

SEGMENTACIÓ: Principal Component Analysis

Assignatura: Visió per Computador
5è Enginyeria en Informàtica

Dept. Electrònica, Informàtica i Automàtica

Què és el PCA?

- Tècnica de descripció estadística per la visualització aproximada de la informació continguda en una taula de dades.
- Es basa en la reducció de la dimensionalitat d'un conjunt de variables
- Intenta identificar quines són les variables importants (*meaningful*)
- Molt utilitzat en imatges Multiespectrals (im. Satèl·lit)

Què és el PCA?

Imatge original

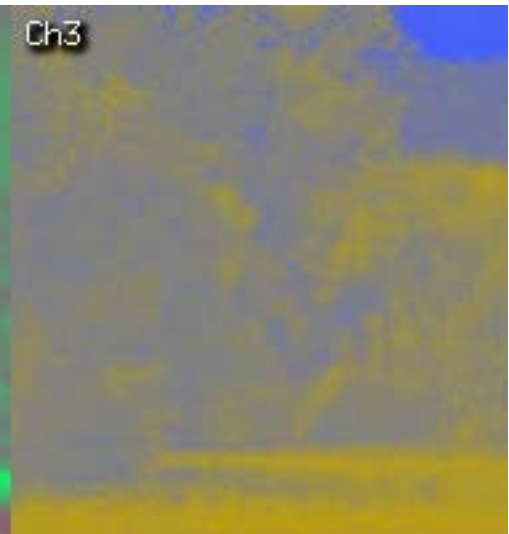
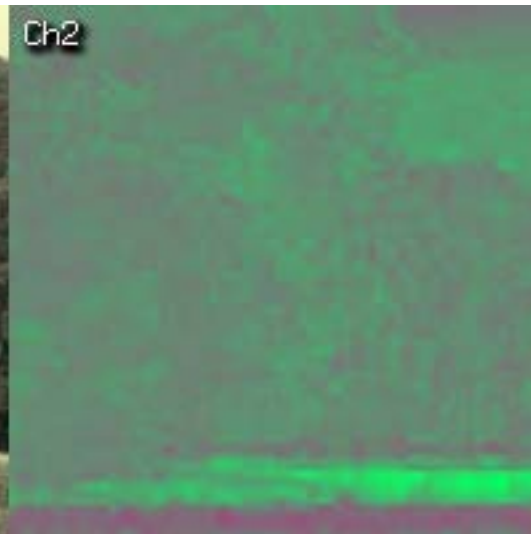


Component Intensitat



Què és el PCA?

Components Principals



http://axon.physik.uni-bremen.de/online_calc/pca/

Què és el PCA?

- Tenim un núvol de punts en un espai de dimensió elevada
- Hem de trobar un pla (primer pla factorial) sobre el que projectem el núvol de punts.
- La configuració obtinguda ha de ser el més semblant possible a la configuració del núvol original. Com?



Fent que globalment, les distàncies entre punts projectats siguin el mes semblants possible a les distàncies reals entre els punts.

Què és el PCA?

Hem de trobar un pla en que el conjunt de les distàncies $d_F(i, i')$ mesurades en el pla factorial per totes les parelles de punts, s'aproximi a les distàncies $d_X(i, i')$ reals mesurades en l'espai original.

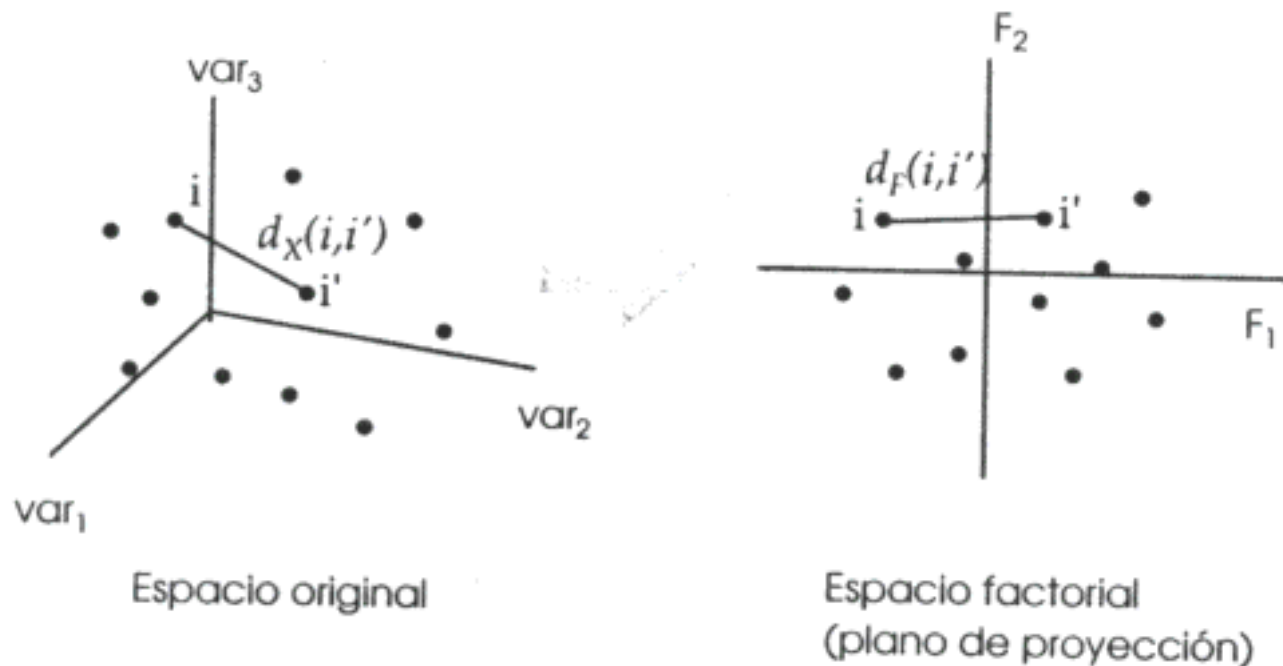
Com definim la distància?

Distància euclidiana entre punts:

$$d^2(i, i') = \sum_{j=1}^p (x_{ij} - x_{i'j})^2$$

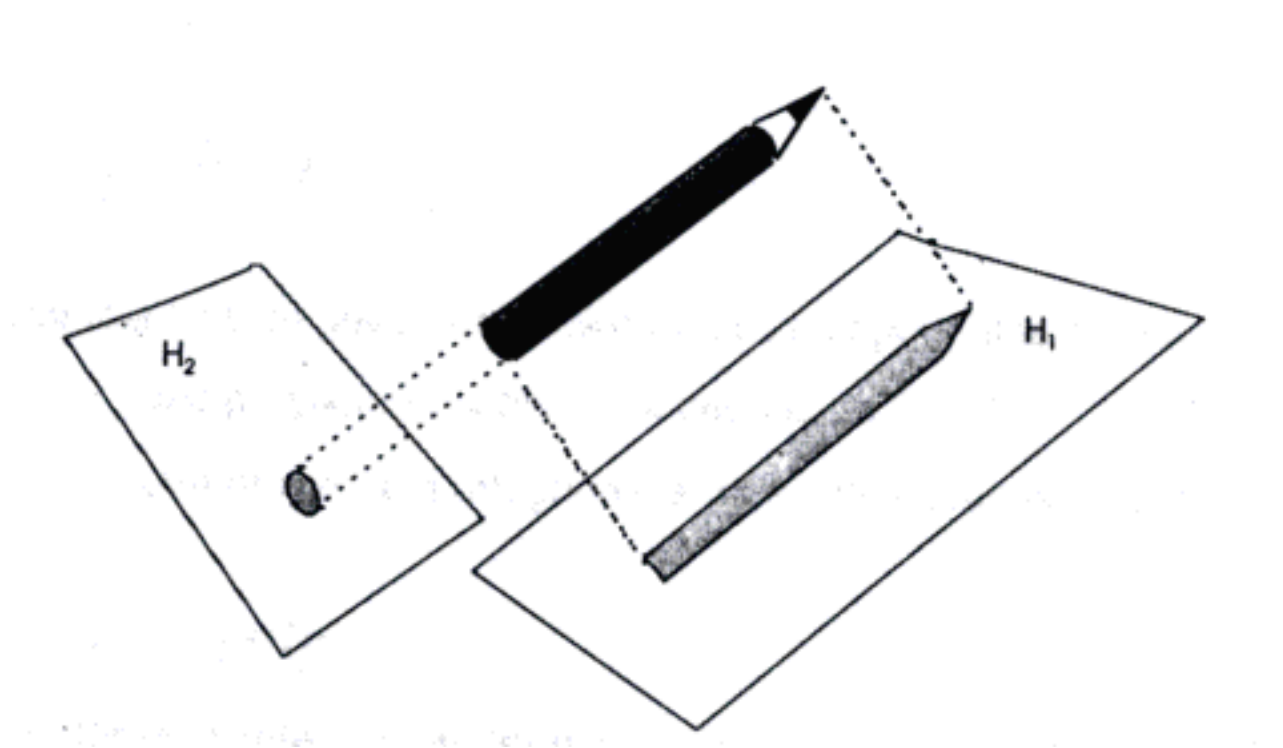
Què és el PCA?

Representació del núvol de punts en el primer pla factorial



Què és el PCA?

- Suposem que el nostre núvol de punts forma un llapis en l'espai original. Quin pla triem?



Què és el PCA?

- Triarem el pla que maximitza la dispersió dels punts projectats

$$\text{Max}_H \sum_i \sum_{i'} d_H^2(i, i')$$

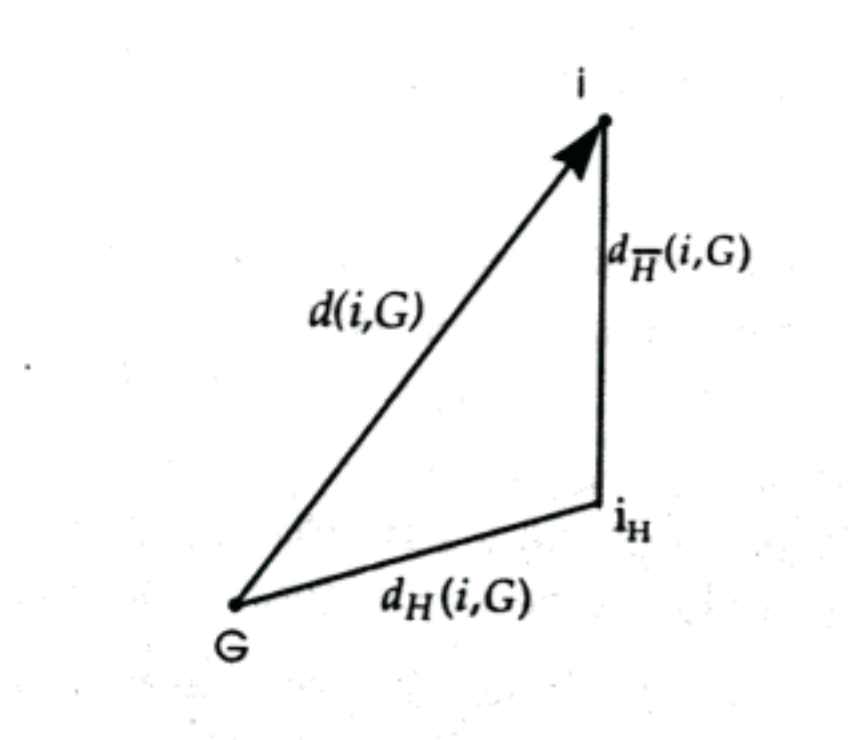
on H és el subespai de projecció.

- La suma de quadrats de les distàncies entre totes les parelles de punts es pot expressar com les distàncies de cada punt al centre de gravetat del núvol.

$$\text{Max}_H \sum_i \sum_{i'} d_H^2(i, i') = \text{Max}_H \sum_i d_H^2(i, G)$$

Què és el PCA?

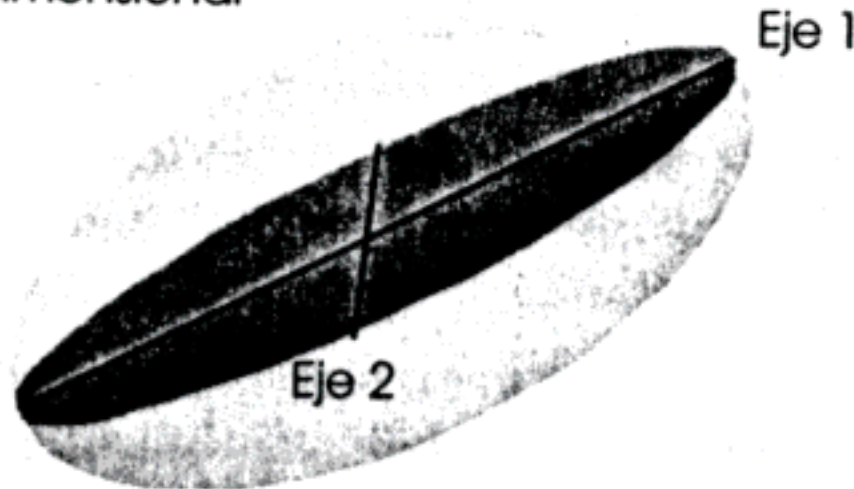
- Descomposició de la distància d'un punt al centre de Gravetat del núvol.



Implementació del PCA

- Primer busquem un subespai de dimensió 1 (una recta que coincideix amb la direcció de major allargament del núvol)

nube multidimensional
 R^p



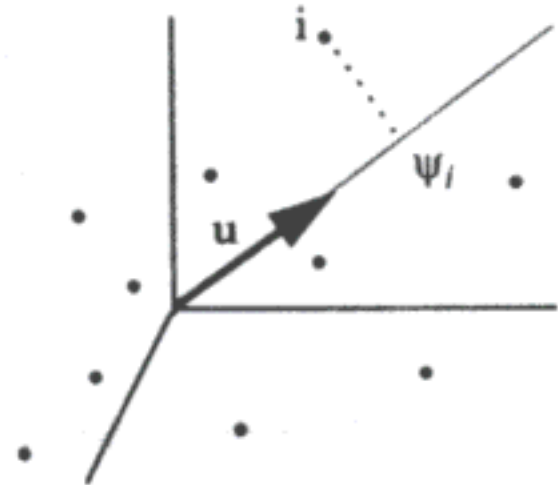
- A continuació busquem un subespai de dim. 2 (pla).

Com implementar el PCA?

- Una recta es defineix un vector de mòdul unitat $\mathbf{u}$.
- Donada una imatge color $\mathbf{X}$, calculem la seva matriu de covariança.
- Obtenim una matriu quadrada 3×3 .
- Diagonalitzem aquesta matriu.
- El vector unitari $\mathbf{u}$, és el *vector propi* associat al *valor propi* més gran, obtingut en la diagonalització de la matriu.
- La direcció ortogonal a $\mathbf{u}$, correspon al vector propi associat al segon valor propi obtingut al diagonalitzar la matriu.

Què és el PCA?

- Projectió d'un punt sobre la direcció definida pel vector unitari $\mathbf{u}$.



Què és el PCA?

- Normalment es treballa amb les variables normalitzades.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma(x_{ij})}$$