UFM UNIVERSIDAD FRANCISCO MARROQUÍN
VERITAS • LIBERTAS • JVSTITIA

FACULTAD de CIENCIAS ECONÓMICAS

Análisis de componentes principales II

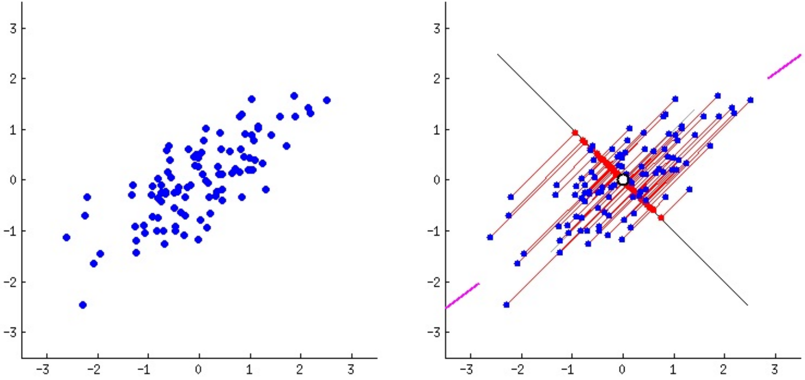
Alan Reyes-Figueroa

Elements of Machine Learning (Aula 07) 07.febrero.2023

Componentes principales

<u>Objetivo</u>: encontrar una estructura subyacente en los datos.

- Proyectar a un subespacio adecuado.



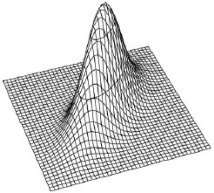
Componentes principales

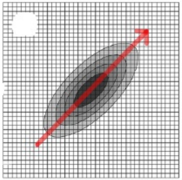
Caso particular 1D: (proyectamos a un subespacio 1-dimensional).

Suponga que proyectamos a un subespacio $\langle \ell \rangle \;\Rightarrow\; \langle \ell, X\rangle = \ell^T X$.
Buscamos maximizar

$$\max_{||\ell||=1} Var(\ell^T X) = \max_{\ell\neq 0} \frac{Var(\ell^T X)}{\ell^T \ell} = \max_{\ell\neq 0} \frac{\ell^T Var(X)\ell}{\ell^T \ell} = \max_{\ell\neq 0} \frac{\ell^T(\mathbb{X}^T\mathbb{X})\ell}{\ell^T \ell}.$$

(cociente de Rayleigh).





UNIVERSIDAD FRANCISCO MARROQUÍN

Teorema (Descomposición en valores singulares (SVD))

Sea $A \in \mathbb{R}^{n\times d}$ una matriz de rango k. Para todo $1 \leq r \leq k$, existen matrices $U \in \mathbb{R}^{n\times r}$, $S \in \mathbb{R}^{r\times r}$, $V \in \mathbb{R}^{d\times r}$, tales que

$$A = USV^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

con

- *las columnas $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{R}^n$ de U son los autovectores de AA^T,*
- *las columnas $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^d$ de V son los autovectores de A^TA, $S = diag(\sigma_1, \ldots, \sigma_r)$, $\sigma_i^2 = \lambda_i$, con λ_i los autovectores de $\mathbf{u}_i$ y de $\mathbf{v}_i$,*
- *Además, $\sigma_i \mathbf{u}_i = A\mathbf{v}_i$ y $\sigma_i \mathbf{v}_i = A^T\mathbf{u}_i$, para $i = 1, 2, \ldots, r$.*

Descomposición SVD

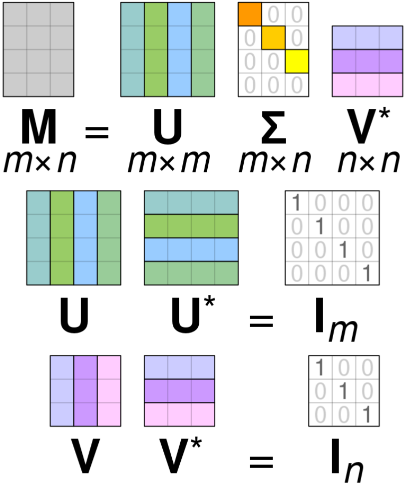
El teorema de descomposición espectral ocurre como un caso particular de la descomposición SVD:

Caso especial: A simétrica

$$A = USU^T = U\Lambda U^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{u}_i^T.$$

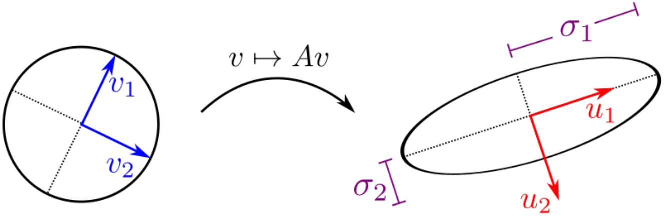
En este caso los autovectores de A y $A^TA = A^2 = AA^T$ coinciden, y los autovalores de A al cuadrado son los autovalores de A^TA.

Direcciones Principales



UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Direcciones Principales



$$A = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} — & v_1^T & — \\ — & v_2^T & — \\ & \vdots & \\ — & v_n^T & — \end{bmatrix}$$

Direcciones Principales

En general, si queremos proyectar los datos $\mathbb{X}$ en $\mathbb{R}^d$ a un subespacio de dimensión r, $0 < r < d$, debemos encontrar las primeras r direcciones principales.

Estas direcciones pueden encontrarse de dos formas:

- las direcciones principales $\mathbf{v}_i$ son los autovectores asociados a los primeros r autovectores de $Cov(\mathbb{X})$.
- también pueden calcularse como los primeros r vectores, asociados a los r mayores valores singulares de $\mathbb{X}$.

Teorema (Eckart-Young)

Sea $A \in \mathbb{R}^{n\times d}$, $n \geq d$, una matriz cuya descomposición SVD está dada por

$$A = USV^T = \sum_{i=1}^{d} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Entonces, la matriz $\widehat{A}_r$ de rango r, $1 \leq r \leq d$, que mejor aproxima A en el sentido de minimizar

$$\min_{rank\ \widehat{A}_r \leq r} ||A - \widehat{A}_r||_F^2$$

se obtiene de truncar la descomposición en valores dingulares de A:

$$\widehat{A}_r = U_r S_r V_r^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Aproximaciones de bajo rango

Teorema (Eckart-Young)

donde

$$U_r = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_r],\ \ S_r = diag(\sigma_1, \sigma_2 \ldots, \sigma_r),\ \ V_r = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_r].$$

En ese caso, el error de aproximación está dado por

$$||A - \widehat{A}_r||_F^2 = \sum_{i=r+1}^{d} \lambda_i,$$

o

$$||A - \widehat{A}_r||_2^2 = \lambda_{r+1}.$$

Aproximaciones de bajo rango

Obs!

- Las direcciones $\mathbf{u}_i$ se llaman las **componentes principales** de $\mathbb{X}$.
- La descomposición SVD proporciona un mecanismo para proyectar los datos al "mejor" subespacio de dimensión $r \leq d$. Dicha proyección se obtiene haciendo
$$\mathbb{X}_{proj} = \mathbb{X}\, V_r^T.$$
- Los autovalores λ_i de $\mathbb{X}^T\mathbb{X}$ nos proporcionan un mecanismo para medir el error, vía $||A - \widehat{A}_r||_F^2 = \sum_{i=r+1}^d \lambda_i$.
- El cociente $\dfrac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^d \lambda_i}, \quad r = 1, 2, \ldots, d$, se interpreta como el porcentaje de variabilidad de los datos $\mathbb{X}$ que es explicada por las primeras r componentes principales.

Ejemplos

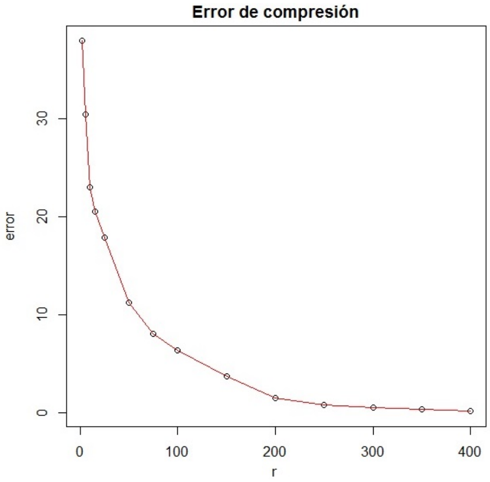
Compresión de imágenes usando PCA.

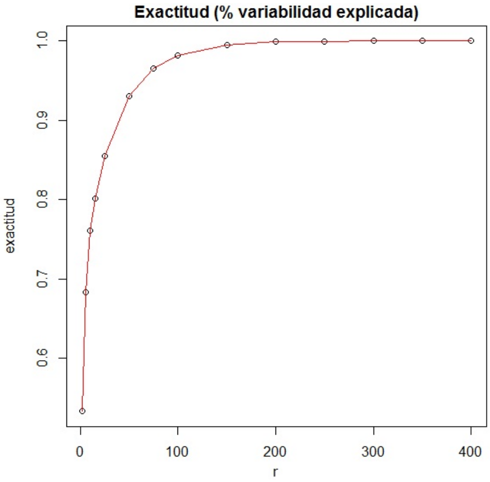
Original $r = 1$ $r = 2$ $r = 4$ $r = 8$ $r = 16$ $r = 32$ $r = 64$

Imagen Original (256×256), aproximaciones con rango = 1, 2, 4, 8, 16, 32, 64.

UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Ejemplos

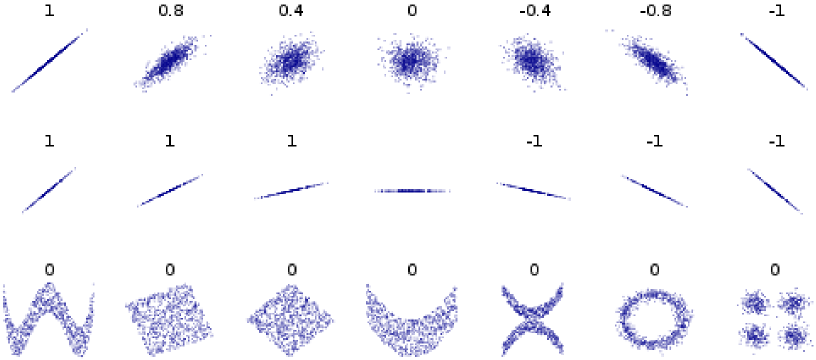


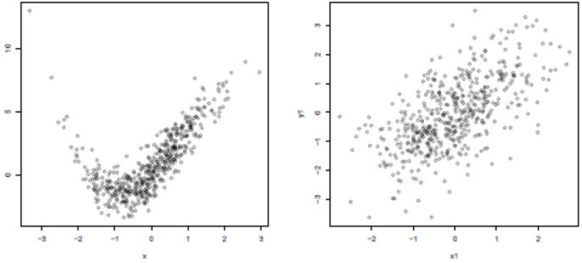


UFM UNIVERSIDAD FRANCISCO MARROQUÍN

Ejemplos

En PCA la estructura de los datos se capta solamente a través de las matrices $Cov(X)$ o $Corr(X)$.



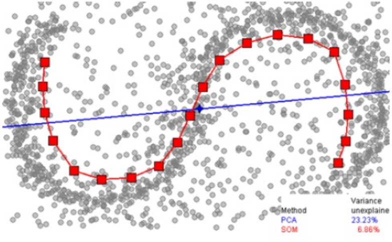


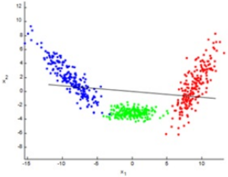
Dos veces misma correlación. =(

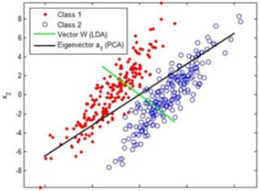
Obs.

- Cuidado con desviaciones fuertes de normalidad.
- Lo ideal es investigar la normalidad de los datos en la práctica, al menos ver si escala es continua, distribución unimodal, simétrica, nubes parecidas a elipses, ...

Contraejemplos







UFM UNIVERSIDAD FRANCISCO MARROQUÍN