Datos Masivos II

Análisis de componentes principales

Blanca Vázquez-Gómez y Gibran Fuentes-Pineda

22 de agosto de 2022

Recordando la maldición de la dimensionalidad

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

La hipótesis de la variedad

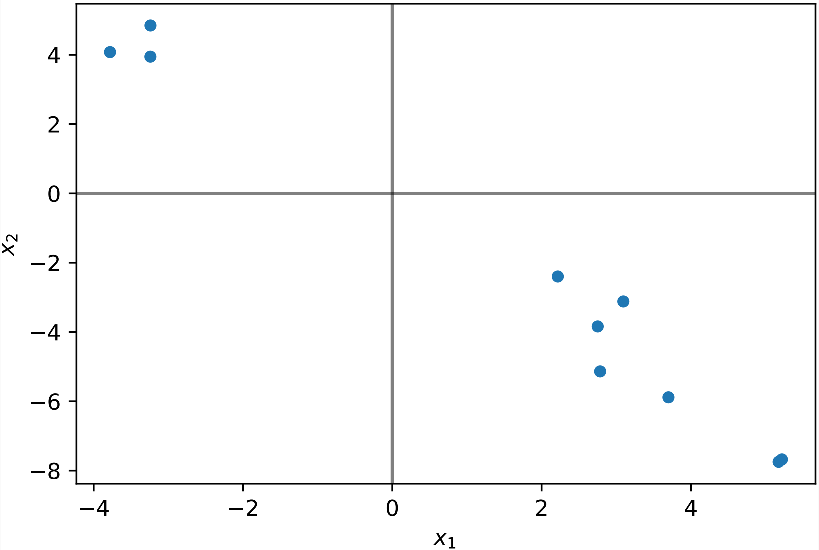
- Ejemplos pueden vivir en una variedad de muchas menores dimensiones que el espacio original

Imagen tomada de Li and Jain, 2005

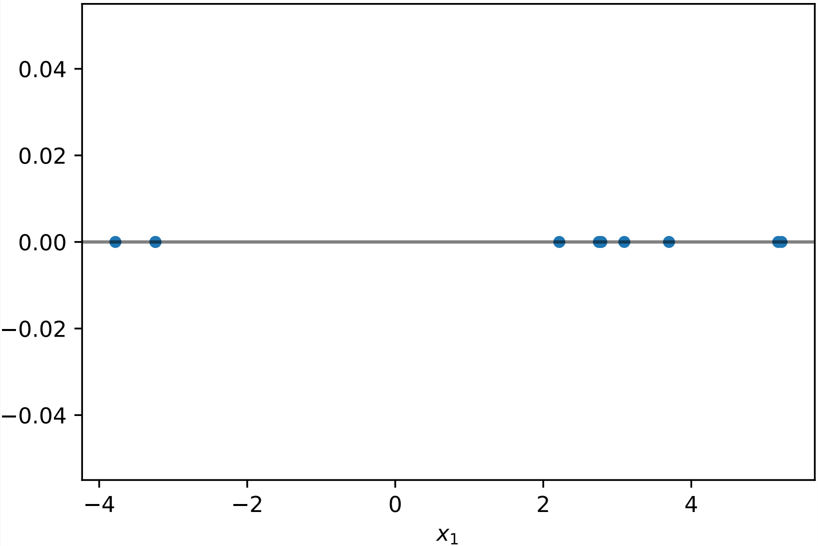
Análisis de componentes principales (PCA)

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

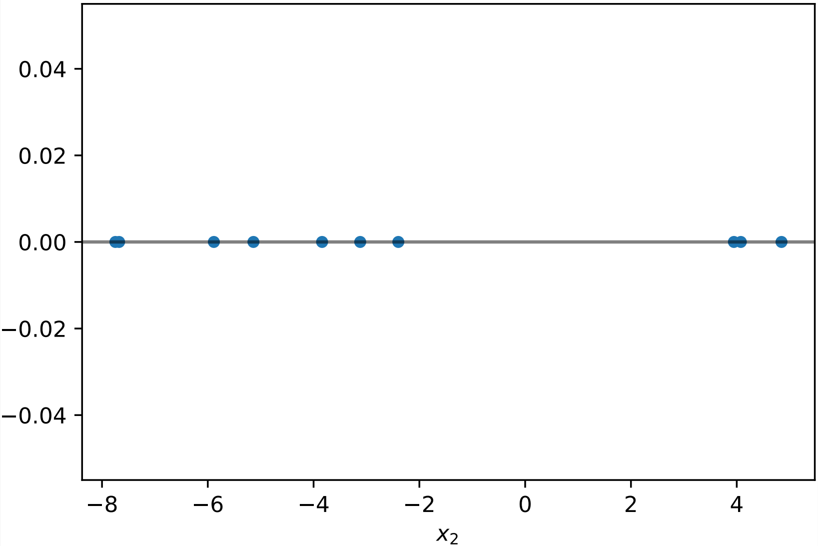
Intuición: datos en 2D



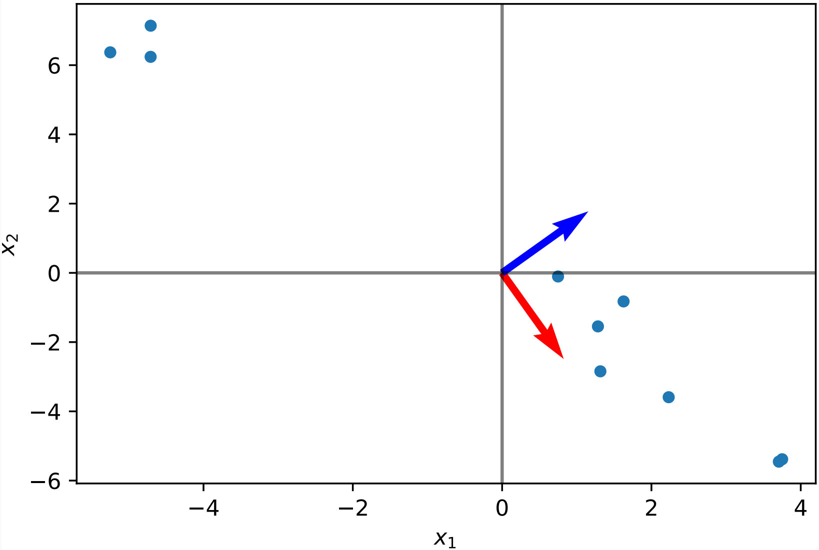
Intuición: datos vistos desde el eje x



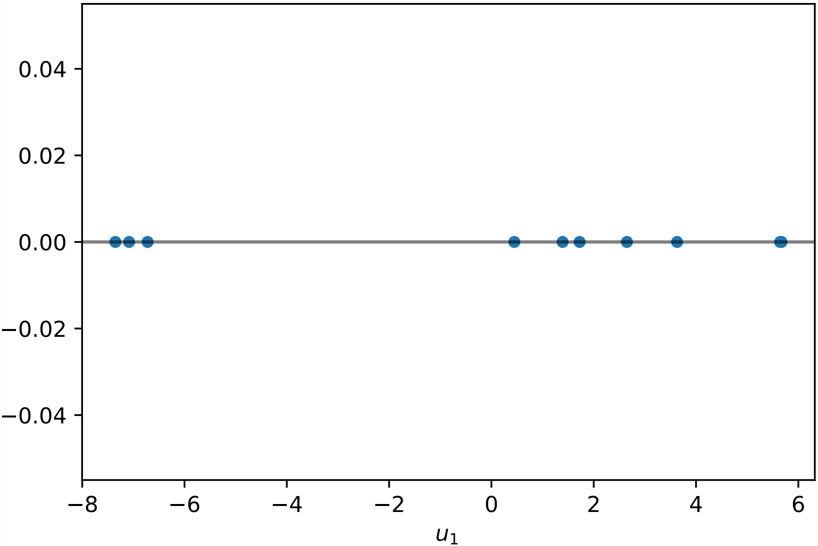
Intuición: datos vistos desde el eje y



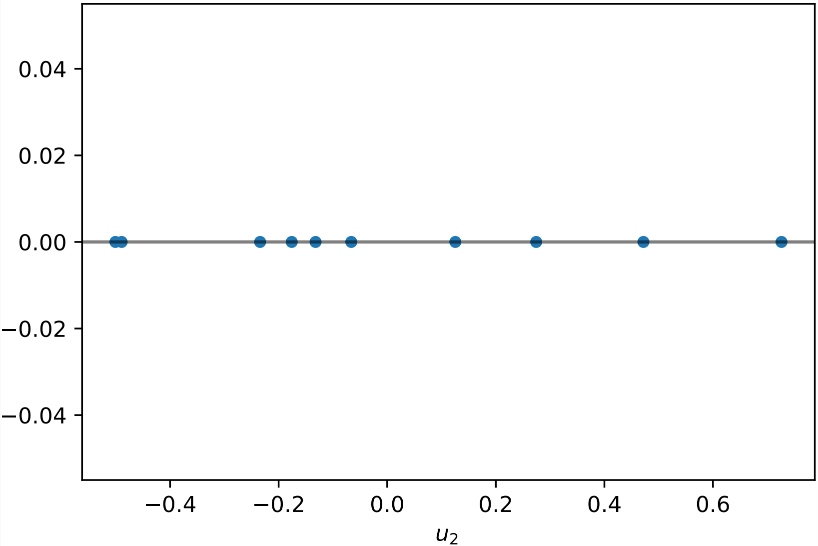
Intuición: nuevos ejes u_1 y u_2



INTUICIÓN: DATOS PROYECTADOS SOBRE EL EJE u_1



Intuición: datos proyectados sobre el eje u_2



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

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

PCA por máxima varianza (1)

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

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

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

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

PCA por máxima varianza (1)

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

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

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

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

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

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

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: $\mathbf{u}_1^\top \mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^\top \mathbf{u}_1)$

- 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.

- Para proyector un vector $\mathsf{x}^{(i)}$ sobre los componentes principales
$$\hat{\mathsf{x}}^{(i)} = \mathsf{U}^\top \left[\mathsf{x}^{(i)} - \bar{\mathsf{x}}\right]$$
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
$$\mathsf{x}_{\text{rec}}^{(i)} = \mathsf{U}\hat{\mathsf{x}}^{(i)} + \bar{\mathsf{x}}$$

- 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

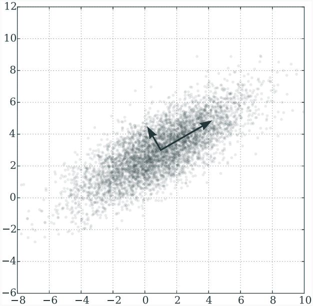
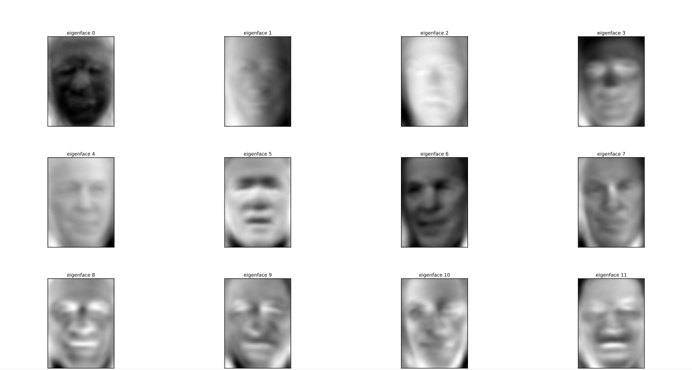


Figura tomada de Wikipedia (*Principal Component Analysis*)

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*