

Universidad de Santiago de Chile
Facultad de Ingeniería
Depto. de Ingeniería Informática.



Análisis de Datos
Capítulo II
“Análisis de Componentes Principales”

Profesor: Dr. Max Chacón

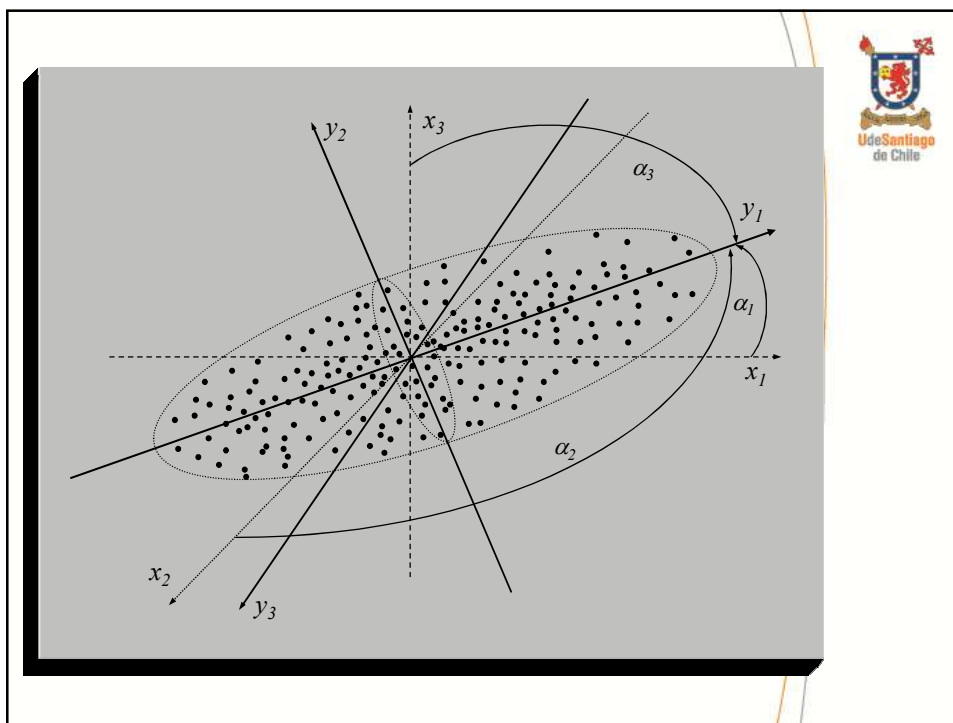
Objetivos:

- Comprender la interpretación geométrica del análisis de componentes principales
- Comprender los conceptos de varianza e información.
- Plantear matemáticamente el problema de ACP
- Resolver matemáticamente usando método de Lagrange.
- Deducir el problema de los valores característicos y utilizar el método para caracterizar nuevo conocimiento.



2.1 Interpretación gráfica

Suponga que se tiene una nube de puntos en \mathfrak{R}^3 , la cual tiene una dispersión mayor en un sentido (mayor varianza). Pero el sistema de coordenadas $[x_1, x_2, x_3]$ no está ubicado en el sentido de la mayor dispersión.





Se requiere una transformación (y_i) que ubique el primer eje en el sentido de la mayor dispersión de puntos, luego el segundo eje en la dirección de la segunda dispersión y así sucesivamente.

Esto permitirá representar de mejor forma la varianza del conjunto de datos y, eventualmente, eliminar las componentes de orden mayor que representen una menor varianza, lo cual es equivalente a contener menor información.



De geometría analítica se tiene que para una transformación ortonormal, los ángulos a_i son los cosenos directores de la transformación:

$$a_{11} = \cos a_1; \quad a_{12} = \cos a_2; \quad a_{13} = \cos a_3$$

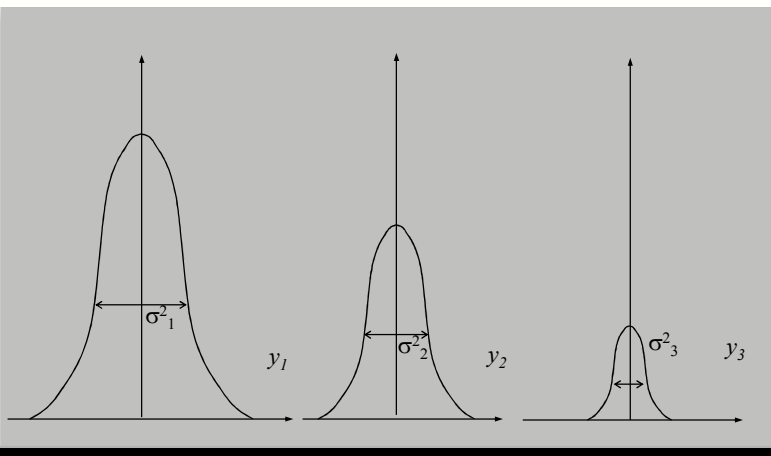
$$a_{11}^2 + a_{12}^2 + a_{13}^2 = 1$$

para una transformación ortonormal

Entonces, un nuevo punto (rotado en torno a las medias) en las coordenadas y_i será:

$$y_1 = a_{11}(x_1 - \bar{x}_1) + a_{12}(x_2 - \bar{x}_2) + a_{13}(x_3 - \bar{x}_3)$$

De esta forma la dispersión de los datos en torno a las nuevas componentes tendrá una distribución decreciente.



- Cálculo algebraico.

Se tiene una matriz X de datos con p columnas (componentes de cada caso) y n filas (número de casos).

Se requiere disminuir la dimensión p del vector \bar{x}_i que representa cada caso a una dimensión q , con $q < p$ y perder la mínima cantidad de información.

Si se trunca directamente el vector \bar{x}_i se producirá un error cuadrático medio igual a la suma de las varianzas de los elementos eliminados de \bar{x}_i .



Se requiere una transformación lineal e invertible T tal que al trincar $T\bar{x}_i$ se produzca una mínima pérdida de varianza.

Se requiere una transformación ortonormal que minimice la varianza de los nuevos y_i para poder truncarlos.

El problema de optimización será: $\bar{y}_j = a_{1j}\bar{x}_1 + a_{2j}\bar{x}_2 + \dots + a_{pj}\bar{x}_p = \bar{a}_j X$

Dada la combinación lineal:

$$\begin{aligned} &\text{minimizar } var(\bar{y}_j) \\ &\text{s.a. } \bar{a}_j^T \bar{a}_j = \delta_{jj} \end{aligned}$$



Sol: Suponiendo que la media de las nuevas componentes es cero, esto es $E[\bar{y}_j] = 0$

$$var(\bar{y}_j) = E[\bar{y}_j \bar{y}_j^T]$$

Pero, de la transformación

Se tiene: $\bar{y}_j = \bar{a}_j X$

$$var(\bar{y}_j) = E[\bar{a}_j^T X X^T \bar{a}_j]$$

$$var(\bar{y}_j) = \bar{a}_j^T E[XX^T] \bar{a}_j$$

Donde $E[XX^T]$ es la matriz de varianzas y co-varianzas, en el caso que los \bar{x}_i estén centrados en la media, es la matriz de correlación $R = E[XX^T]$



Así: $Var(\bar{y}_j) = \bar{a}_j^T \mathbf{R} \bar{a}_j$

Para resolver el problema de optimización con restricciones se hace uso del Lagrangeano aumentado, esto es, incluir las restricciones igualadas a cero, ponderadas por una constante (λ_j), en la función objetivo.

$$L(\bar{a}_j) = \bar{a}_j^T \mathbf{R} \bar{a}_j - \lambda_j(\bar{a}_j^T \bar{a}_j - 1)$$

Para minimizar $\frac{\partial L(\bar{a}_j)}{\partial \bar{a}_j} = 0$, Usando la derivada de una

forma cuadrática $\mathbf{R}\bar{a}_j - \lambda_j \mathbf{I}\bar{a}_j = 0$ o $(\mathbf{R} - \lambda_j \mathbf{I})\bar{a}_j = 0$

Esta ecuación corresponde al problema de los valores propios de la matriz \mathbf{R} .



Tiene solución para $\bar{a}_j \neq 0$ para valores especiales λ_j denominados valores propios de \mathbf{R} . Los \bar{a}_j son los vectores propios de \mathbf{R} .

Si los \bar{a}_j son vectores columna al colocarlos en una matriz, esta será la matriz de transformación

$$\mathbf{T} = [\bar{a}_1, \bar{a}_2, \dots, \bar{a}_p]$$

Como la matriz de correlación es simétrica y de valores reales positivos, sus valores propios serán reales y positivos.

Matricialmente: $\mathbf{R}\mathbf{T} = \mathbf{T}\mathbf{L}$



Con L la matriz de valores propios: $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_p \end{bmatrix}$

Como los vectores de T satisfacen la condición de ortonormalidad $\vec{a}_i^T \vec{a}_j = \delta_{ij}$ entonces $T^T T = I$, esto es $T^T = T^{-1}$.

Pre multiplicando por T^T la forma matricial se tiene que:

$T^T R T = L$ y en términos vectoriales $\vec{a}_j^T R \vec{a}_j = \lambda_j$

Como $\text{var}(y_j) = \vec{a}_j^T R \vec{a}_j$ entonces $\text{var}(y_j) = \lambda_j$



Esto significa que si se ordenan los valores propios en forma decreciente, es posible obtener lo buscado:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$$

Con λ_1 el valor propio de mayor valor asociada a la primera componente principal y_1

Si la matriz R es la matriz de correlación entonces los

valores propios son normalizados y $\sum_{i=1}^p \lambda_i = P$.

Esto significa que cada λ_i representa un porcentaje de la varianza de las nuevas coordenadas y_i .





Para obtener el vector original basta con multiplicar por la matriz de transformación transpuesta.

Como $Y = T^T X$ al pre-multiplicar por T se tiene: $X = TY$.

Los \bar{y}_j son llamadas las componentes principales y tienen la misma dimensión que el vector original \bar{x} .

-2.2. Reducción de dimensionalidad.



Debido a que las componentes principales ahora están ordenadas en orden decreciente de varianza es posible eliminar las últimas perdiendo el mínimo de información.

La información esta representada por los valores propios λ_j .

Si se reduce la dimensión a un punto $q < p$ el error cometido al truncar en la componente q se puede evaluar sumando la contribución de las componentes eliminadas

$$\%e = \frac{100}{P} \sum_{i=q+1}^p \lambda_i$$

Este valor esta dado en % si los vectores y valores propios son extraídos de la matriz R . Si son extraídos de la matriz de varianzas-covarianzas S , se requiere dividir la suma valores propios para obtenerlos en porcentaje.

Una forma de examinar el error cometido es obtener la estimación del vector \hat{x} después de haber truncado en q las componentes de los vectores \bar{y}_j .

$$\hat{x} = \sum_{i=1}^q \bar{y}_i \bar{a}_i \text{ con } q < p$$

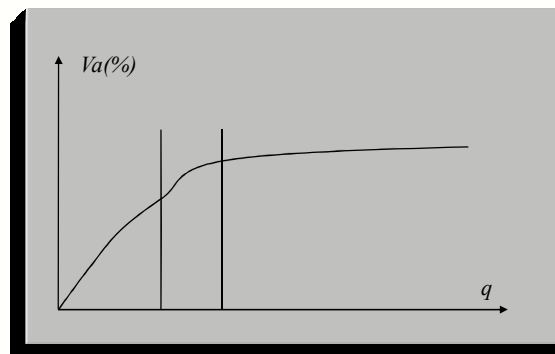


El problema de determinar el valor q en general depende del problema, puesto que al reducir q se reduce la dimensionalidad del nuevo espacio de características pero también aumenta la perdida de información.

Una forma de encontrar un valor para q es graficar la varianza acumulada en función de q de las componentes principales.



Esto es $Va(\%) = \sum_{i=1}^q \lambda_i$



Debido a que la disminución de los valores propios no tiene porque ser monótona es posible encontrar un punto en que el aporte de las últimas componentes sea poco significativo.

