Datos Masivos II

Análisis de componentes principales

Blanca Vázquez-Gómez y Gibran Fuentes-Pineda

29 de agosto de 2022

Recordando la maldición de la dimensionalidad

- Objetos cada vez más dispersos conforme aumenta el número de dimensiones

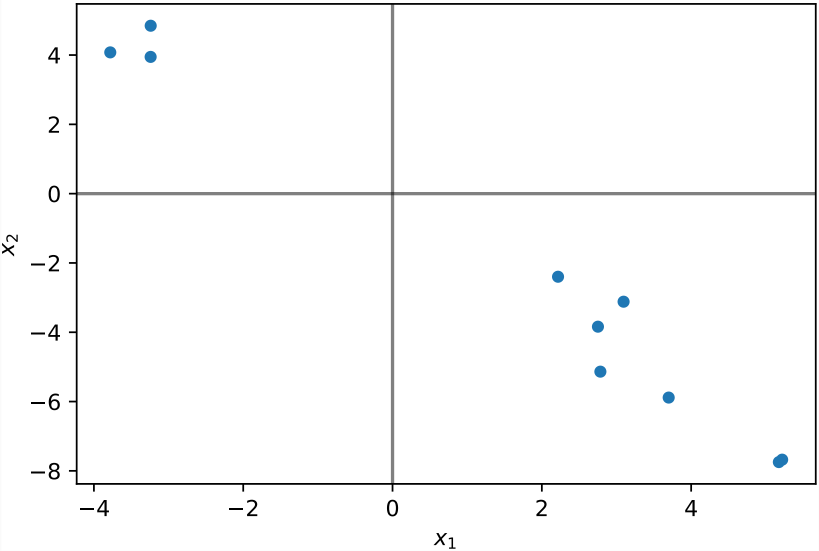
- Ejemplos pueden vivir en una variedad de muchas menores dimensiones que el espacio original

Imagen tomada de Li and Jain, 2005

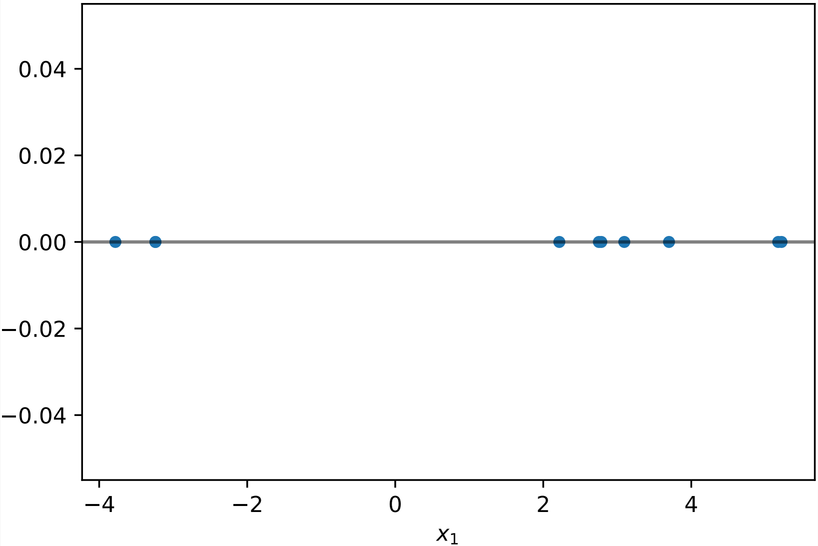
Análisis de componentes principales (PCA)

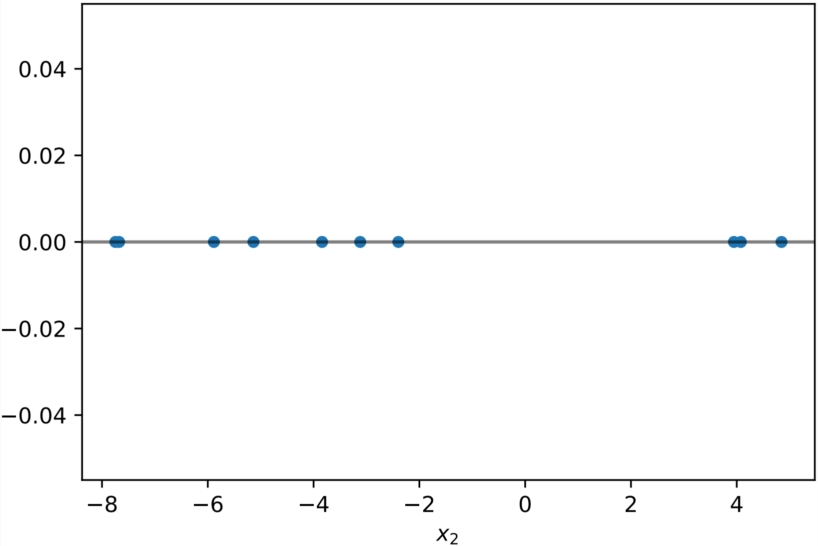
- Proyección ortogonal de un conjunto de vectores
- Genera una nueva vista
- Aplicaciones
  - Visualización
  - Extracción de características
  - Reducción de dimensionalidad
  - Compresión

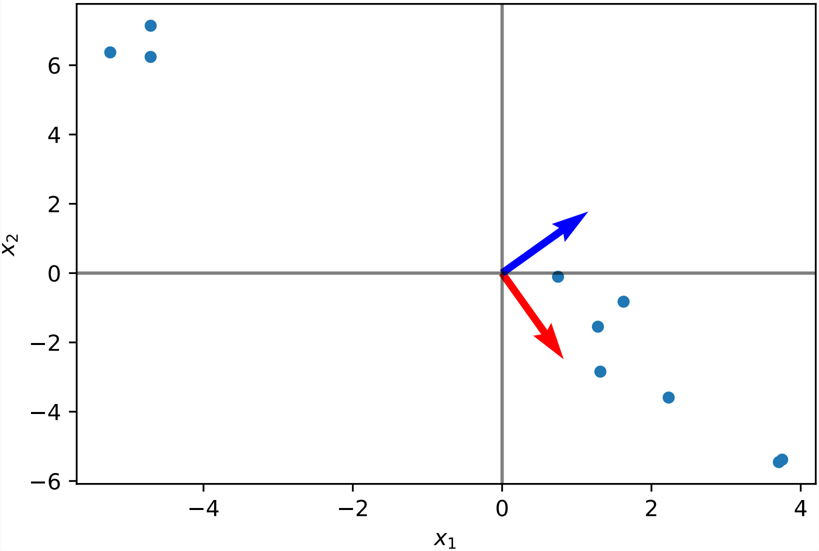
Intuición: datos en 2D



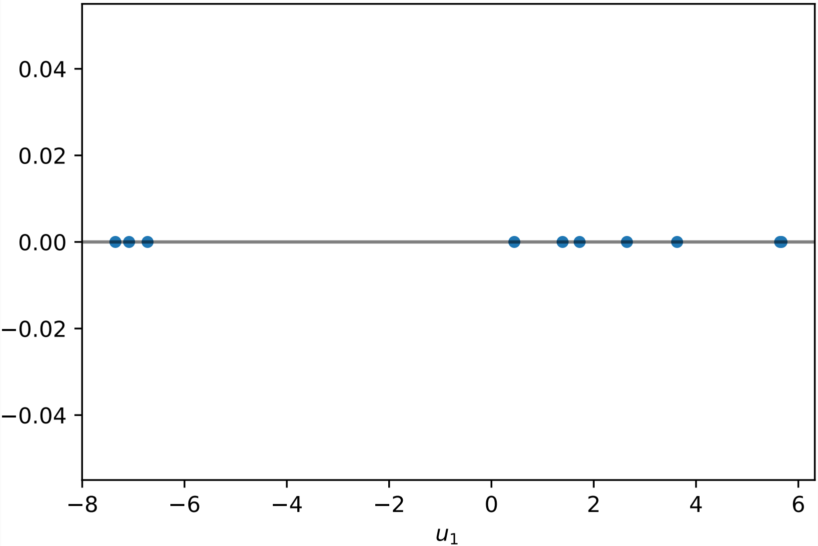
Intuición: datos vistos desde el eje x

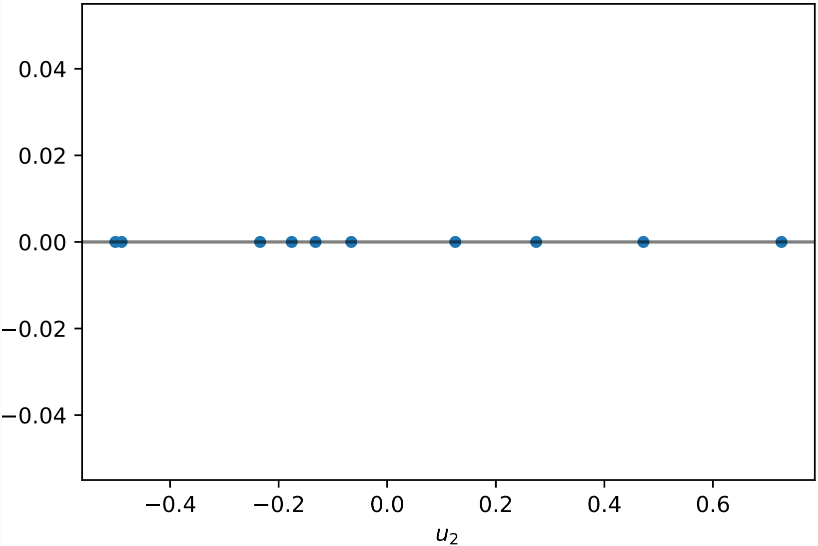



0.04
0.02
0.00
−0.02
−0.04
−8
−6
−4
−2
0
2
4
x_2


6
4
2
0
−2
−4
−6
x_2
−4
−2
0
2
4
x_1

Intuición: datos proyectados sobre el eje u_1




0.04
0.02
0.00
−0.02
−0.04
−0.4
−0.2
0.0
0.2
0.4
0.6
u2

PCA: formulación de máxima varianza

- Dado un conjunto de vectores $\mathsf{X} = \{\mathsf{x}^{(1)}, \ldots, \mathsf{x}^{(n)}\}$ de d dimensiones, el primer componente principal es el vector $\mathbf{u}_1$ que maximice la varianza de los datos proyectados, donde $\mathbf{u}_1$ es un vector de d dimensiones

PCA POR MÁXIMA VARIANZA (1)

- La proyección de un vector sobre el componente principal está dado por

$$\hat{\mathbf{x}}^{(i)} = \mathbf{u}_1^\top \mathbf{x}^{(i)}$$

PCA por máxima varianza (1)

- La proyección de un vector sobre el componente principal está dado por

$$\hat{\mathrm{x}}^{(i)} = \mathbf{u}_1^\top \mathrm{x}^{(i)}$$

- La media de los datos proyectados es $\mathbf{u}_1^\top \bar{\mathrm{x}}$, donde

$$\bar{\mathrm{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{x}^{(i)}$$

- La proyección de un vector sobre el componente principal está dado por

$$\hat{\mathrm{x}}^{(i)} = \mathrm{u}_1^\top \mathrm{x}^{(i)}$$

- La media de los datos proyectados es $\mathrm{u}_1^\top \bar{\mathrm{x}}$, donde

$$\bar{\mathrm{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}^{(i)}$$

- La varianza es $\frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{u}_1^\top \mathrm{x}^{(i)} - \mathrm{u}_1^\top \bar{\mathrm{x}}\right] = \mathrm{u}_1^\top \mathrm{S}\mathrm{u}_1$, donde

$$\mathrm{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}^{(i)} - \bar{\mathrm{x}})(\mathrm{x}^{(i)} - \bar{\mathrm{x}})^\top$$

PCA por máxima varianza (2)

- Queremos encontrar el vector $\mathbf{u}_1$ que maximice la varianza de los datos proyectados $\mathbf{u}_1^\top \mathbf{S} \mathbf{u}_1$, con la restricción que $\mathbf{u}_1^\top \mathbf{u}_1 = 1$

PCA por máxima varianza (2)

- Queremos encontrar el vector $\mathbf{u}_1$ que maximice la varianza de los datos proyectados $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1$, con la restricción que $\mathbf{u}_1^\top \mathbf{u}_1 = 1$

- Podemos formular esta restricción usando multiplicadores de Lagrange: $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^\top \mathbf{u}_1)$

PCA por máxima varianza (2)

- Queremos encontrar el vector $\mathbf{u}_1$ que maximice la varianza de los datos proyectados $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1$, con la restricción que $\mathbf{u}_1^\top \mathbf{u}_1 = 1$

- Podemos formular esta restricción usando multiplicadores de Lagrange: $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^\top \mathbf{u}_1)$

- Derivando e igualando a cero, tenemos

$$\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

PCA por máxima varianza (2)

- Queremos encontrar el vector $\mathbf{u}_1$ que maximice la varianza de los datos proyectados $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1$, con la restricción que $\mathbf{u}_1^\top \mathbf{u}_1 = 1$

- Podemos formular esta restricción usando multiplicadores de Lagrange: $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^\top \mathbf{u}_1)$

- Derivando e igualando a cero, tenemos

$$\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

- Esto es, $\mathbf{u}_1$ es un vector propio de $\mathbf{S}$, donde $\lambda_1 = \mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1$ es su valor propio que se corresponde con la varianza de los datos proyectados

PCA POR MÁXIMA VARIANZA (3)

- El siguiente componente principal es el vector propio que maximice la varianza de los datos proyectados entre el conjunto de vectores ortogonales a los que ya han sido elegidos.

  Este proceso se realiza de forma incremental hasta obtener los K componentes principales.

- El conjunto de K componentes principales forman una base ortonormal de funciones.

Proyección y reconstrucción con PCA

- Para proyector un vector $\mathsf{x}^{(i)}$ sobre los componentes principales

$$\mathsf{z}^{(i)} = \mathsf{U}^\top \left[\mathsf{x}^{(i)} - \bar{\mathsf{x}}\right]$$

  donde $\mathsf{U} = [\mathsf{u}_1, \mathsf{u}_2, \ldots, \mathsf{u}_K]$ es una matriz cuyas columnas se corresponden con los K componentes principales.

- La reconstrucción está dada por

$$\hat{\mathsf{x}}^{(i)} = \mathsf{U}\mathsf{z}^{(i)} + \bar{\mathsf{x}}$$

PCA por vectores y valores propios

- Busca subespacio de K dimensiones que maximiza varianza (o minimiza error) de los ejemplos
    - Definido por eigenvectores $\mathbf{u}_1, \ldots, \mathbf{u}_K$ con eigenvalores más grandes $\lambda_1, \ldots, \lambda_K$ de la matriz de covarianza

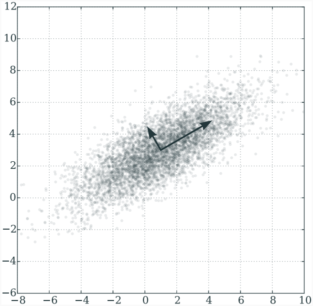
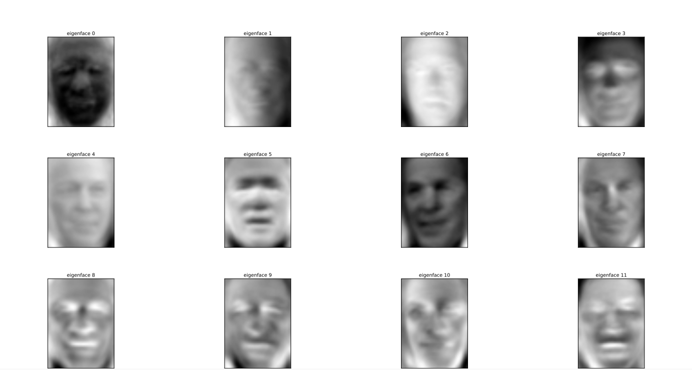


Figura tomada de Wikipedia (*Principal Component Analysis*)

PCA aplicado a imágenes de rostros

- Componentes principales se toman como base (**eigenfaces**)
- Nuevos rostros se proyectan en subespacio encontrado para ser comparados



Ejemplo de *http://scikit-learn.org*

Solución por SVD (1)

- Para cualquier matriz $\mathbf{X}$ de $n \times d$, la descomposición SVD está dada por

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$

- Substituyendo en $\mathbf{X}\top\mathbf{X}$ tenemos

$$\mathbf{X}^\top\mathbf{X} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right)^\top\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right) = \mathbf{V}\boldsymbol{\Sigma}^\top\underbrace{\mathbf{U}^\top\mathbf{U}}_{\mathbf{I}}\boldsymbol{\Sigma}\mathbf{V}^\top$$

$$= \mathbf{V}\underbrace{\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}}_{\mathbf{D}}\mathbf{V}^\top = \mathbf{V}\mathbf{D}\mathbf{V}^\top$$

- Si $\mathbf{C} = \mathbf{X}^\top\mathbf{X}$ y multiplicamos ambos lados por $\mathbf{V}$

$$\mathbf{C}\mathbf{V} = \mathbf{V}\mathbf{D}\underbrace{\mathbf{V}^\top\mathbf{V}}_{\mathbf{I}} = \mathbf{V}\mathbf{D}$$

- Entonces, los vectores propios de $\mathbf{C}$[1] se corresponden con $\mathbf{V}$ y sus valores propios asociados con la diagonal de $\mathbf{D}$. Cuando $\mathbf{X}$ está centrado, esto corresponde a PCA sin

Solución por SVD (2)

- De forma equivalente

$$\mathbf{X}\mathbf{X}^\top = \left(\mathbf{U}\Sigma\mathbf{V}^\top\right)\left(\mathbf{U}\Sigma\mathbf{V}^\top\right)^\top = \mathbf{U}\Sigma\underbrace{\mathbf{V}^\top\mathbf{V}}_{\mathbf{I}}\Sigma^\top\mathbf{U}^\top$$
$$= \mathbf{U}\underbrace{\Sigma\Sigma^\top}_{\mathbf{D}}\mathbf{U}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$

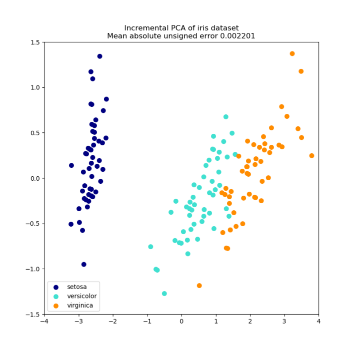
- Si $\mathbf{C} = \mathbf{X}\mathbf{X}^\top$ y multiplicamos ambos lados por $\mathbf{U}$

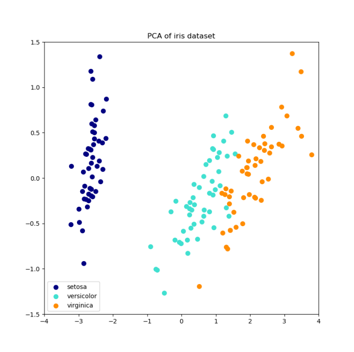
$$\mathbf{C}\mathbf{U} = \mathbf{U}\mathbf{D}\underbrace{\mathbf{U}^\top\mathbf{U}}_{\mathbf{I}} = \mathbf{U}\mathbf{D}$$

- Entonces, los vectores propios de $\mathbf{C}$[2] se corresponden con $\mathbf{U}$ y sus valores propios asociados con la diagonal de $\mathbf{D}$

[2] Si $\mathbf{X}$ está centrada se correspondería con la matriz de covarianza sin la división entre $n-1$

PCA incremental





Ejemplo de *http://scikit-learn.org*