EST-46114 Métodos Multivariados y Datos Categóricos

Sesion 07 - Analisis de Componentes Principales - Parte 3/3

Juan Carlos Martinez-Ovando

Maestria en Ciencia de Datos

Objetivos

- Estudiaremos PCA aplicado a trasformaciones funcionales de los datos originales (usualmente conocidas como trasnformaciones por kernels).
- Exploraremos un procedimiento para corregir por *raleza* (*sparcity*) en PCA.

PCA | Preliminar 1/

Recordemos que PCA puede verse como un procedimiento de ortogonalizacion de una matriz de datos $X_{(n\times p)}$, con $n >> p$, basada en la descomposicion en valores singulares,

$$X_{(n\times p)} = U_{(n\times p)} D_{(p\times p)} V_{(p\times p)}.$$

A partir de esta descomposicion, podemos calcular las siguientes matrices cuadraticas

$$\begin{aligned} S_{(p\times p)} &= X'_{(n\times p)} X_{(n\times p)} = V'_{(p\times p)} D^2_{(p\times p)} V_{(p\times p)}, \\ K_{(n\times n)} &= X_{(n\times p)} X'_{(n\times p)} = U_{(n\times p)} D^2_{(p\times p)} U'_{(p\times n)}. \end{aligned}$$

PCA | Preliminar 2/

- La matrix S corresponde a la suma de cuadrados de X –cuando los datos han sido estandarizados previamente–,
- La matriz K es referida como la *matriz de Gram*.

PCA | Preliminar 3/

Recordemos que el primer SPC esta dado por la siguiente transformacion

$$
\begin{aligned}
c_1 &= X v_1 \\
&= U D V' v_1 \\
&= u_1 d_1,
\end{aligned}
$$

donde v_1 es un vector de dimension $(p \times 1)$ correspondiente al eigenvector asociado con el primer eigenvalor de X.
Sucesivamente, los demas componentes principales $(c_2, \ldots, c_p)$ se obtienen con base en la misma proyeccion anadiendo las restricciones ortogonales anidadas correspondientes con los c_js previos.

PCA | Preliminar 4/

Retomemos, el primer componente principal, bajo SPCA, puede obtenerse de tres formas alternaivas:

a. Como el producto de X con el primer eigenvector de S
b. A partir de la descomposicion en valores singulares de X (descrito lineas arriba)
c. A traves de la descomposicion singular de K.

PCA | Preliminar 5/

Asi, pues no se necesita saber X directamente, sino que basta con conocer S o K para producir los componentes principales de un conjunto de datos.

–Este resultado fue invocado la sesion anterior para realizar PCA Inferencial–.

PCA | Preliminar 6/

En particular, el primer componente principal de un vector p-dimensional $\boldsymbol{x}_{i\cdot}$ puede obtenerse como la proyeccion sobre el eje del primer componente, i.e.

$$c_{i1} = v_1' \boldsymbol{x}_{i\cdot}.$$

la expresion anterior puede calcularse directamente, o *indirectamente* empleando la expresion alternativa

$$c_{i1} = u_1' X \boldsymbol{x}_{i\cdot}/d_1 = \sum_{i=1}^{n} \left(\frac{u_{i1}}{d_1} \right) y_i' X.$$

Expresiones semjantes se obtienen de manera analoga para los demas componentes principales (no solo el primero).
De esta forma podemos ver que es necesario conocer los productos interiores $(\boldsymbol{x}_{i\cdot}' X)_{i=1}^{n}$ solamente.

PCA | Preliminar 7/

Como antes mencionamos, el calculo de $c_{.1}$ puede obtenerse de dos formas:

Forma.1

Calcular $S = X'X$, obteniendo el primer `eigenvector` de esta matriz, v_1 de S, y calcular

$$c_1 = Xv_1.$$

Forma.2

Empleando la *matriz de Gram*, calculando XX', obteniendo el primer `eigenvector` de esta matriz, u_1 y su correspondiente `eigenvalor`, d_1, y calculando

$$c_1 = u_1 d_1.$$

> *La **Forma.1** es particularmente util cuando $n >> p$, mientras que la **Forma.2** lo es para el caso $n << p$.*

PCA | Funcional 1/

Al final del dia, SPCA descansa en el calculo de los prodcutos interiores

$$(\boldsymbol{x}_i'\boldsymbol{X})_{i=1}^{n},$$

el cual puede interpretarse como una 'medida de similaridad euclidiana' entre objetos p-dimensionales.

> *La idea entonces de **PCA Funcional** (o **Kernel PCA**) es la de relajar el supuesto de `similaridad euclidiano` para otras medidas de similaridad.*

Esto en particular cuando los objetos/renglones de X residan en *sub-espacios no lineales* de $\mathbb{R}^p$ (curvas, superficies o *manifolds*).

PCA | Funcional 2/

¿Como son los sub-espacios no lienales?

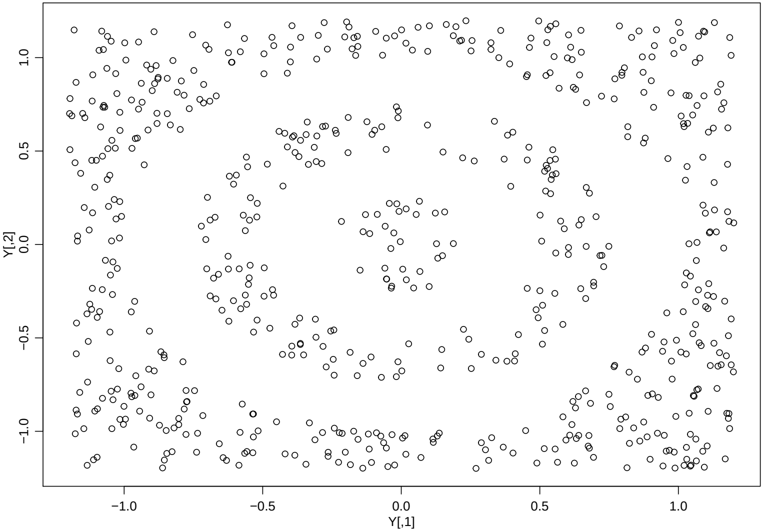
PCA | Funcional 3/

Veamos un ejemplo con el siguiente diagrama de datos sinteticos en R.

```
set.seed(1)
n <- 1000
Y <- matrix(runif(n*2,-1.2,1.2),n,2)
r <- sqrt(apply(Y^2,1,sum))
Y <- Y[  r<.25 | (r>.5 & r<.75) | r>1   ,]
r <- sqrt(apply(Y^2,1,sum))

r <- sqrt(apply(Y^2,1,sum))
clr <- rgb( (r/max(r))^.7,(1-r/max(r))^.7,.5)
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
#plot(Y,col=clr)
plot(Y)
```

PCA | Funcional 4/



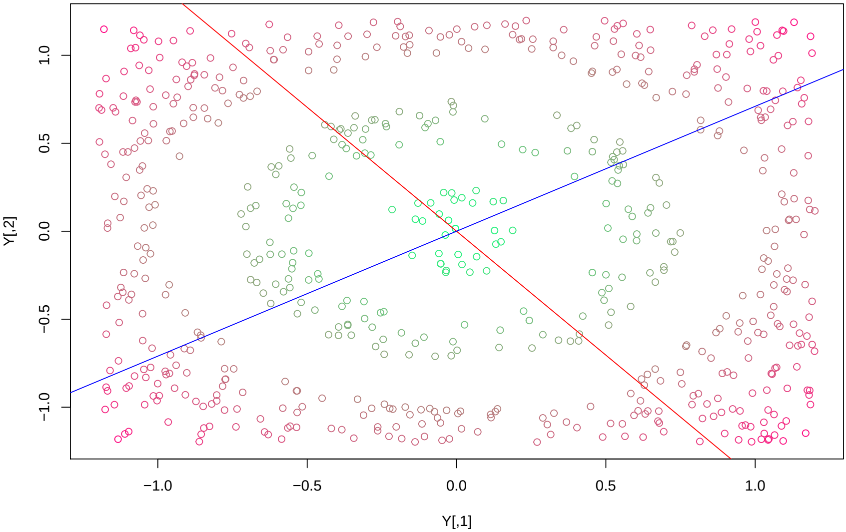
PCA | Funcional 5/

En este caso, la *matriz de Gram* deseable sera tal que mida alternativamente la similaridad entre los objetos/renglones $\boldsymbol{x}_{i\cdot}$s de la matriz $\boldsymbol{X}$.

PCA | Funcional 6/

En caso de realizar PCA convencional en este conjunto de datos, se obtendrian resultados confusos, pues al parecer no habria ortogonalizacion que realizar. **Al menos, ortogonalizacion lienal**. Veamos los siguientes resultados:

PCA | Funcional 7/

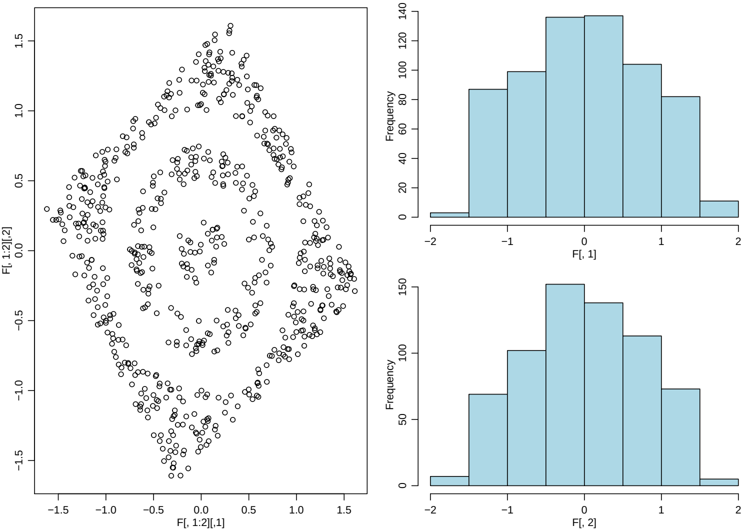


PCA | Funcional 8/

- En el grafico anterior, las rectas representan los *dos ejes principales del SPCA*.
- Como observamos, los datos de *componentes principales* son iguales a X.

Esto es porque la *medida de similaridad* empleada es la euclidiana. El resultado es, en este caso, una rotacion de la matriz X solamente.

PCA | Funcional 9/



PCA | Funcional 10/

El resultado anterior se obtuvo adoptando la matriz

$$K = XX',$$

como la matriz de Gram basada en la nocion de similaridad euclidiana (i.e. similaridad entre datos/renglones $\boldsymbol{X}_{i\cdot}$*).*

PCA | Funcional 11/

Ahora, si *modificamos la no nocion de similaridad* por la siguiente metrica,

$$d(\boldsymbol{x}_{i\cdot}, \boldsymbol{x}_{j\cdot}) = \left(\boldsymbol{x}_{i\cdot}' \boldsymbol{x}_{i\cdot} + 1\right)^2 ,$$

la *matriz de Gram* asociada seria

$$\tilde{K} = \left(XX' + 1\right)^2 .$$

Noten que las modificaciones de similaridad pueden ser distintas.

En este caso, realizando la descomposicion singular de $\tilde{K}$ resultaria en el siguiente SPCA.

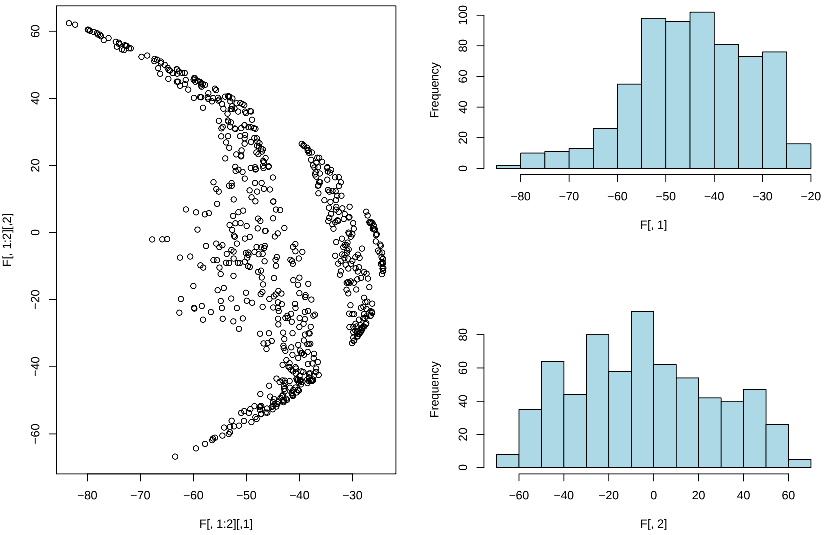
```
K <- (tcrossprod(Y) + 1)^2

sK <- svd(K)
F <- sK$u%*%diag(sK$d)

layout(matrix(c(1,1,2,3),2,2))
plot(F[,1:2])

hist(F[,1],main="",col="lightblue")
hist(F[,2],main="",col="lightblue")
```

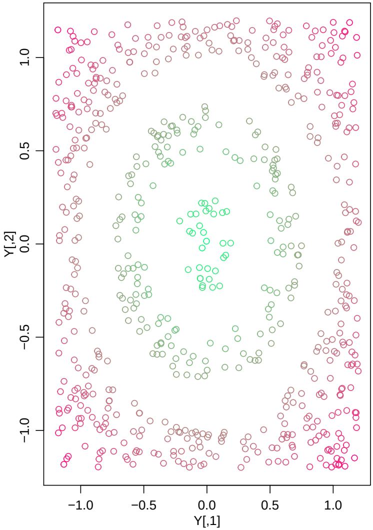
PCA | Funcional 13/

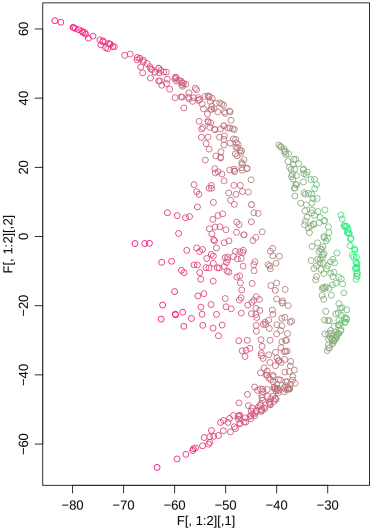


PCA | Funcional 14/

- Como es evidente, los datos tranformados (panel izquierdo en la grafica), responden la evidente separacion de anillos que visualiamos en los datos originales.
- Las proyecciones de los datos en los *nuevos ejes principales* es tambien distinta, mostrando sesgo en este caso.
- La asociacion entre las regiones de los datos originales, X, y los datos transformados, F, se muestra en el siguiente grafico.

PCA | Funcional 15/





PCA | Funcional 16/

Al modificar la nocion de similaridad como antes, se introduce la nocion de *kernel*.

- En un contexto general, el *kernel* servira como la medida de similaridad.

Asi, entre dos puntos y_i y y_j la similaridad de objetos/renglones con base en el *kernel* se definira como

$$k(\boldsymbol{x}_{i\cdot}, \boldsymbol{x}_{j\cdot}) = \phi(\boldsymbol{x}_{i\cdot})'\phi(\boldsymbol{x}_{j\cdot}),$$

donde

$$\phi : \mathbb{R}^p \to \mathbb{R}^q,$$

donde tipicamente $q >> p$.
El *kernel*, asi definido, puede interpretarse como el **producto interior no euclidiano** no de los datos originales $\boldsymbol{x}_{i\cdot}$s, sino de los datos modificados por una cierta funcion $\phi(\cdot)$.

> *La eleccion de la funcion $\phi(\cdot)$ es generalemente caso particular. Revisen las lecturas complementarias para darse una idea general de las aplicaciones posibles.*

PCA | Funcional 17/

En el ejemplo anterior, la funcion $\phi(\cdot)$ empleada es:

$$\phi(\boldsymbol{x}_{i\cdot}) = \left(1, \sqrt(2)x_{i1}, \ldots, \sqrt(2)x_{ip}, x_{i1}^2, \ldots, x_{ip}^2, \sqrt{2}x_{i1}x_{i2}, \ldots, \sqrt{2}x_{i(p-1)}x_{ip}\right),$$

para $i = 1, 2, \ldots, p$.

PCA | Funcional 18/

¿En que contextos **PCA Funcional** es util?

- Procesamiento de textos
- Procesamiento de imagenes
- Procesamiento de audio

Ejercicio analitico

- Verifiquen que $d(\boldsymbol{x}_{i\cdot}, \boldsymbol{x}_{j\cdot})$ visto antes es igual a $\phi(\boldsymbol{x}_{i\cdot})'\phi(\boldsymbol{x}_{j\cdot})$ para todo $\boldsymbol{x}_{i\cdot}$ y $\boldsymbol{x}_{j\cdot}$ en $\mathbb{R}^p$.

Observaciones

- En R se pueden encontrar varias implementaciones confiables en el paquete `kernlab`.
- Casi todos los procedimientoes vistos en este curso donde se emplea el producto interior euclidiano pueden modificarse para definirse en terminos de *kernels*.

  Regresion es un ejemplo, derivando en modelos de expansiones de bases de kernels. Eso lo veremos en la siguiente semana.

Lecturas complementarias

- Wang (2014) "Kernel Principal Component Analysis and its Applications in Face Recognition and Shape Models." `est46114_s07_suplemento1.pdf`
- Erichson et al (2018) "Sparse Principal Component Analysis via Variable Projection." `est46114_s07_suplemento2.pdf` y en esta liga.

Table of Contents