

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

<b>Introducción</b>	<b>3</b>
<b>1. Análisis de Componentes Principales</b>	<b>7</b>
1.1. Introducción . . . . .	7
1.2. Definición y cálculo de las Componentes . . . . .	8
<b>Bibliografía</b>	<b>11</b>



# Introducción



# Capítulo 1

## Análisis de Componentes Principales

### 1.1. Introducción

Sea una población de la que se han tomado  $n$  muestras de las cuales hemos medido  $p$  variables. Del conjunto de datos que resulta, podemos dar la matriz de datos  $\mathbf{X}$  de tamaño  $n \times p$ .

En consecuencia, las muestras  $y_i$ , con  $i = 1 \dots n$  se pueden interpretar como elementos de  $\mathbb{R}^p$ , pero si  $p > 3$ , la representación gráfica de estos datos no se puede realizar

El análisis de componentes principales, aunque normalmente se utiliza con el objetivo de representar los datos de manera sencilla, también podría llegar a ser útil en la reducción de la dimensionalidad para evitar el overfitting en el aprendizaje automático.

Este método es una técnica matemática que dado un vector aleatorio  $\mathbf{X}^T = [X_1, \dots, X_p]$ , con vector de media  $\mu$  y matriz de covarianzas  $\Sigma$ , utiliza transformaciones ortogonales para conseguir las componentes principales.

Estas componentes principales son combinaciones lineales de las variables que forman el vector, de manera que estando correladas las iniciales, las componentes no lo están y se busca calcularlas con la máxima varianza posible.

Esta transformación se busca ya que si varias variables están altamente correladas, entonces están aportando la misma información, siempre que no sean dependientes. Podemos encontrar por esta razón variables que no estén

correladas, lo que implica que no dan la misma información y que en consecuencia pueden aportar mayor información acerca de la variación de los datos, sin tener que ser compartida o repetida por varias variables.

Al finalizar este Análisis de Componentes Principales, obtendremos un conjunto de variables nuevas no correladas entre sí, que son combinación lineal de las iniciales que maximizan la varianza en cada paso.

Añadir que si las variables no están correladas o están cerca de no estarlo, el Análisis de Componentes Principales no tiene sentido ya que el conjunto de componentes principales será parecido a las variables iniciales, con la única diferencia que estarán ordenadas por orden creciente de varianza.

## 1.2. Definición y cálculo de las Componentes

Sea un vector aleatorio  $\mathbf{X}^T = [X_1, \dots, X_p]$  con media  $\mu$  y matriz de covarianzas  $\Sigma$ .

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

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

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

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

El cálculo de la primera componente principal se lleva a cabo con un proceso de optimización de la función  $Var(\mathbf{Y}_1)$  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  $g(\mathbf{x}) = g(x_1 \dots x_p) = c$  entonces existe una constante  $\lambda$  de manera 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  $f(\mathbf{x})$



Sea ahora la función  $L(\mathbf{x}) = f(\mathbf{x}) - \lambda[g(x) - c]$  entonces podemos simplificar la expresión anterior a:

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

En nuestro caso particular  $L(\mathbf{a}_1) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 - \lambda[\mathbf{a}_1^T \mathbf{a}_1 - 1]$ . Al derivarla obtenemos que :

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

Igualando a 0 tenemos la siguiente ecuación:

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

Para que la ecuación tenga una solución que no sea la trivial, tenemos que elegir  $\lambda$  de manera que  $|\Sigma - \lambda I| = 0$ . Luego  $\lambda$  es uno de los valores propios de la matriz. Generalmente una matriz  $(p \times p)$  tiene  $p$  valores propios  $\{\lambda_1, \dots, \lambda_p\}$  y como  $Var(\mathbf{Y}_1) = Var(\mathbf{a}_1^T \mathbf{X}) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 = \mathbf{a}_1^T \lambda \mathbf{a}_1 = \lambda$  que es la variable a maximizar, elegimos  $\lambda = \max\{\lambda_1, \dots, \lambda_p\}$ , por tanto, el vector  $\mathbf{a}_1$  es el vector propio con valor propio  $\lambda = \lambda_1$  reordenando si es necesario.

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

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

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

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

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 sepamos sus valores propios.

**Corolario 1.2.1.** Las componentes principales son todas perpendiculares entre sí.

# 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.