

# Componentes Principales

Nicolás Kossacoff

Octubre 2024

## 1. Introducción

### 1.1. Transformación Lineal

Definimos una **transformación lineal** como una función  $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$  y que cumple con dos propiedades:

- Sean  $v$  y  $w$  dos vectores en  $\mathbb{R}^n$ , se cumple que  $T(v + w) = T(v) + T(w)$ .
- Para todo  $\lambda \in \mathbb{R}$  y  $v \in \mathbb{R}^n$ , se cumple que  $T(\lambda v) = \lambda T(v)$ .

#### 1.1.1. Matriz Asociada

Si  $T$  es una transformación lineal definida en un espacio de **dimensión finita**, entonces podemos asegurar que existe una **matriz asociada**  $A$ , tal que aplicar la transformación lineal a cualquier vector  $v$  es equivalente a multiplicarlo por dicha matriz,  $T(v) = Av$ .

### 1.2. Autovalores

Sea  $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$  una transformación lineal, decimos que  $v \neq 0$  es un **autovector** de  $T$  con **autovalor**  $\lambda_v \in \mathbb{R} - \{0\}$ , si aplicarle la transformación lineal  $T$  es equivalente a multiplicarlo por su autovalor,  $T(v) = \lambda_v v$ .

Dado que  $T$  se encuentra definida en un espacio de dimensión finita, entonces tiene asociada una matriz  $A$ . Si  $A$  es una simétrica (i.e.,  $A = A'$ ) y positiva (i.e.  $z'Az > 0, \forall z \neq 0$ ), entonces tenemos una base ortonormal de autovectores para nuestro espacio vectorial.

Esto quiere decir que podemos construir cualquier vector en nuestro espacio utilizando una combinación lineal de los autovectores.

## 2. Componentes Principales

**Componentes Principales** es una técnica de **reducción de dimensiones** que se basa en construir una transformación lineal que lleve nuestro conjunto de datos de

$\mathbb{R}^p$  a  $\mathbb{R}^q$ , con  $q \ll p$ .

## 2.1. Definición

Sea  $X \in \mathbb{R}^p$  un vector aleatorio con  $E(X) = \mu$  y  $Var(X) = \Sigma$ . La matriz  $\Sigma$  es simétrica y positiva, por lo tanto, tenemos una base de autovectores ortonormales. Esto quiere decir que tenemos  $p$  autovectores de  $\Sigma$ ,  $\{\gamma_1, \dots, \gamma_p\}$ , asociados a un conjunto de autovalores,  $\lambda_1 \geq \dots \geq \lambda_p$  que podemos utilizar para construir cualquier vector dentro del espacio vectorial.

Dicho esto, podemos escribir a  $X$  como:

$$X = \mu + \sum_{j=1}^p \gamma_j' (X - \mu) \gamma_j = \mu + \sum_{j=1}^p \langle X - \mu, \gamma_j \rangle \cdot \gamma_j \quad (1)$$

donde:

$$v = \sum_{j=1}^p \langle X - \mu, \gamma_j \rangle \gamma_j = (X - \mu) \gamma_1 + \dots + (X - \mu) \gamma_p = v_1 + \dots + v_p \quad (2)$$

Las coordenadas de  $v$ ,  $\{v_1, \dots, v_p\}$ , son los componentes principales de la observación  $X$  (i.e., tenemos  $p$  componentes principales por observación, uno por cada feature).

**Observación.** La  $j$ -ésima componente principal,  $v_j = \gamma_j' (X - \mu) = \langle X - \mu, \gamma_j \rangle$ , es la proyección ortogonal de  $(X - \mu)$  sobre la dirección del autovector  $\gamma_j$ .

## 2.2. Propiedades

**Propiedad 1.** Los componentes principales,  $\{v_1, \dots, v_p\}$ , son no correlacionados y  $Var(v_j) = \lambda_j$ . Es decir:

$$Var(v) = diag(\lambda_1, \dots, \lambda_p)$$

De la [Propiedad 1](#) se desprende que cada componente principal aporta información que los componentes anteriores no aportaron.

**Propiedad 2.** Sea  $H_0$  el sub-espacio que generamos con los  $q$  primeros autovectores y sea  $H$  otro sub-espacio de la misma dimensión. Llamemos  $\pi(X, H)$  a la proyección de  $X$  en el sub-espacio  $H$ . Entonces:

$$E[||X - \pi(X, H_0)||^2] \leq E[||X - \pi(X, H)||^2]$$

Esta propiedad nos dice que los componentes principales nos dan el mejor ajuste lineal sobre sub-espacio de menor dimensión.

### 2.2.1. Relación con la varianza

Supongamos que queremos encontrar el vector  $a$  que genere la combinación lineal,  $a'X$ , con mayor varianza, es decir, cuya dirección aporta más información. Por propiedad de componentes principales se cumple que ese vector es el primer autovector,  $\gamma_1$ , el cual está relacionado con el primer componente:

$$\max_a Var(a'X) = Var(v_1) \text{ s.a. } \|a\| = 1 \quad (3)$$

Ahora, queremos encontrar el segundo vector con mayor varianza pero que no aporte la misma información que el vector anterior,  $\gamma_1$ . Para eso planteamos el siguiente problema de maximización:

$$\max_a Var(a'X) = Var(v_2) \text{ s.a. } \|a\| = 1 \wedge Cov(a'X, v_1) = 0 \quad (4)$$

Al igual que antes, la solución a este problema de maximización es el segundo autovector,  $\gamma_2$ . La [Ecuación \(4\)](#) se puede generalizar para los siguientes casos de la siguiente manera:

$$\begin{aligned} \max_a Var(a'X) &= Var(v_k) \\ \text{s.a.} \\ \|a\| &= 1 \wedge Cov(a'X, v_j) = 0, \forall j \in [1, k-1] \end{aligned}$$

## 2.3. Número de componentes

Por la [Propiedad 1](#) sabemos que la varianza del componente principal  $v_j$  es igual a su autovalor  $\lambda_j$ . También sabemos que para obtener la varianza total simplemente tenemos que sumar los autovalores:

$$Var(v) = \sum_{j=1}^p \lambda_j = \text{traza}(\Sigma)$$

Una criterio para elegir la cantidad  $q$  de componentes principales que vamos a agregar es mirar la proporción de la variabilidad que está siendo explicada por esos componentes:

$$\text{prop} = \frac{\sum_{j=1}^q \lambda_j}{\text{traza}(\Sigma)}$$

## 2.4. Inferencia

En la práctica, la media poblacional ( $\mu$ ) y la matriz de covarianza ( $\Sigma$ ) son desconocidos. En estos casos tenemos que estimarlos utilizando una muestra aleatoria

$X_1, \dots, X_n$ , tal que:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$Q = \sum_{i=1}^n (X_i - \hat{\mu})(X_i - \hat{\mu})'$$

$$\hat{\Sigma} = \frac{Q}{n}$$

Una vez calculada la matriz  $\hat{\Sigma}$  podemos calcular los autovectores y calcular los componentes principales al igual que en la versión poblacional.