

Universidad de Salamanca
Grado en Matemáticas

REVISIÓN DE MÉTODOS
MULTIVARIANTES
SUPERVISADOS Y NO SUPERVISADOS

Trabajo Fin de Grado



VNiVERSiDAD
D SALAMANCA

Alumno: Pedro Ángel Fraile Manzano
Tutoras: Ana Belén Nieto Librero y Nerea González García

Salamanca, Julio de 2023

Índice general

1	Métodos no supervisados	1
1.1	Introducción	1
1.2	Análisis de Componentes Principales	1
1.2.1	Definición y cálculo de las Componentes	1
1.3	PCA en matrices de datos	4
	Bibliografía	5

Capítulo 1

Métodos no supervisados

1.1. Introducción

1.2. Análisis de Componentes Principales

El análisis de componentes principales fue en primera instancia desarrollado a principios del siglo XX por el estadístico Pearson (1901). Fue una de las primeras técnicas de análisis multivariante. Como muchos otros métodos de este tipo no tuvo un verdadero desarrollo y expansión hasta que la capacidad de computación fue suficiente para manejar cantidades de datos considerables.

1.2.1. Definición y cálculo de las Componentes

Sea un vector aleatorio $\mathbf{x}^T = [X_1, \dots, X_p]$ con vector de medias μ y matriz de covarianzas Σ .

Definición 1.2.1. Las componentes principales son combinaciones lineales de las variables $X_1 \dots X_p$

$$\mathbf{z}_j = a_{1j}X_1 + \dots + a_{pj}X_p = \mathbf{a}_j^T \mathbf{x} \quad (1.1)$$

Donde \mathbf{a}_j es un vector de constantes y la variable \mathbf{z}_j cumple lo siguiente:

- Si $j = 1$ $Var(\mathbf{z}_1)$ es máxima restringido a $\mathbf{a}_1^T \mathbf{a}_1 = 1$
- Si $j > 1$ debe cumplir:
 - $Cov(\mathbf{z}_j, \mathbf{z}_i) = 0 \quad \forall i \neq j$
 - $\mathbf{a}_j^T \mathbf{a}_j = 1$
 - $Var(\mathbf{z}_j)$ es máxima.

De esta manera lo que se busca es una nueva base que reúna las direcciones de máxima variación

El cálculo de la primera componente principal se lleva a cabo con un proceso de optimización de la función $Var(\mathbf{z}_1)$ sujeto a la restricción de que $\mathbf{a}_1^T \mathbf{a}_1 = 1$.

Aplicando el método de los multiplicadores de Lagrange, dada una función $f(\mathbf{x}) = f(x_1, \dots, x_p)$ diferenciable con una restricción $g(\mathbf{x}) = g(x_1, \dots, x_p) = c$, existe una constante λ de manera que la ecuación:

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0 \quad i = 1, \dots, p \quad (1.2)$$

Tiene como solución los puntos estacionarios de $f(\mathbf{x})$. Además, si se define la función $L(\mathbf{x}) = f(\mathbf{x}) - \lambda[g(\mathbf{x}) - c]$ es posible simplificar la expresión anterior a:

$$\frac{\partial L}{\partial \mathbf{x}} = 0 \quad (1.3)$$

Para el caso de las componentes principales, la función objetivo es la varianza de la combinación lineal, es decir, $f(\mathbf{x}) = \mathbf{x}^T \Sigma \mathbf{x}$ y la restricción aplicada es $g(\mathbf{x}) = \mathbf{x}^T \mathbf{x} = 1$.

Tomando $\mathbf{x} = \mathbf{a}_1$ se puede establecer $L(\mathbf{a}_1) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 - \lambda[\mathbf{a}_1^T \mathbf{a}_1 - 1]$. Que al derivarla se obtiene:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_1} &= 2\Sigma \mathbf{a}_1 - 2\lambda \mathbf{a}_1 \\ &= 2(\Sigma - \lambda) \mathbf{a}_1 \end{aligned}$$

Igualando a 0 tenemos la siguiente ecuación:

$$(\Sigma - \lambda I) \mathbf{a}_1 = 0 \quad (1.4)$$

Para que \mathbf{a}_1 sea un vector no trivial, tenemos que elegir λ de tal manera que $|\Sigma - \lambda I| = 0$, es decir, λ es un vector propio de la matriz de covarianzas, Σ . Al ser ésta una matriz semidefinido positiva y simétrica, los valores propios son reales y positivos. Por tanto, \mathbf{a}_1 es un vector propio de la matriz de covarianza.

La función a maximizar es $Var(\mathbf{z}_1) = Var(\mathbf{a}_1^T \mathbf{x}) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 = \mathbf{a}_1^T \lambda \mathbf{a}_1$, y para maximizarla basta tomar $\lambda = \max\{\lambda_1 \dots \lambda_p\}$. reordenando si es necesario, se tiene que $\lambda = \lambda_1$

Una vez calculada la primera componente principal \mathbf{z}_1 , la segunda componente se calcula de manera análoga, maximizando $Var(\mathbf{z}_2) = Var(\mathbf{a}_2^T \mathbf{x})$ condicionada por $\mathbf{a}_2^T \mathbf{a}_2 = 1$. A esta restricción tenemos que añadir la restricción $Cov(\mathbf{z}_1, \mathbf{z}_2) = 0$

Proposición 1.2.1. La condición $Cov(\mathbf{z}_1, \mathbf{z}_2) = 0$ equivale a la condición $\mathbf{a}_2^T \mathbf{a}_1 = 0$.

Demostración. Utilizando que $\mathbf{z}_j = \mathbf{a}_j^T \mathbf{x} \quad \forall j$, tenemos entonces que:

$$\begin{aligned} Cov(\mathbf{z}_2, \mathbf{z}_1) &= Cov(\mathbf{a}_2^T \mathbf{x}, \mathbf{a}_1^T \mathbf{x}) \\ &= \mathbb{E}(\mathbf{a}_2^T (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \mathbf{a}_1) \\ &= \mathbf{a}_2^T \mathbb{E}((\mathbf{x} - \mu)(\mathbf{x} - \mu)^T) \mathbf{a}_1 \\ &= \mathbf{a}_2^T \Sigma \mathbf{a}_1 \\ &= \mathbf{a}_2^T \lambda_1 \mathbf{a}_1 \end{aligned}$$

De manera que, si $\mathbf{a}_2^T \lambda_1 \mathbf{a}_1 = 0 \Rightarrow \mathbf{a}_2^T \mathbf{a}_1 = 0$, luego son vectores ortogonales entre sí. \square

Observación: Esta proposición se puede extender de manera simple al caso de tener que calcular la i -ésima componente principal habiendo calculado las anteriores de las cuales se sepan los valores propios asociados.

Corolario 1.2.1. Las componentes principales son todas ortogonales entre sí.

Para $k = 2$, se dan dos restricciones, $\mathbf{a}_2^T \mathbf{a}_2 = 1$ y además $\mathbf{a}_1^T \mathbf{a}_2 = 0$. Para este caso existen λ, ϕ de manera que la función a maximizar es:

$$L(\mathbf{a}_2) = \mathbf{a}_2^T \Sigma \mathbf{a}_2 - \lambda[\mathbf{a}_2^T \mathbf{a}_2 - 1] - \phi(\mathbf{a}_1^T \mathbf{a}_2) \quad (1.5)$$

Que al ser derivado respecto \mathbf{a}_2 obtenemos:

$$2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \phi \mathbf{a}_1 = 0 \quad (1.6)$$

Que al multiplicar todo por \mathbf{a}_1^T obtenemos que $\phi = 0$. De esta manera, se obtiene en la ecuación lo mismo que en el cálculo de la primera.

Por tanto, $\lambda = \lambda_2$ que es el segundo valor propio más grande, y \mathbf{a}_2 es el vector propio de valor propio λ_2 .

Un proceso similar se puede seguir para calcular el resto de componentes principales.

Por tanto, se obtiene que las componentes principales viene dado por los vectores propios de la matriz de covarianzas Σ . Además sabemos que $Var(\mathbf{a}_k^T \mathbf{x}) = \lambda_k$ donde λ_k es el k -ésimo valor propio más grande.

Sea ahora la matriz \mathbf{A} cuyas columnas son los \mathbf{a}_k . Entonces el vector \mathbf{z} que contiene a las componentes principales viene dado por la transformación:

$$\mathbf{z} = \mathbf{A}^T \mathbf{x} \quad (1.7)$$

Se deduce rápidamente que la matriz \mathbf{A} es ortonormal, de manera que $\mathbf{A}^T \mathbf{A} = \mathbf{I}$

Una vez descompuesta de esta manera la matriz de covarianzas tenemos que $\Sigma = \mathbf{A}^T \Delta \mathbf{A}$

Proposición 1.2.2. *Demostración.* □

1.3. PCA en matrices de datos

Sea \mathbf{x} el vector aleatorio de longitud p , tomamos n observaciones de ese vector y obtenemos $\mathbf{x}_1 \dots \mathbf{x}_n$. Esta recopilación de observaciones nos permite construir la matriz de datos \mathbf{X} cuyas filas son cada una de las observaciones. Esta matriz \mathbf{X} es de tamaño $n \times p$. Para este caso, las componentes principales se definen de manera análoga:

Definición 1.3.1. Dado un vector aleatorio de longitud p del cual hemos extraído n observaciones podemos definir las componentes principales como:

$$\tilde{z}_{ij} = \mathbf{a}_j^T \mathbf{x}_i \quad (1.8)$$

Donde \mathbf{a}_j es un vector de constantes de longitud p que cumple lo siguiente:

- Si $j = 1$, entonces \mathbf{a}_1 maximiza la varianza de la muestra, es decir maximiza $\frac{1}{n-1} \sum_{i=1}^n (\tilde{z}_{i1} - \bar{z}_1)^2$. Además debe cumplir que $\mathbf{a}_1^T \mathbf{a}_1 = 1$
- Si $j > 1$ debe cumplir:
 - Los vectores \mathbf{a}_j son ortogonales entre sí.
 - $\mathbf{a}_j^T \mathbf{a}_j = 1$
 - La varianza muestral es máxima.

Es decir el factor \tilde{z}_{ij} es la transformación de la observación j -ésima por la i -ésima componente principal.

Por tanto, el proceso que se detalla para un vector aleatorio \mathbf{x} con matriz de covarianzas Σ se puede extender a este caso en el que conocemos la matriz de covarianzas muestrales \mathbf{S} .

Con el objetivo de hacer las demostraciones más sencillas y compactas tomaremos la matriz \mathbf{X} la matriz centrada $\bar{\mathbf{X}}$, es decir:

$$\bar{x}_{ij} = x_{ij} - \bar{x}_j \quad (1.9)$$

Donde \bar{x}_j es la media muestral de la j -ésima variable. Esto hace que $\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$. Esto permite hablar de los valores y vectores propios de \mathbf{S} y de $\mathbf{X}^T \mathbf{X}$ indistintamente ya que los vectores son los mismos y los valores propios son proporcionales.

Bibliografía

- [1] **Chatfield,C y Collins A.J (1989).** *Introduction to multivariate analysis*, Chapman and Hall.
- [2] **Jolliffe I.T.(1986).** *Principal Component Analysis*, Springer-Verlag.
- [3] **Hastie, T.,Tibshirani, R. y Friedman J. (2001),** *The Elements of Statistical Learning, Data Mining, Inference and Prediction* Springer
- [4] **Cuadras, C.M. (2014),** *Nuevos métodos de Análisis Multivariante*, CMC Editions, Barcelona.
- [5] **Abdi, H., y Williams, L. J. (2010).** *Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics* 2(4), 433–459.