Dimensionality Reduction

LaPI
LABORATORIO AVANZADO DE PROCESAMIENTO DE IMÁGENES

1

Data Visualization

Example:

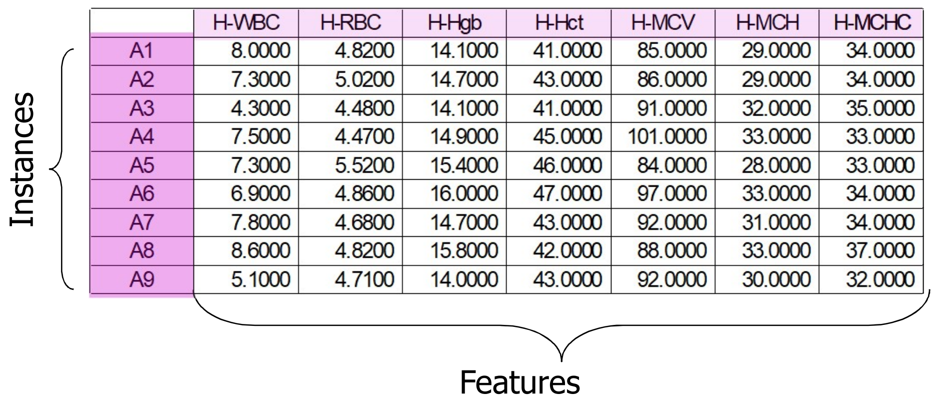
- Given 53 blood and urine samples (features) from 65 people.
- How can we visualize the measurements?

2

2

Data Visualization

- Matrix format (65x53)


Instances
Features

	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000

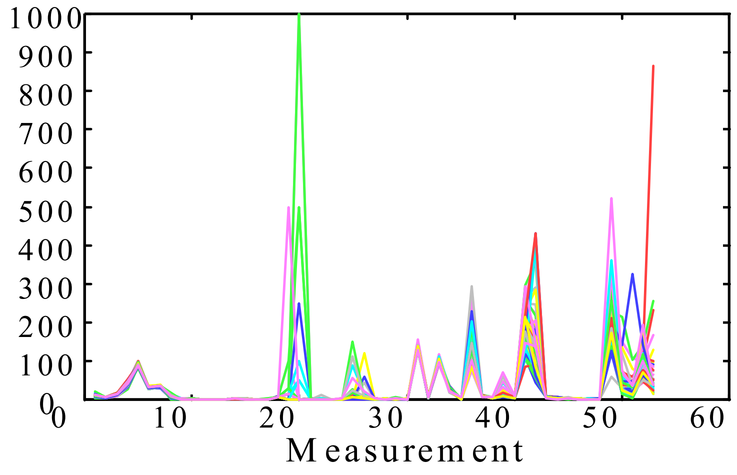
Difficult to see the correlations between the features...

3

3

Data Visualization

- Spectral format (65 pictures, one for each person)



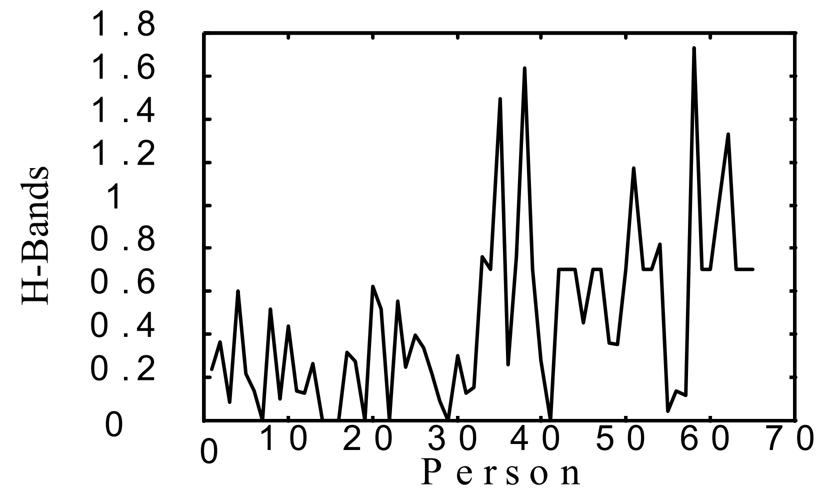
Difficult to compare the different patients...

4

4

Data Visualization

- Spectral format (53 pictures, one for each feature)

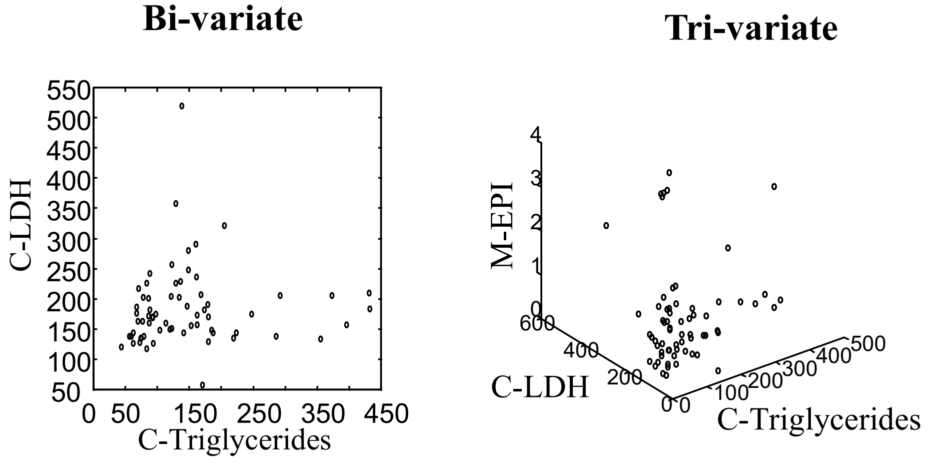


Difficult to see the correlations between the features...

5

5

Data Visualization



How can we visualize the other variables???

… difficult to see in 4 or higher dimensional spaces...

6

6

Data Visualization

- Is there a representation better than the coordinate axes?
- Is it really necessary to show all the 53 dimensions?
  - … what if there are strong correlations between the features?
- How could we find the *smallest* subspace of the 53-D space that keeps the *most information* about the original data?
- A solution: **Principal Component Analysis**

7

7

Problema

- **Problema:** ¿Podemos describir la "información" contenida en estos datos mediante algún conjunto de variables menor que el de variables originales?
- **Idea:** Si una variable es función de otras, contiene información redundante.
- Por tanto, si las p variables observadas están **fuertemente correlacionadas**, será posible sustituirlas por menos variables sin gran pérdida de "información".

8

8

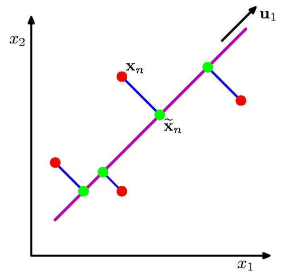
Análisis de componentes principales

- Método que se encarga de ver obtener los components principales de un dataset mediante técnicas de factorización matricial.
- Descomponer una matriz en un producto de submatrices.
- Los componentes principales son aquellas direcciones en las cuales la varianza de los elementos del dataset es máxima (los elementos están mas separados los unos de los otros)
- Al descomponer los N componentes principales podemos reducir la dimensionalidad conservando la mayor parte de la información (la varianza) contenida en el dataset original

Esta reducción de la dimensión va a permitir:

- Simplificar análisis posteriores, que se harán a partir de un menor número de variables que el original.
- Una representación grafica de los individuos en dimensión reducida (generalmente, 1 o 2).
- Examinar e interpretar las relaciones entre las variables observadas.

Principal Component Analysis



PCA:

Orthogonal projection of data onto lower-dimension linear space that...

- maximizes variance of projected data (purple line)
- minimizes mean squared distance between
  - data point and
  - projections (sum of blue lines)

11

11

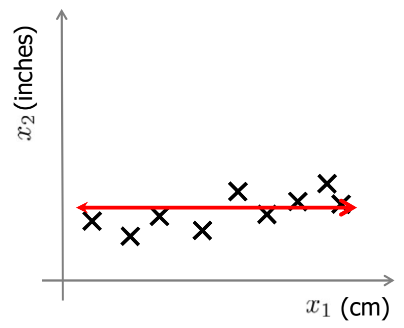
The Principal Components

- **Vectors** originating from the center of mass
- Principal component #1 points in the direction of the **largest variance**.
- Each subsequent principal component…
  - is **orthogonal** to the previous ones, and
  - points in the directions of the **largest variance of the residual subspace**

12

12

Data Compression



Reduce data from

2D to 1D

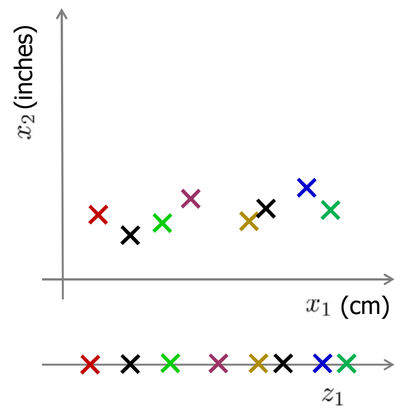
$$\frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

$$\boxed{\bar{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i}$$

13

13

Data Compression



Se pierde la información en x1

Reduce data from

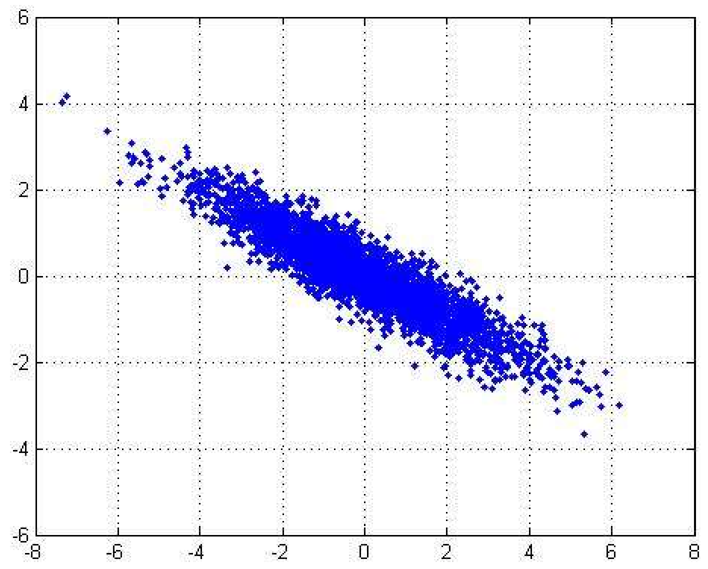
2D to 1D

$$x^{(1)} \rightarrow z^{(1)}$$
$$x^{(2)} \rightarrow z^{(2)}$$
$$\vdots$$
$$x^{(m)} \rightarrow z^{(m)}$$

14

14

2D Gaussian dataset

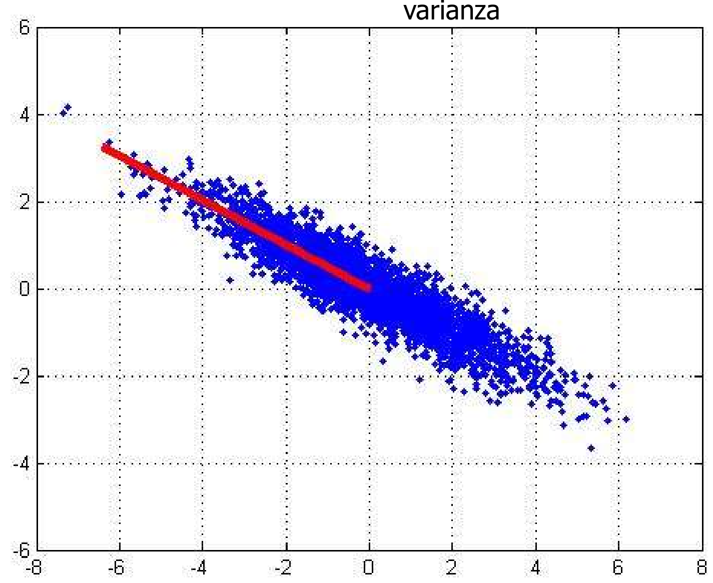


15

15

1st PCA axis

Componente principal 1 mayor varianza

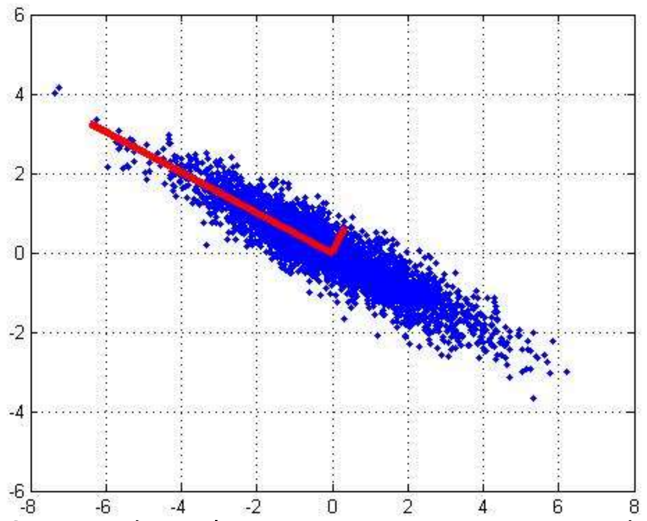


16

16

2nd PCA axis

Componente principal2,
2º en varianza

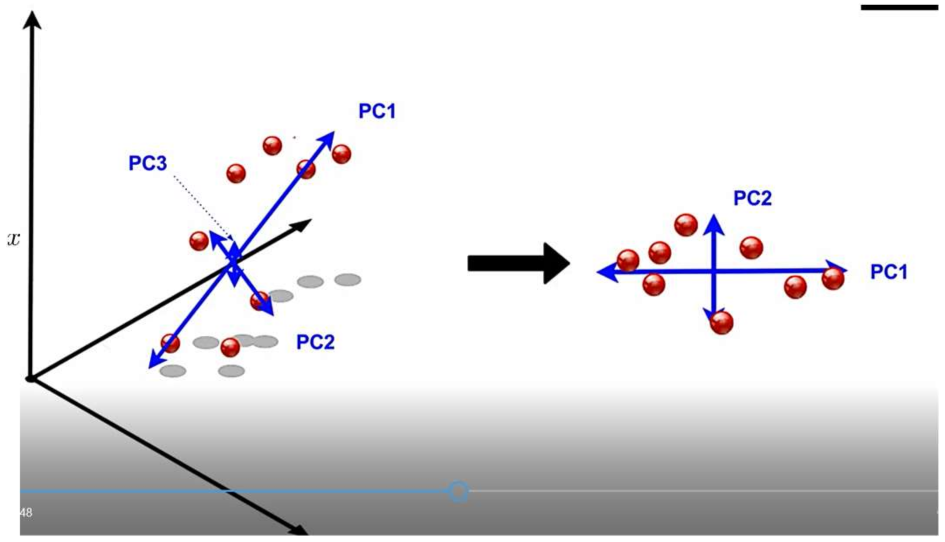


Si me quedo con los primeros componentes me quedo con la mayor parte de la varianza (energía) de los datos

17

17

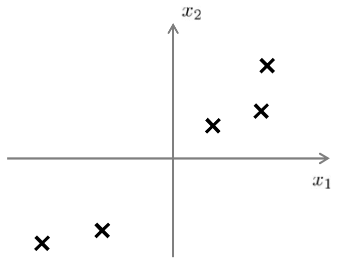
$3D \to 2D$

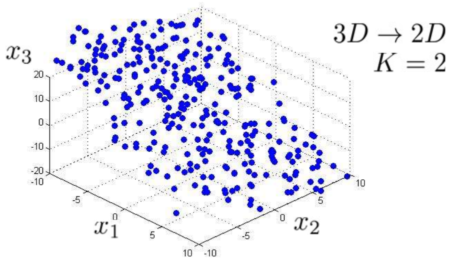


18

18

Principal Component Analysis (PCA) problem formulation





Reduce from 2-dimension to 1-dimension: Find a direction (a vector $u^{(1)} \in \mathbb{R}^n$) onto which to project the data so as to minimize the projection error.

Reduce from n-dimension to k-dimension: Find k vectors $u^{(1)}, u^{(2)}, \ldots, u^{(k)}$ onto which to project the data, so as to minimize the projection error.

19

19

Variance and Covariance

$$\frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T$$

- Variance and Covariance:
  - Measure of the "spread" of a set of points around their center of mass(mean)
- Variance:
  - Measure of the deviation from the mean for points in one dimension
- Covariance:
  - Measure of how much each of the dimensions vary from the mean with **respect to each other**

- **Covariance is measured between two dimensions**
- **Covariance sees if there is a relation between two dimensions**
- **Covariance between one dimension is the variance**

20

20

Método PCA

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T$$

Para calcular los componentes principales de una matriz hay que descomponer su matriz de covarianza en sus valores propios. La matriz de covarianza de una matriz es otra matriz donde la diagonal es la varianza de cada una de las variables y el resto de elementos son las covarianzas entre variables. Expresado de otra forma, la matriz de covarianza (Σ):

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

$$\mu_i = \mathrm{E}(X_i)$$

Con E siendo la media aritmética. Antes de calcular la matriz de covarianza se centran los datos restandoles su media (media de cada variable).

21

Usando métodos de factorización de matrices (descomposición espectral, svd) podemos descomponer la matriz de covarianza en:

$$\Sigma = Q \Lambda Q^T$$

Q Matriz ortogonal cuyas columnas son los componentes principales de **Σ** *(matriz ortogonal es aquella que multiplicada por su transpuesta da la matriz identidad)*

Λ Matriz diagonal cuyos valores son las varianzas que explica cada componente principal

Una vez tenemos los componentes principales, seleccionamos los que N mayores que queramos

*svd= Singular value decomposition
Al generar la matriz diagonal los componentes se ordenan de mayor a menor varianza

22

PCA algorithm II

PCA algorithm($\mathbf{X}$, k): top k eigenvalues/eigenvectors

```
    % X = N × m data matrix,
    % … each data point $\mathbf{x}_i$ = column vector, i=1..m
• $\underline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i$
• $\mathbf{X}$ ← subtract mean $\underline{\mathbf{x}}$ from each column vector $\mathbf{x}_i$ in $\underline{\mathbf{X}}$
• $\Sigma \leftarrow \mathbf{X}\mathbf{X}^T$   … covariance matrix of $\mathbf{X}$
• $\{ \lambda_i, \mathbf{u}_i \}_{i=1..N}$ = eigenvectors/eigenvalues of $\Sigma$
    ... $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$
• Return $\{ \lambda_i, \mathbf{u}_i \}_{i=1..k}$
    % top $k$ principle components
```

Principal Component Analysis (PCA)

- Stack N datapoints column-wise in matrix **X’**
- De-mean the dataset $\mathbf{X}=\mathbf{X}'-\mathbf{X}_m$
- Obtain new orthogonal basis from **X**:
  - Via SVD —> $[\mathbf{U},\mathbf{D},\mathbf{V}^T]$=SVD($\mathbf{X}^T$/sqrt(N-1))
  - Via covariance matrix (eigenvectors) $\mathbf{C}=\mathbf{X}\mathbf{X}^T=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T=\mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T$
- Principal components: columns of $\mathbf{V}^T$
- Variance on the diagonal of $\mathbf{D}^{\wedge 2}$
- Now one can represent data in the new subspace as $\mathbf{Y}=\mathbf{X}^T\,\mathbf{V}$
- It’s possible to
  1. **Reduce** the dimensionality of the data and
  2. **Synthesise** high dimensional data that is coherent with modes of variation discovered by PCA

Principal Component Analysis

Goal: Find r-dim projection that best preserves variance

1. Compute mean vector μ and covariance matrix Σ of original points
2. Compute eigenvectors and eigenvalues of Σ
3. Select top r eigenvectors
4. Project points onto subspace spanned by them:

$$y = A(x - \mu)$$

where y is the new point, x is the old one, and the rows of A are the eigenvectors

25

25

PCA Applications

- Data Visualization
- Data Compression
- Noise Reduction
- Data Classification
- Data Recognition
- ...

26

26

Image Compression

27

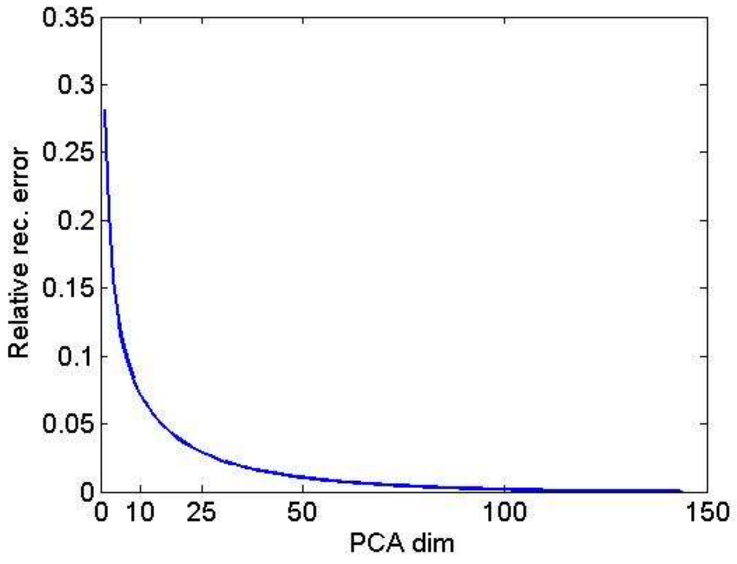
Original Image

- Divide the original 372x492 image into patches:
  - Each patch is an instance that contains 12x12 pixels on a grid
- View each as a 144-D vector

28

28

L_2 error and PCA dim



29

29

PCA compression: 144D → 60D

30

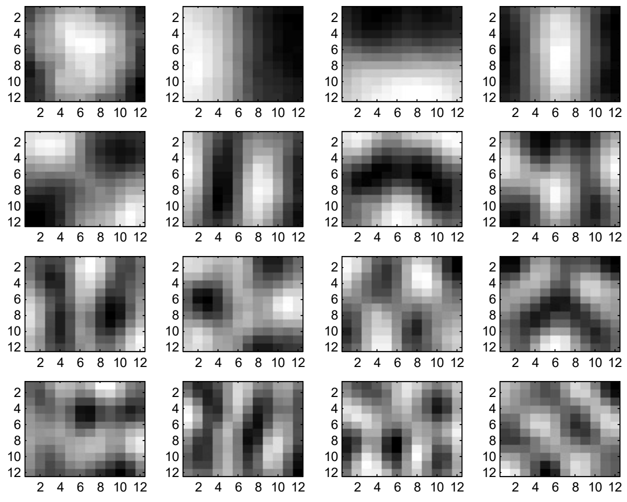
30

PCA compression: 144D → 16D

31

31

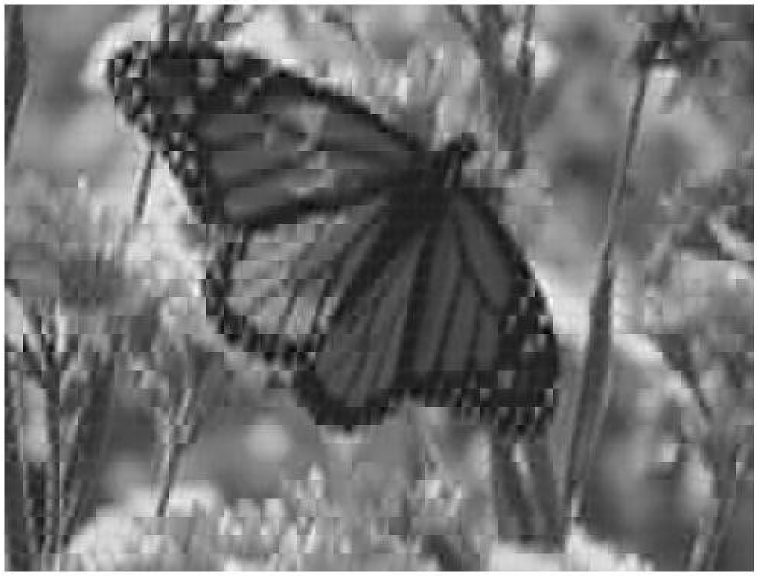
16 most important eigenvectors



32

32

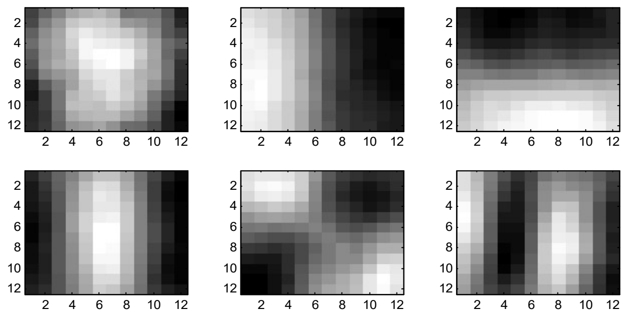
PCA compression: 144D → 6D



33

33

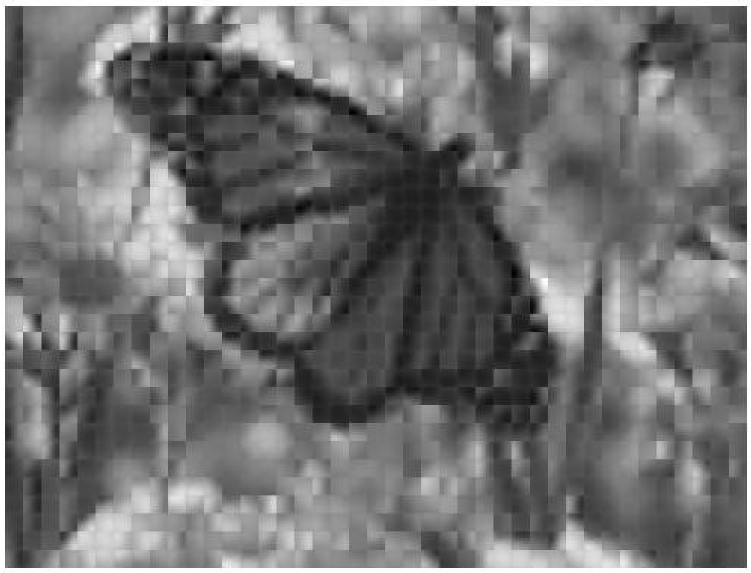
6 most important eigenvectors



34

34

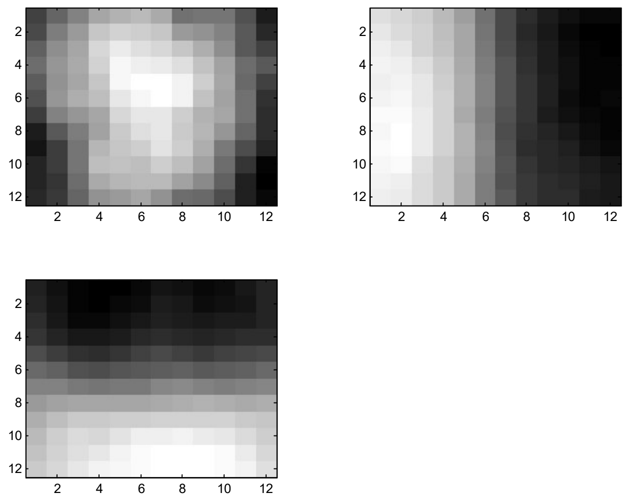
PCA compression: 144D → 3D



35

35

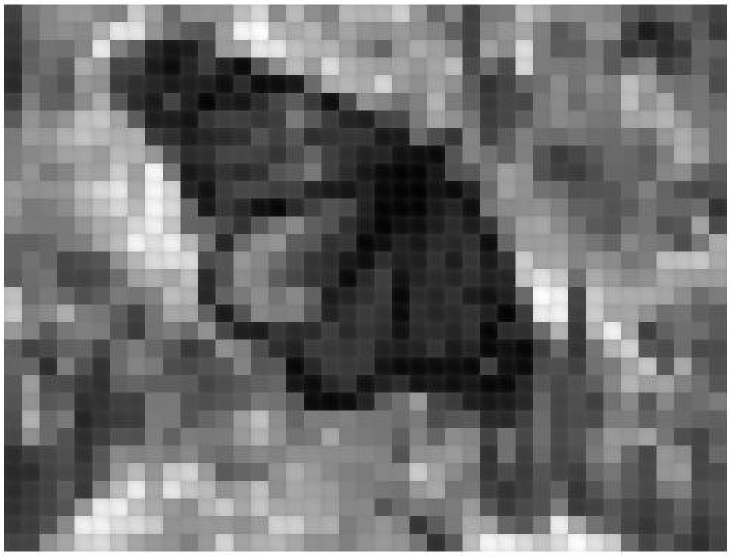
3 most important eigenvectors



36

36

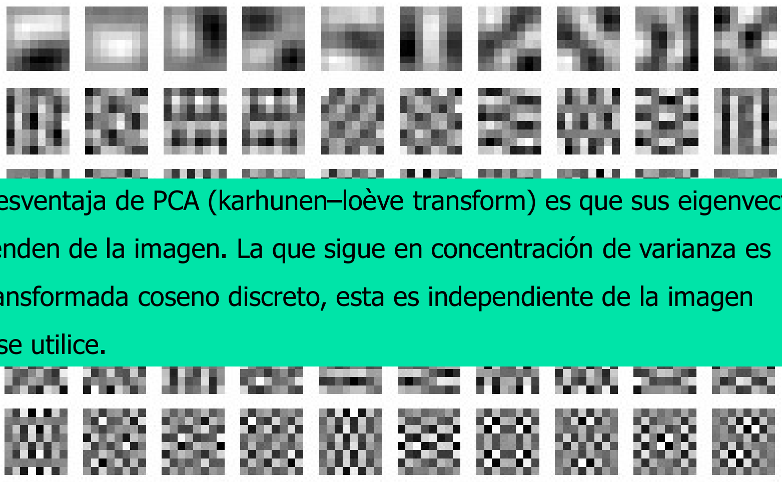
PCA compression: 144D → 1D



37

37

60 most important eigenvectors



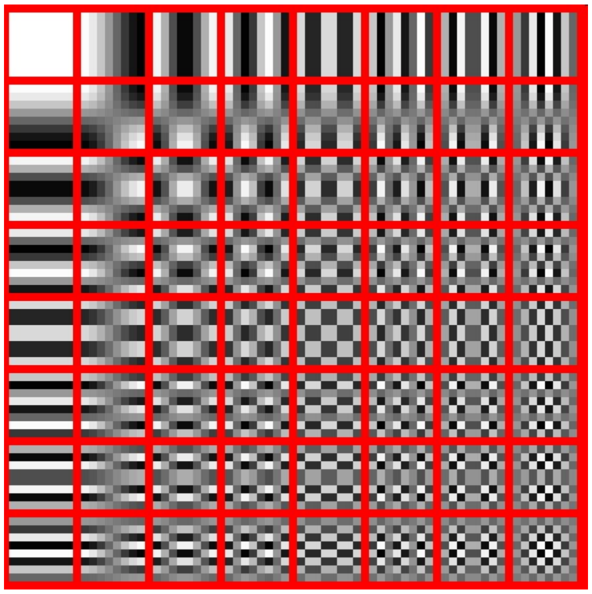
La desventaja de PCA (karhunen–loève transform) es que sus eigenvectores dependen de la imagen. La que sigue en concentración de varianza es la transformada coseno discreto, esta es independiente de la imagen que se utilice.

Looks like the discrete cosine bases of JPG!...

38

38

2D Discrete Cosine Basis

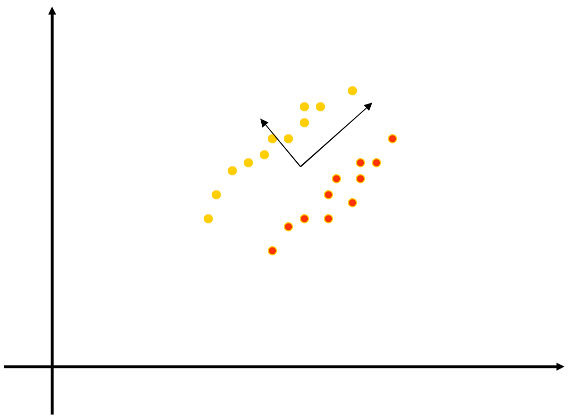


http://en.wikipedia.org/wiki/Discrete_cosine_transform

39

39

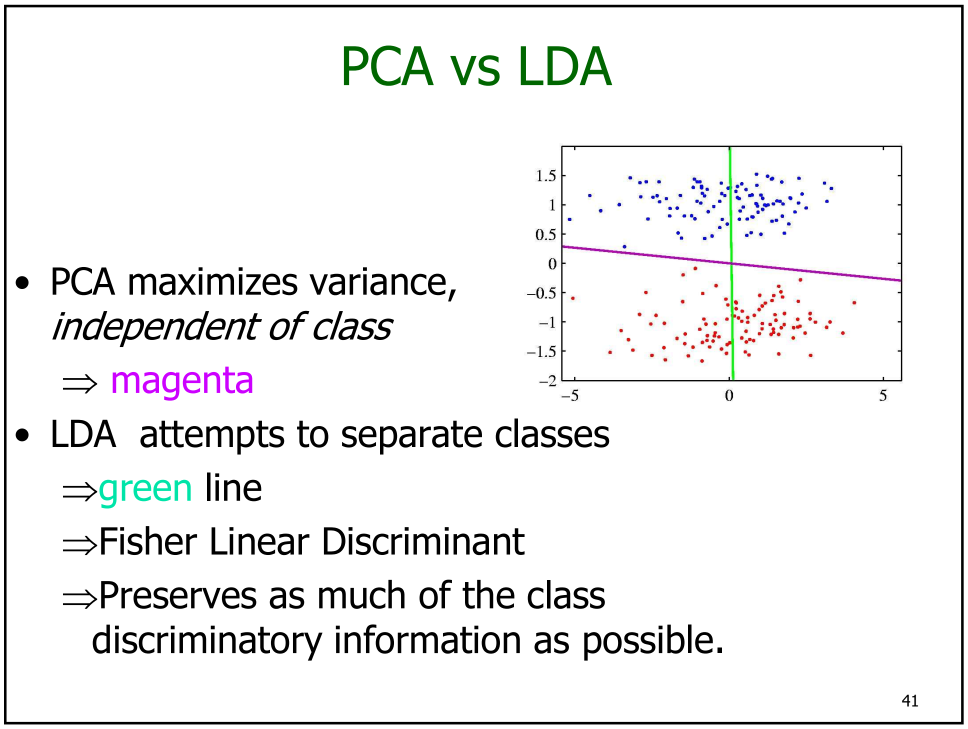
PCA, a Problematic Data Set



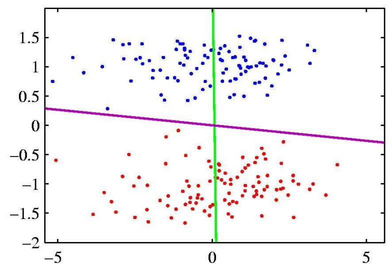
PCA doesn't know labels!

40

40



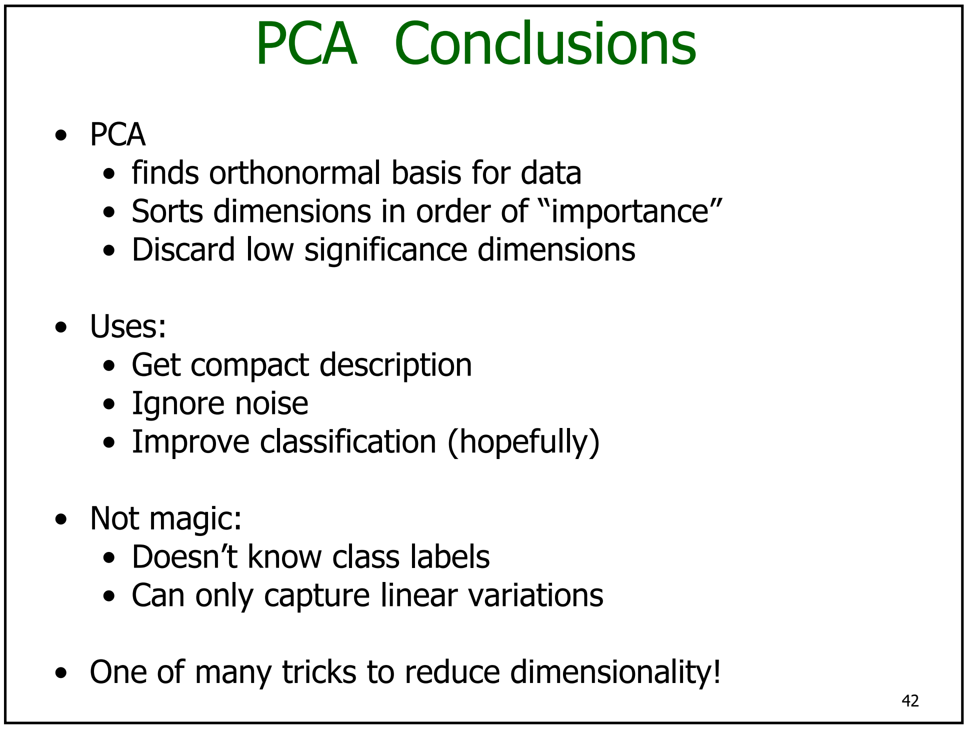
PCA vs LDA



- PCA maximizes variance, *independent of class*
  - ⇒ magenta
- LDA attempts to separate classes
  - ⇒green line
  - ⇒Fisher Linear Discriminant
  - ⇒Preserves as much of the class discriminatory information as possible.

41

41



PCA Conclusions

- PCA
  - finds orthonormal basis for data
  - Sorts dimensions in order of "importance"
  - Discard low significance dimensions
- Uses:
  - Get compact description
  - Ignore noise
  - Improve classification (hopefully)
- Not magic:
  - Doesn't know class labels
  - Can only capture linear variations
- One of many tricks to reduce dimensionality!

42

42

Dimensionality reduction

- PCA (Principal Component Analysis) :
  - Find projection that maximize the variance (linear)
- ICA (Independent Component Analysis):
  - Very similar to PCA except that it assumes non-Guassian features
- Multidimensional Scaling:
  - Find projection that best preserves inter-point distances
- LDA(Linear Discriminant Analysis):
  - Maximizing the component axes for class-separation (linear)
- t-SNE (non-parametric/ nonlinear)
  - A pesar de las muy buenas propiedades que tiene el *PCA*, sufre de algunas limitaciones
    - Solo tiene en cuenta combinaciones lineales de las variables originales.
    - En determinados escenarios, el no poder considerar otro tipo de combinaciones supone perder mucha información.
  - En 2008, *Geoffrey Hinton* y *Laurens van der Maaten* publicaron el método no lineal *t-distributed stochastic neighbor embedding (t-SNE)* que consigue superar esta limitación y por lo tanto supera al *PCA* en muchos escenarios.

43

43

- https://sandipanweb.wordpress.com/2018/01/06/eigenfaces-and-a-simple-face-detector-with-pca-svd-in-python/

44

44