ , cuyos elementos son los productos puntos  $\mathbf{x}_i'\mathbf{x}_j$ , ( $i, j = 1, \dots, n$ ). Además la proyección ortogonal de  $\mathbf{x}$  a lo largo de las direcciones de máxima variabilidad se calcula con los mismos coeficientes  $\alpha_{ij}$  y los productos punto  $\mathbf{x}_i'\mathbf{x}$ .

El ACP con kernels es un ACP de los datos transformados  $\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_n)$ , donde  $\Phi: \mathbb{R}^p \rightarrow \mathcal{H}$  es un mapeo no lineal a un espacio de Hilbert  $\mathcal{H}$  con producto interno  $\langle \cdot, \cdot \rangle$ . A  $\mathcal{H}$  se le llama el espacio de las características (Schölkopf *et al.*, 1999). En el ACPK se debe calcular el eigensistema de la matriz de varianzas y covarianzas de los datos transformados

$$\mathbf{S}_\Phi = \frac{1}{n} \sum_{i=1}^n \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)'$$

y luego proyectar ortogonalmente a las observaciones sobre los eigenvectores de  $\mathbf{S}_\Phi$ . Este procedimiento se puede realizar sin necesidad de conocer explícitamente al mapeo  $\Phi$  del siguiente modo. Sea  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  una función que evalúa el producto interno en  $\mathcal{H}$  de las imágenes bajo  $\Phi$  de observaciones en el espacio original  $\mathcal{V}$ , es decir,  $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$  para todo  $\mathbf{x}$  y  $\mathbf{y}$  en  $\mathcal{V}$ . Consideremos a la matriz

$$\mathbf{K} = \begin{bmatrix} \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_1) \rangle & \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2) \rangle & \dots & \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_n) \rangle \\ \langle \Phi(\mathbf{x}_2), \Phi(\mathbf{x}_1) \rangle & \langle \Phi(\mathbf{x}_2), \Phi(\mathbf{x}_2) \rangle & \dots & \langle \Phi(\mathbf{x}_2), \Phi(\mathbf{x}_n) \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \Phi(\mathbf{x}_n), \Phi(\mathbf{x}_1) \rangle & \langle \Phi(\mathbf{x}_n), \Phi(\mathbf{x}_2) \rangle & \dots & \langle \Phi(\mathbf{x}_n), \Phi(\mathbf{x}_n) \rangle \end{bmatrix}.$$

De modo que los elementos de  $\mathbf{K}$  están dados por  $k(\mathbf{x}_i, \mathbf{x}_j)$ , ( $i, j = 1, \dots, n$ ). La matriz  $\mathbf{K}$  se llama matriz Gram (Taylor *et al.*, 2004). Por otro lado, de (1) vemos que los eigenvectores de  $\mathbf{S}_\Phi$  están dados por  $\mathbf{v}_j = \sum_{i=1}^n \alpha_{ij} \mathbf{x}_i$  donde  $\alpha_{ij} = (\mathbf{u}_j)_i / \lambda_j$  donde ahora  $\mathbf{K}\mathbf{u}_j = \lambda_j^2 \mathbf{u}_j$ . De (2) vemos que las proyecciones ortogonales de las observaciones de entrenamiento en el espacio de las características  $\Phi(\mathbf{x}_i)$  a lo largo de los eigenvectores de  $\mathbf{S}_\Phi$  son

$$P_{\mathbf{v}_j}(\Phi(\mathbf{x}_l)) = \langle \mathbf{v}_j, \Phi(\mathbf{x}_l) \rangle = \sum_{i=1}^n \alpha_{ij} k(\mathbf{x}_i, \mathbf{x}_l), \quad l = 1, \dots, n.$$

**Tabla 1.** Algoritmo de ACPK.

- 
1. Entrada: Los datos  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{V} \subseteq \mathbb{R}^p$  y un kernel  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ .
  2. Se construye la matriz  $\mathbf{K}$  con elementos  $k(\mathbf{x}_i, \mathbf{x}_j)$ , ( $i, j = 1, \dots, n$ ).
  3. Se centra a la matriz  $\mathbf{K}$  calculando  $\mathbf{K}' = \mathbf{K} - \mathbf{1}_n \mathbf{K} - \mathbf{K} \mathbf{1}_n + \mathbf{1}_n \mathbf{K} \mathbf{1}_n$  donde  $\mathbf{1}_n$  es una matriz  $n \times n$  con entradas  $1/n$ .
  4. Se calcula el eigensistema de  $\mathbf{K}'$  dado por los pares  $(\lambda_j^2, \mathbf{u}_j)$ , ( $j = 1, \dots, n$ ).
  5. Se calculan los primeros  $r$  vectores  $\alpha_j = \mathbf{u}_j / \lambda_j$ , ( $j = 1, \dots, r$ ), con los  $r$  eigenvalores más grandes.
  6. Se calculan las componentes principales en el espacio de las características
$$z_{lj} = P_{\mathbf{v}_j}(\Phi(\mathbf{x}_l)) = \sum_{i=1}^n \alpha_{ij} k(\mathbf{x}_i, \mathbf{x}_l), \quad (j = 1, \dots, r, l = 1, \dots, n)$$
  7. Salida: La matriz  $n \times r$   $\mathbf{Z} = [z_{lj}]$  cuyas columnas son las primeras componentes principales en el espacio de las características.
- 

De manera similar, si  $\mathbf{x} \in \mathcal{V}$ , entonces  $P_{\mathbf{v}_j}(\Phi(\mathbf{x})) = \sum_{i=1}^n \alpha_{ij} k(\mathbf{x}_i, \mathbf{x})$  es la proyección ortogonal de  $\mathbf{x}$  a lo largo de  $\mathbf{v}_j$ . De esta forma vemos que para realizar el ACP en el espacio de las características es

suficiente con poder evaluar los productos internos entre las observaciones de entrenamiento en  $\mathcal{H}$ . La función  $k$  se le llama kernel y es la que permite evaluar estos productos internos.

Se deben hacer dos consideraciones. Primero, a partir de un mapeo  $\Phi$  no lineal del espacio original  $\mathcal{V}$  a un espacio  $\mathcal{H}$  con producto interno  $\langle \cdot, \cdot \rangle$  definamos  $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$ . Es fácil mostrar que esta función  $k$  es un kernel y por lo tanto podemos realizar el ACP en  $\mathcal{H}$  diagonalizando a la matriz  $\mathbf{K}$  en lugar de  $\mathbf{S}_\Phi$ . Segundo, supongamos que existe una función  $k: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  con la propiedad de que la matriz  $\mathbf{K}$ , cuyas entradas son  $k(\mathbf{x}_i, \mathbf{x}_j)$  para cualesquiera  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , es positiva definida, a esta función  $k$  se le conoce como kernel positivo definido. Se puede demostrar que existe un espacio  $\mathcal{H}$  con un producto interno  $\langle \cdot, \cdot \rangle$  y un mapeo  $\Phi$  tal que  $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$ . Es decir, dado un kernel positivo definido  $k$ , resulta que éste evalúa los productos internos de algún mapeo no lineal  $\Phi$  hacia algún espacio de producto interno  $\mathcal{H}$ .

El mecanismo de evaluar los productos internos en el espacio de las características mediante el kernel  $k$  se conoce como el truco kernel. Por ejemplo, para realizar el ACPK se tiene el siguiente algoritmo que usa el truco kernel. Es importante notar que en el paso 3 se centran las observaciones en el espacio de las características.

### 3. MEJORA DEL KERNEL PARA ACPK

Consideremos un conjunto fijo y arbitrario de  $\delta$  kernels  $k_i$  y definamos al conjunto de kernels (Kim *et al.*, 2006).

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

Proponemos 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, buscamos maximizar

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

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 y consideremos al conjunto 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 determinan a  $G^*$  a partir de los kernels  $k_i$ . La función objetivo queda dada por

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

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 de eigenvalores en un espacio convexo de matrices simétricas y positivas definidas. Existen muchos algoritmos numéricos de programación convexa semidefinida para resolver este problema de optimización. En este trabajo se propone un algoritmo genético para resolver la optimización (Montano, 2011).

**Tabla 2.** Población inicial de  $\theta_i, i = 1, \dots, M$ .

$\theta_1$	$\theta_{11}$	$\theta_{12}$	$\dots$	$\theta_{1\delta}$
$\theta_2$	$\theta_{21}$	$\theta_{22}$	$\dots$	$\theta_{2\delta}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
$\theta_M$	$\theta_{M,1}$	$\theta_{M,2}$	$\dots$	$\theta_{M,\delta}$

Se fijan  $\delta$  kernels  $k_i$  con los que se construye el espacio de matrices Gram  $\mathcal{G}$ . A continuación se generan de forma aleatoria los coeficientes  $\theta$ 's con las restricciones  $\sum_{i=1}^{\delta} \theta_i = 1$  y  $\theta_i \geq 0$ . Estos pasos se repiten  $M$  veces. La población inicial de kernels está dada por los valores en la Tabla 2.

En el siguiente paso se elige al renglón que explique mayor proporción de varianza. Para esto cada hilera se evalúa para obtener la nueva matriz Gram

$$G_i^* = \theta_{i1} G_1 + \theta_{i2} G_2 + \dots + \theta_{i\delta} G_\delta, \quad i = 1, \dots, M.$$

De estas matrices Gram se calculan sus eigenvalores para así conocer la proporción de varianza explicada por cada kernel. Hasta aquí se conoce la máxima varianza explicada por 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. El objetivo es crear nuevos kernels con al menos igual proporción de varianza explicada. Para esto se eligen las  $h$  matrices Gram con la mayor proporción de varianza explicada y se procede a aplicar el operador cruce aritmético del siguiente modo. Para cada uno de los  $\binom{h}{2}$  pares de vectores de ponderaciones se construye un nuevo vector de ponderaciones dado por

$$T_\ell = (r_\ell \theta_{i1} + (1 - r_\ell) \theta_{j1}, r_\ell \theta_{i2} + (1 - r_\ell) \theta_{j2}, \dots, r_\ell \theta_{i\delta} + (1 - r_\ell) \theta_{j\delta}),$$

donde  $t_{\ell 1} = r_\ell \theta_{i1} + (1 - r_\ell) \theta_{j1}$ ,  $t_{\ell 2} = r_\ell \theta_{i2} + (1 - r_\ell) \theta_{j2}, \dots, t_{\ell \delta} = r_\ell \theta_{i\delta} + (1 - r_\ell) \theta_{j\delta}$ ,  $i, j = 1, \dots, h$ ,  $i \neq j$ ,  $\ell = 1, \dots, \binom{h}{2}$ . De esta manera se obtienen  $\binom{h}{2}$  nuevos vectores de ponderaciones que producen  $\binom{h}{2}$  matrices nuevas de Gram

$$G_\ell = t_{\ell 1} G_1 + t_{\ell 2} G_2 + \dots + t_{\ell \delta} G_\delta, \ell = 1, \dots, \binom{h}{2}.$$

En un algoritmo genético se procedería con un operador mutación. Existen varios operadores de 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  $\mathcal{G}$ . Al aplicar la mutación por intercambio, 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.

#### 4. APLICACIÓN

Para ejemplificar la propuesta se hace uso de los 10 indicadores socioeconómicos de los 210 Municipios del Estado de Veracruz que se utilizan para calcular el índice de marginación en el año 2005 (Tabla 3). Estos indicadores se construyen con la información del II Censo de Población y Vivienda 2005 (INEGI) y la Encuesta Nacional de Ocupación y Empleo (ENOE) 2005. El Consejo Nacional de Población (CONAPO) define al Índice de Marginación como el primer componente principal de los datos  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{210}$  en  $\mathbb{R}^{10}$ .

**Tabla 3.** Indicadores usados en el cálculo de Índice de Marginación.

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

La Tabla 4 muestra las proporciones de varianza del ACP aplicado a los datos estandarizados. Vemos que el Índice de Marginación se calcula en base a una explicación del 56% de la varianza total de los datos.

**Tabla 4.** Porcentajes de varianza explicada del ACP a los datos.

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

La Tabla 5 muestra los diez valores de los  $\sigma$ 's de la combinación lineal convexa del kernel así como los diez valores óptimos de los  $\theta$ 's que se obtuvieron al aplicar el algoritmo genético. Se usaron 10 kernels Gaussianos

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

Con este kernel, el primer componente principal explica el 99% de la variabilidad total en el espacio de las características.

**Tabla 5.** Valores de los  $\sigma$ 's y de los  $\theta$ 's obtenidos con el algoritmo genético.

$j$	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	91.418	92.649
$\theta$	0.00480	0.0123	0.02418	0.0684	0.0880	0.1076	0.11329	0.1646	0.2030	0.2136
	1	4	7	1	3	4	7		4	1

ACPK permite también extraer patrones de máxima variabilidad mediante las proyecciones ortogonales sobre los primeros eigenvectores principales de las imágenes  $\Phi(\mathbf{x})$  de observaciones  $\mathbf{x}$  en el espacio original. 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

$$v_l = \sum_{i=1}^{210} \alpha_{i1} k(\mathbf{x}_i, \mathbf{x}_l) = \sum_{i=1}^{210} \alpha_{i1} \left[ \sum_{j=1}^{10} \theta_j \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_l\|^2}{2\sigma_j^2}\right) \right], \quad l = 1, \dots, 210,$$

y  $\alpha_1 = (\alpha_{11}, \alpha_{21}, \dots, \alpha_{210,1})$  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 6 en el Apéndice. Se tiene que entre los municipios con los valores más pequeños están Veracruz, Xalapa, Coahuila, 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, Tlalixcoyan, 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.

## 5. CONCLUSIÓN Y RECOMENDACIONES.

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.

En el ejemplo para calcular el Índice de Marginación con la metodología del CONAPO se requieren los 6 primeros componentes para explicar el 93% de la varianza total. Con la metodología propuesta, se obtiene un porcentaje de varianza explicado del 99% con los 2 primeros componentes. El algoritmo genético proporciona 4 kernels óptimos, uno de los cuales se identifica desde la segunda iteración. Cabe señalar que los kernels identificados como óptimos explican con el primer componente el 99.93% de la varianza total. Este hecho se debe considerar en el cálculo de un nuevo índice de marginación.

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.
- 2) La selección de kernel se puede extender a espacios de kernels más complejos que el que se usa en este trabajo.
- 3) Ya se ha iniciado investigación para entender 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 patrones 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 otras técnicas multivariadas está abierto, por ejemplo en correlación canónica. Una posibilidad para abordarlo es seguir una estrategia análoga a la que se propone en este artículo.

RECEIVED, OCTOBER 2013  
REVISED FEBRUARY, 2014

## REFERENCIAS

- [1] JOHNSON, R., and WICHERN, D. (1999): **Applied Multivariate Statistical Analysis**. Prentice Hall, Fourth Edition. New Jersey.
- [2] KIM, S., MAGNANI, A. and BOYDD, S. (2006): **Optimal Kernel Selection in Kernel Fisher Discriminant Analysis**. Department of Electrical Engineering. Stanford University. Stanford, C.A. USA.
- [3] MIKA, S., RÄTSCH, G., WESTON, J., SCHÖLKOPF, B., and MÜLLER, K. (2000): **Fisher Discriminant Analysis with Kernels**. Springer, Berlin.
- [4] MIKA, S. (2002): **Kernel Fisher Discriminants**. Von der Fakultät IV – Elektrotechnik und Informatik der Technischen Universität Berlin. Berlin, Germany.

- [5] MONTANO, J.A. Y CANTÚ, M. (2011): **Algoritmos Genéticos en la Discriminación. Una Solución Óptima para Clasificación.** Editorial Académica Española, Madrid.
- [6] SCHÖLKOPF, B., and SMOLA, A. (2002): **Learning with Kernels.** Cambridge. The MIT Press Cambridge.
- [7] SCHÖLKOPF, B., SMOLA, A. and MÜLLER, K. (1998): Nonlinear component analysis as a kernel eigenvalue problem. **Neural Computation** 10,1299-1319
- [8] 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.
- [9] TAYLOR, J., and CRISTIANINI, N. (2004): **Kernel Methods for Pattern Analysis.** New York. Cambridge.

## APENDICE

**Tabla 8.** Ordenamiento de los municipios de Veracruz obtenido con ACPK.

<b>Municipio</b>	<b>Municipio</b>	<b>Municipio</b>
Veracruz	Comapa	Tempoal
Xalapa	Chiconamel	Espinal
Mixtla de Altamirano	Texhuacán	Landero y Coss
Tehuipango	Pueblo Viejo	Huayacocotla
Coatzacoalcos	Minas, Las	Actopan
Filomeno Mata	Carlos A. Carrillo	Tonayán
Soteapan	Papantla	Ixhuatlán del Sur
Boca Del Río	Chumatlán	Pánuco
Orizaba	Naranjos Amatlán	Cazones
Poza Rica	Fortín	Zentla
Nanchital	Magdalena	Coyutla
Texcatepec	Calcahualco	Chiconquiaco
Mecayapan	Cosoleacaque	Perote
Zozocolco	Coatepec	Rafael Delgado
Río Blanco	Aquila	San Andrés Tux.
Carrillo Puerto	Tenampa	Temapache
Mecatlán	Nogales	Apazapan
Córdoba	Coahuilán	Emiliano Zapata
Tantoyuca	Camarón d Tejada	Cuitláhuac
Ilamatlán	Tatatila	Tlaltetela
Soledad Atzompa	Tlachichilco	Chicontepec
Zontecomatlán	Cosamaloapan	Tantima
Astacinga	Martínez d la Torre	Alto Lucero
Tatahuicapan	Zaragoza	Tenochtitlán
Atlahuilco	Ixhuatlán deMadero	Acula
Lerdo de Tejada	Alpatláhuac	Higo, El
Reyes, Los	Coetzala	Jalcomulco
Minatitlán	Amatitlán	Chontla
Zacualpan	Acatlán	Jilotepec
Ursulo Galván	Huiloapan	Ixtaczoquitlán
Pajapan	Tuxtilla	Teocelo
Banderilla	Ixhuacán d los Reyes	Tezonapa
Cerro Azul	Coatzintla	Miahuatlán
Coxquihui	Choapas, Las	Naranjal
Ayahualulco	Jáltipan	Tecolutla
Túxpam	Tlacolulan	Tihuatlán
Camerino Z. Mendoz	Alvarado	Tampico Alto
Antigua, La	Puente Nacional	Yanga



Perla, La	Ixcatepec	Xico
Tequila	Rafael Lucio	Tlacotepec
Xoxocotla	Oluta	Atzalan
Oteapan	Chalma	Tepetlán
Zongolica	Cotaxtla	Vega d Alatorre
San Andrés Te.	Hidalgotitlán	Naolinco
Tlaquilpa	Tepatlaxco	Tamiahua
Agua Dulce	Ozuluama	Villa Aldama

Continuación de la Tabla 8.

<b>Municipio</b>	<b>Municipio</b>	<b>Municipio</b>
Cosautlán de Carv.	Tlapacoyan	Tlilapan
Chinampa	Atzacan	Ixhuatlán Del Café
Paso de Ovejas	Nautla	Tlacojalpan
Coscomatepec	Tepetzintla	Huatusco
Saltabarranca	Las Vigas d Ramírez	Juchique De Ferrer
Benito Juárez	Manlio Fabio Altam.	Chinameca
Medellín	Chacaltianguis	San Juan Evangelist
Tierra Blanca	Maltrata	Santiago Tuxtla
Tomatlán	Coacoatzintla	Ixhuatlancillo
Otatitlán	Acajete	José Azueta
Colipa	Amatlán de los Reyes	Angel R. Cabada
Castillo de Teayo	Cuichapa	Catemaco
Citlaltépetl	Tres Valles	Tamalín
Sochiapa	Playa Vicente	Isla
Uxpanapa	Paso Del Macho	Soledad De Doblado
Ignacio de la Llave	Altotonga	Jesús Carranza
Tlalnelhuayocan	Chocamán	Soconusco
Atoyac	Jalacingo	Yecuatla
Acayucan	Platón Sánchez	Acultzingo
Mariano Escobedo	Moloacán	Misantla
Gutiérrez Zamora	Tancoco	Tlalixcoyan
Ixmatlahuacan	Texistepec	Juan Rodríguez Clara
Tlacotalpan	Totutla	Sayula De Alemán
		Hueyapan de
Omealca	Jamapa	Ocampo