MACHINE LEARNING (ML-15)

Dr. NEERAJ GUPTA, Department of CEA, GLA University, Mathura

AGENDA

- Principal Component Ananlysis (PCA)

MOTIVATION

Clustering

- One way to summarize a complex real-valued data point with a single categorical variable

Dimensionality reduction

- Another way to simplify complex high-dimensional data
- Summarize data with a lower dimensional real valued vector

MOTIVATION

Clustering

- One way to summarize a complex real-valued data point with a single categorical variable

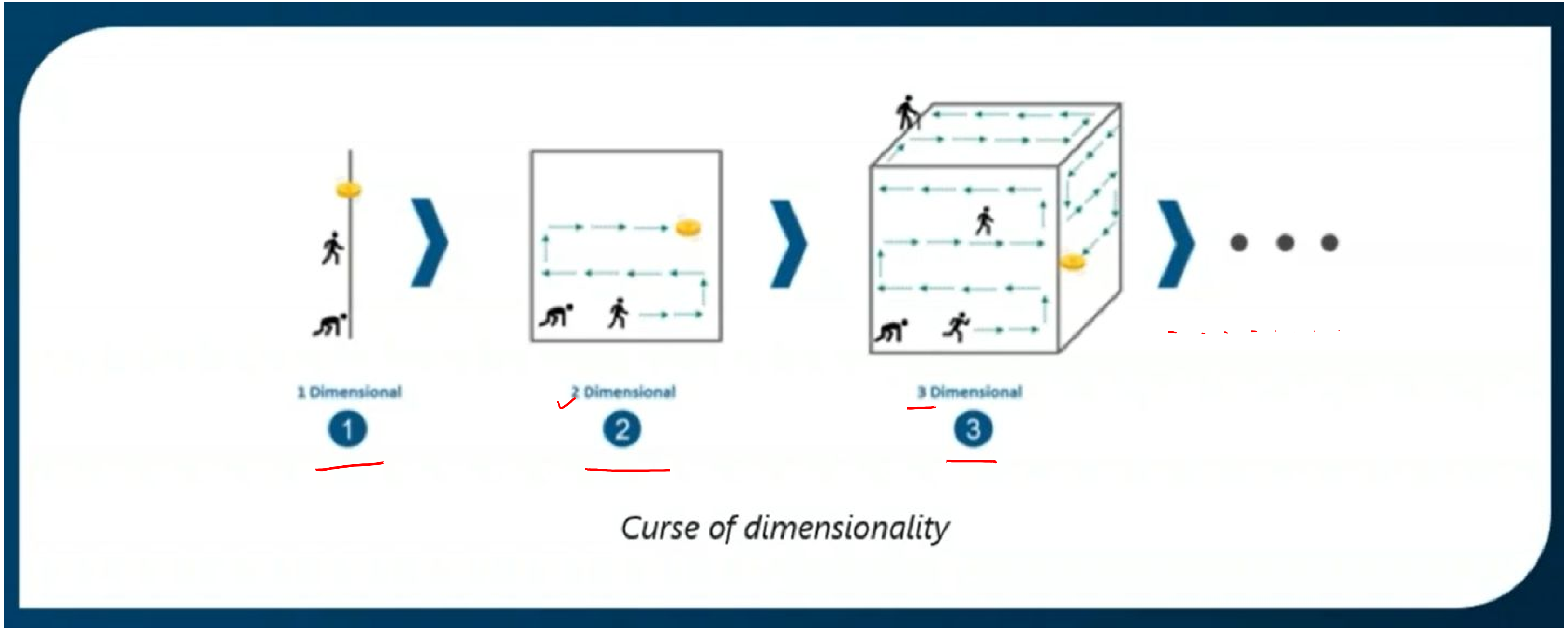
Dimensionality reduction

- Another way to simplify complex high-dimensional data
- Summarize data with a lower dimentional real valued vector

$r < d$

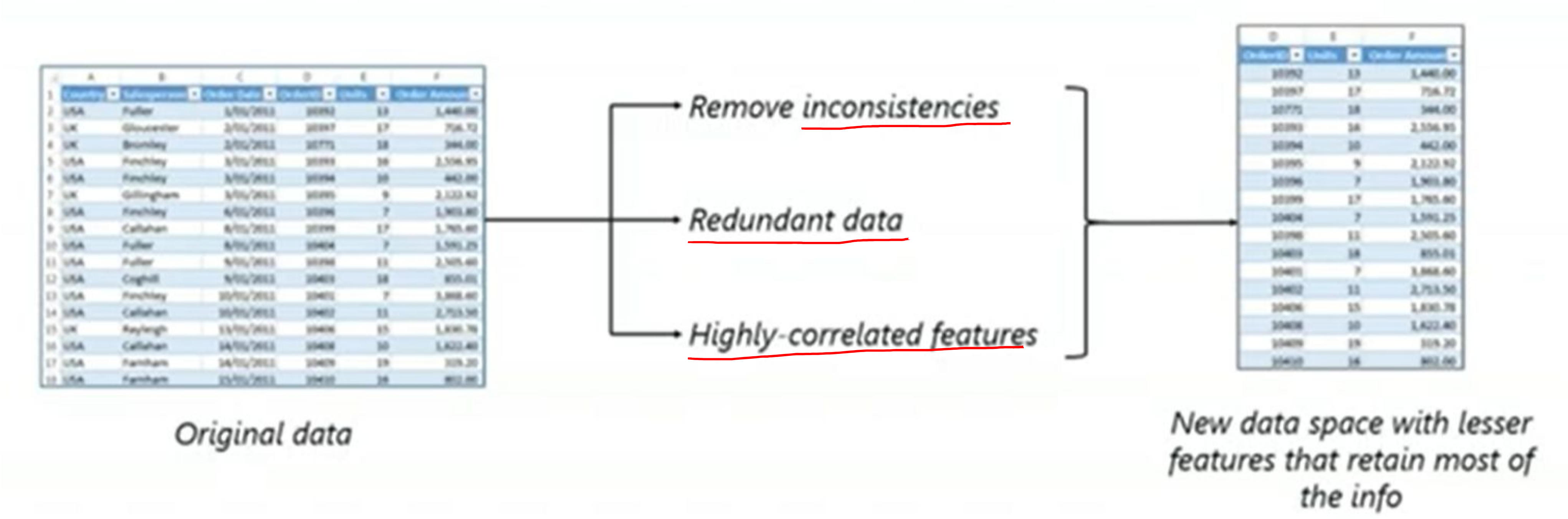
- Given data points in *d* dimensions
- Convert them to data points in $r<d$ dimensions
- With minimal loss of information

NEED FOR PCA

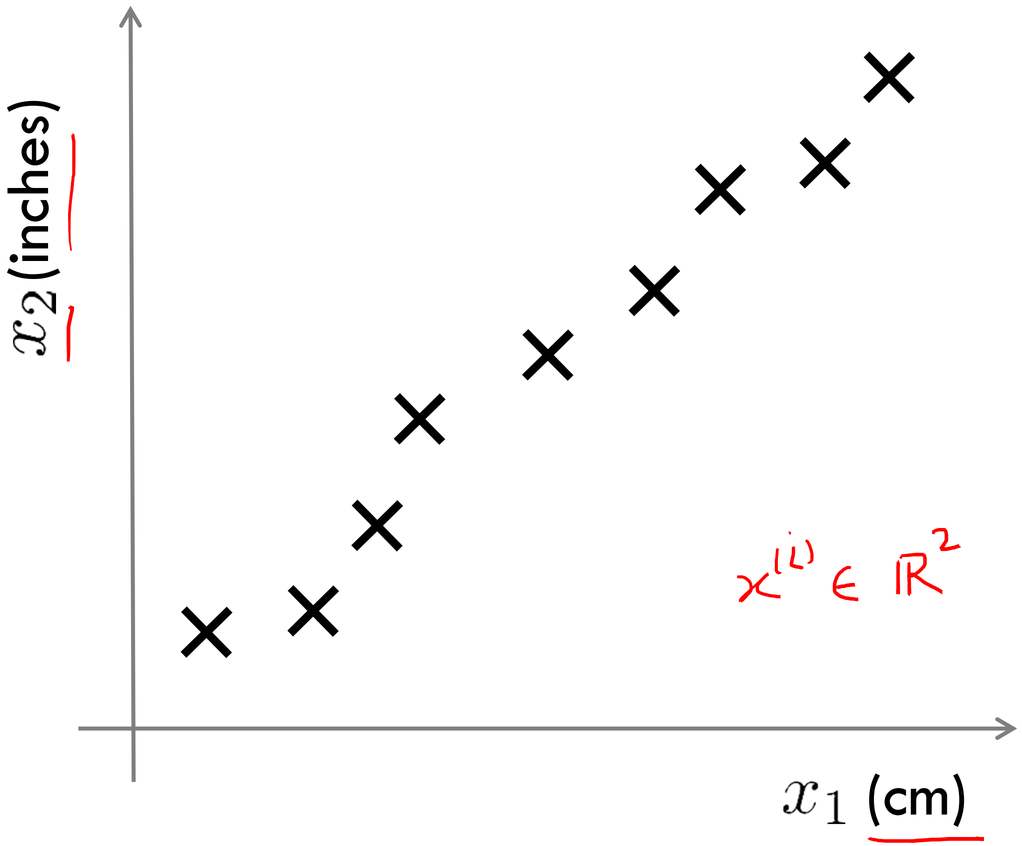


High dimension data is extremely complex to process due to inconsistencies in the feature which increase the computation time.

NEED FOR PCA


Remove inconsistencies
Redundant data
Highly-correlated features
Original data
New data space with lesser features that retain most of the info

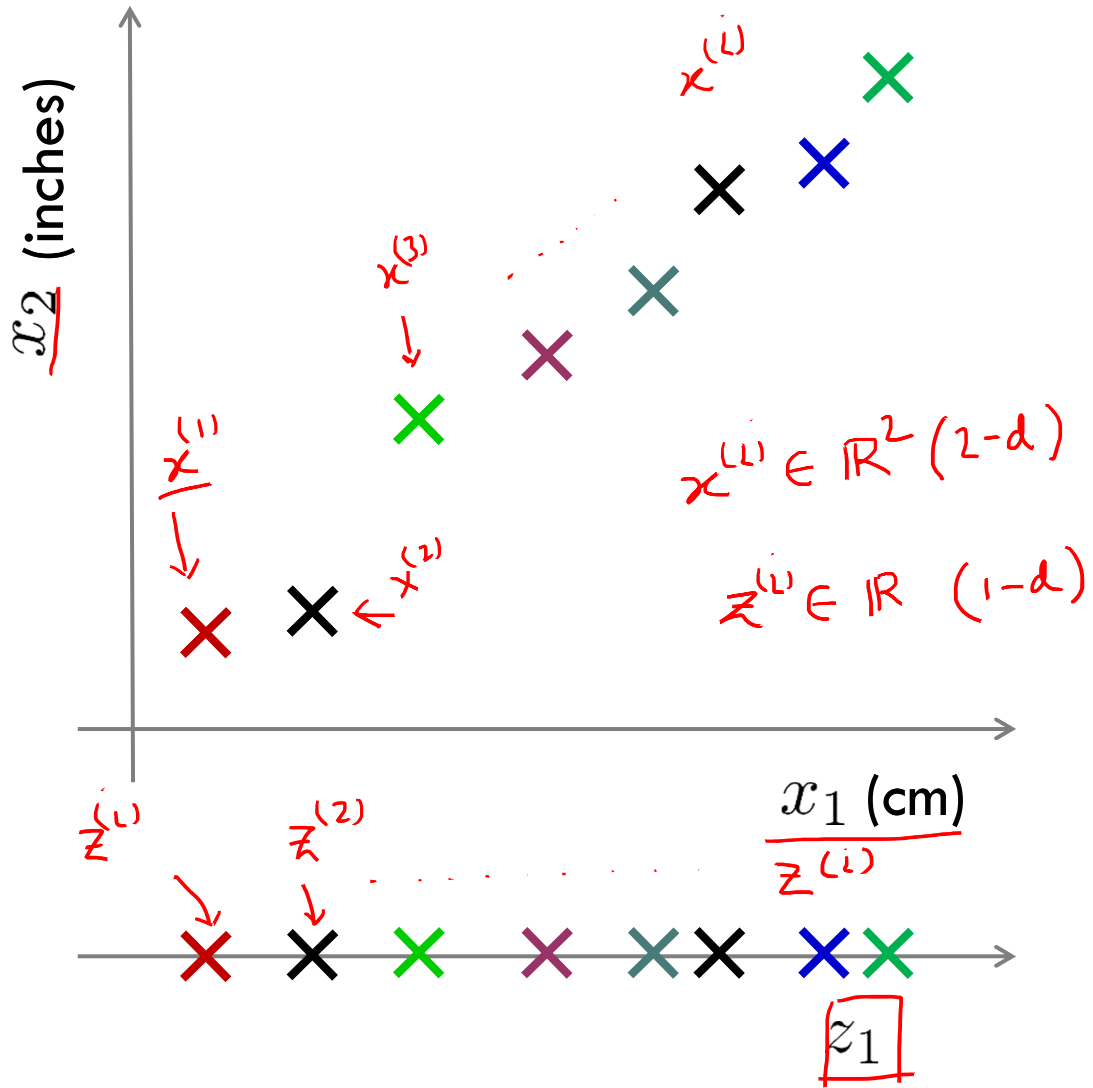
Data Compression



Reduce data from 2D to 1D

Andrew Ng

Data Compression

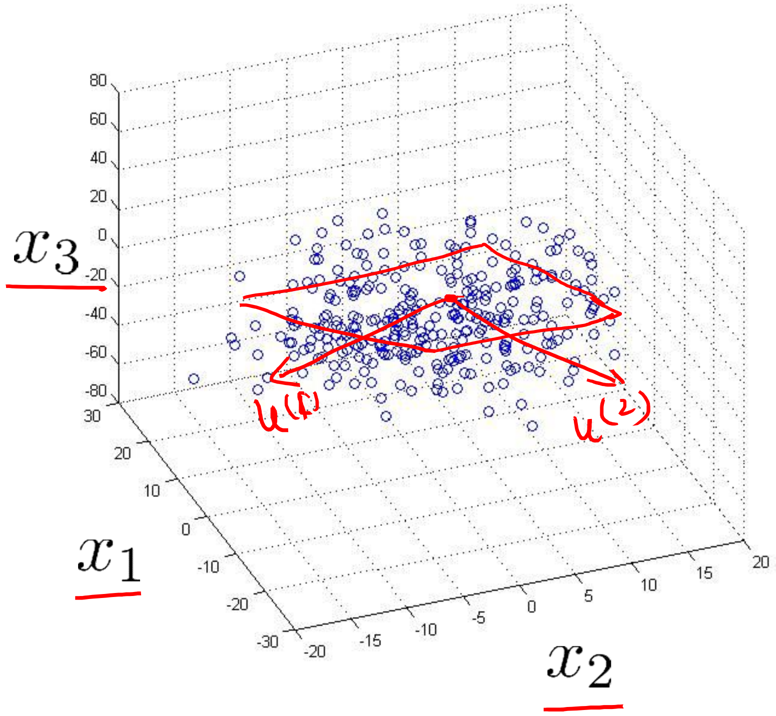


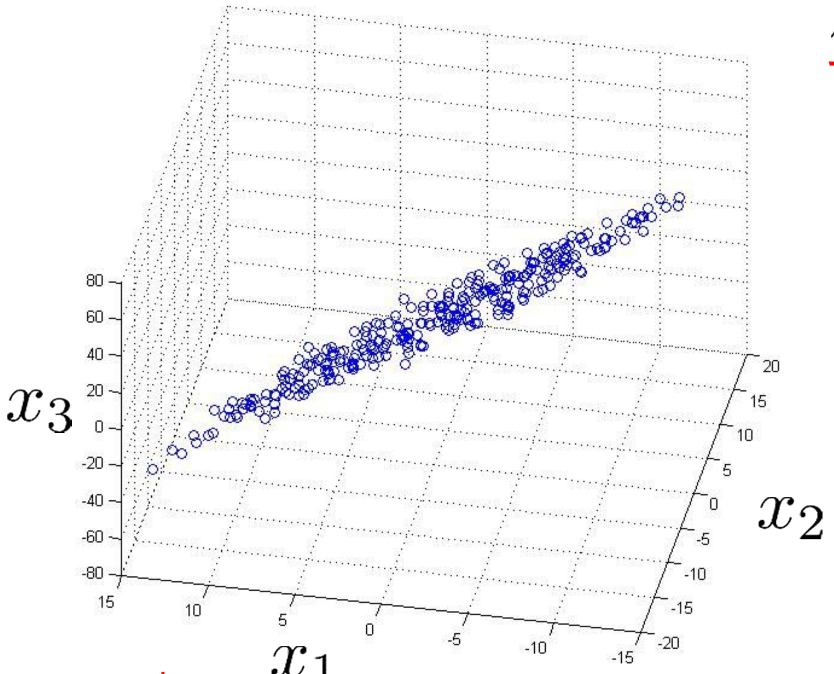
Reduce data from 2D to 1D

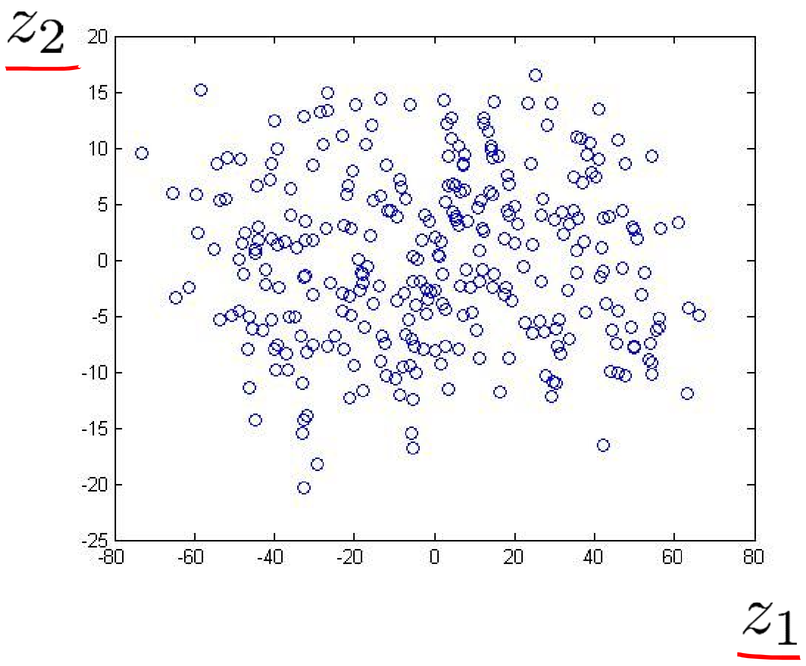
$$x^{(1)} \rightarrow z^{(1)}$$
$$x^{(2)} \rightarrow z^{(2)}$$
$$\vdots$$
$$x^{(m)} \rightarrow z^{(m)}$$

Data Compression

Reduce data from 3D to 2D





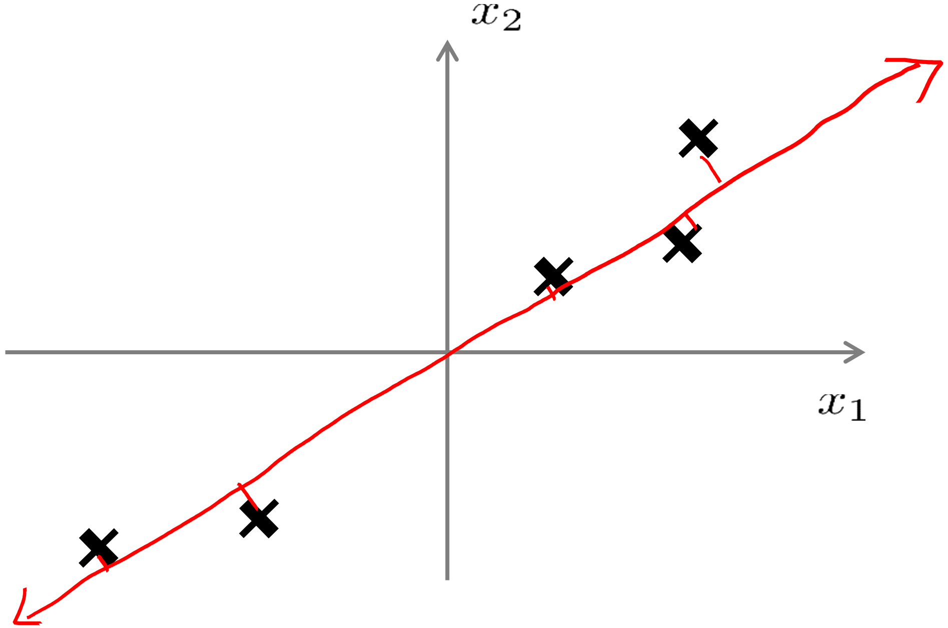


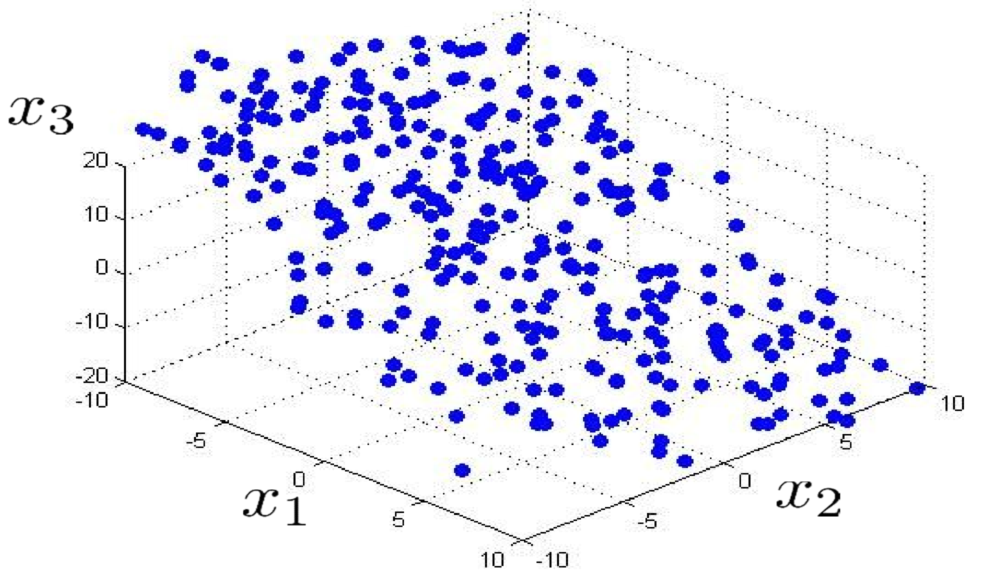
$x^{(i)} \in \mathbb{R}^3$ (3-dim) $\longrightarrow$ $z^{(i)} \in \mathbb{R}^2$

$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$

$3D \rightarrow 2D$
$K = 2$

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.

Andrew Ng

STEP BY STEP PCA

Age	Salary
$\boxed{23}$	10000
24	50000
26	60000
29	15000

$$\boxed{z = \frac{\text{value} - \text{mean}}{\text{stand. devi}}}$$

1. Standardization of the data (1)
2. Computing the covariance matrix (2)
3. Calculating the eigenvectors and eigenvalues (3)
4. Computing the Principal Components (4)
5. Reducing the dimensions of the data (5)

$f_1, f_2 \ldots f_n$

$p \times p$ (symmetric matrix). p-dimension

3-dim 3×3

2 dim 2×2

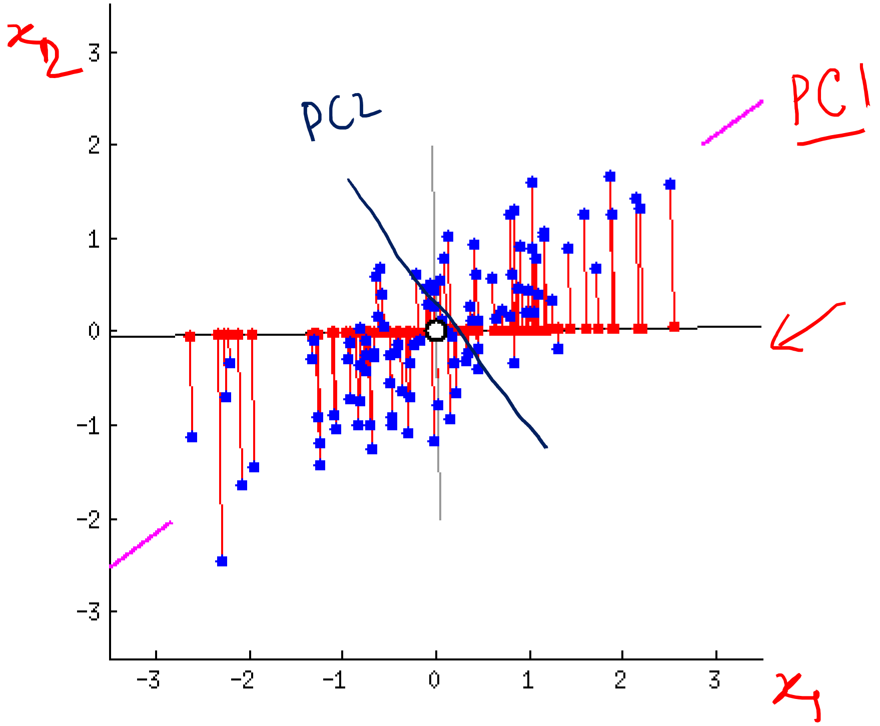
x, y

$$p = \begin{bmatrix} Cov(x,x), & Cov(x,y) \\ Cov(y,x), & Cov(y,y) \end{bmatrix}$$

+ve (Correlated)

−ve (Inversely Correlated)

Principal Components


x2
PC2
PC1
x1
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

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

COVARIANCE

Variance and Covariance:

- Measure of the “spread” of a set of points around their center of mass(mean)

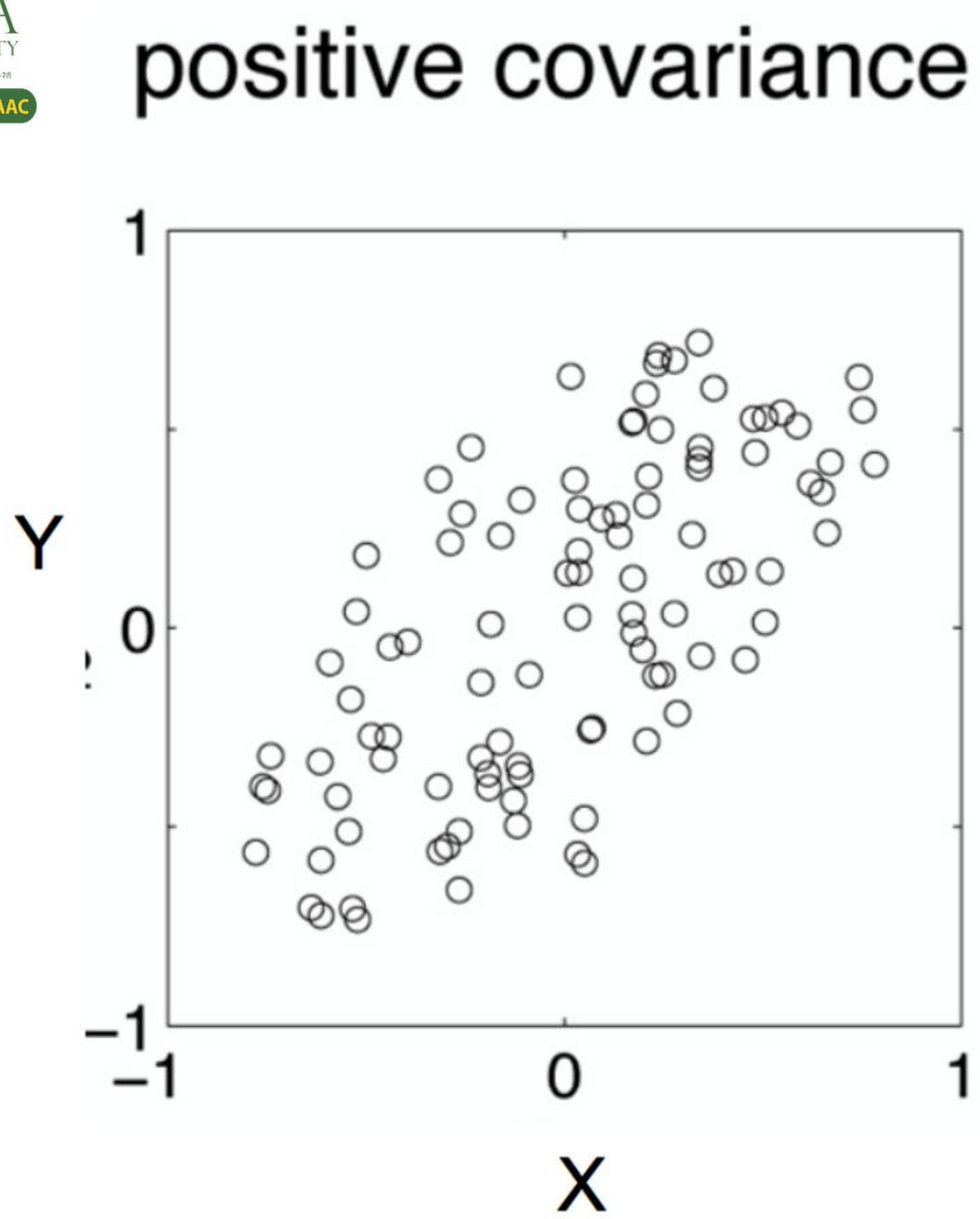
Variance:

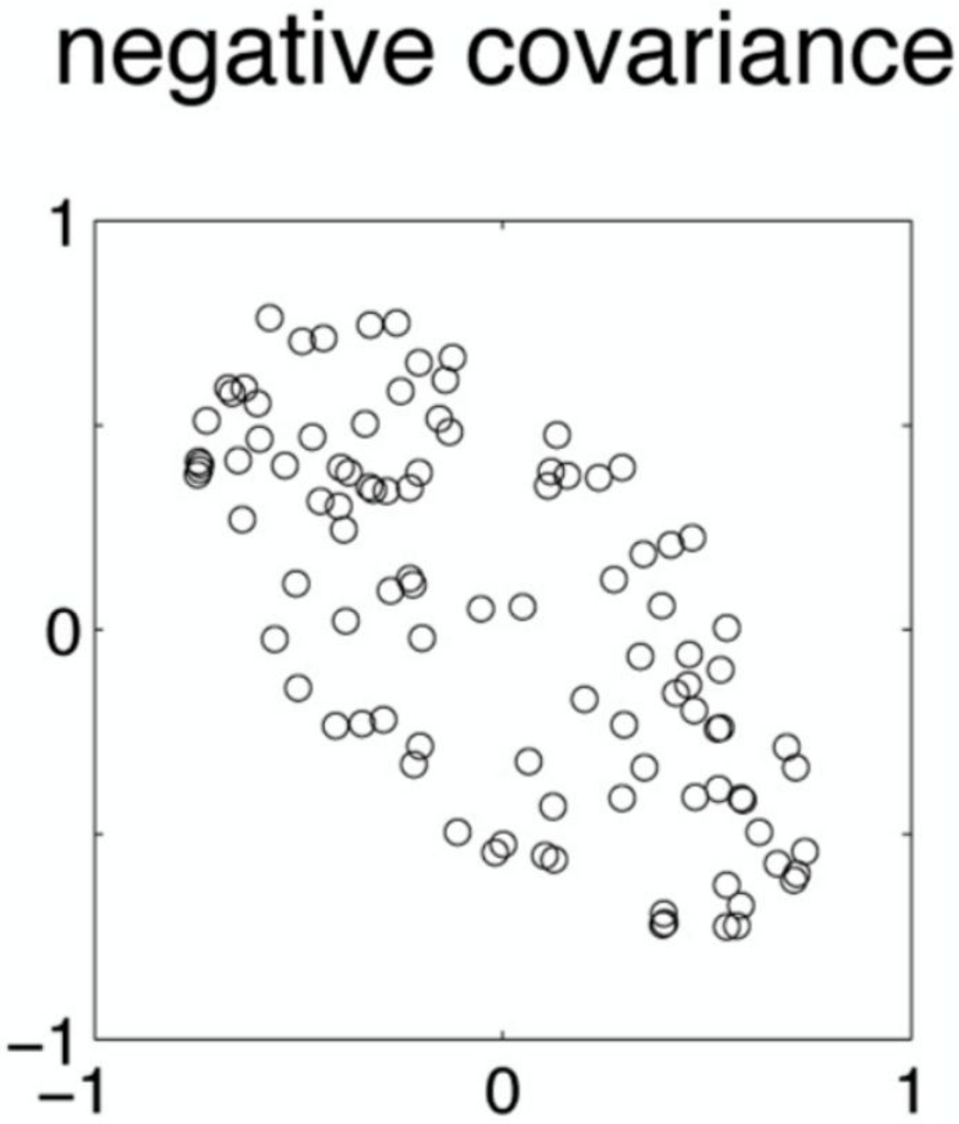
- Measure of the deviation from the mean for points in one dimension

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**





Positive: Both dimensions increase or decrease together

Negative: While one increase the other decrease

COVARIANCE

Used to find relationships between dimensions in high dimensional data sets

mean mean

$$\underline{q_{jk}} = \frac{1}{N}\sum_{i=1}^{N}(X_{ij} - \underline{E(X_j)})(X_{ik} - \underline{E(X_k)})$$

The Sample mean

EIGENVECTOR AND EIGENVALUE

$$Ax = \lambda x$$

A: Square Matirx

λ: Eigenvector or characteristic vector

X: Eigenvalue or characteristic value

- *The zero vector can not be an eigenvector*
- *The value zero can be eigenvalue*

EIGENVECTOR AND EIGENVALUE

$Ax = \lambda x$

A: Square Matirx

λ: Eigenvector or characteristic vector

X: Eigenvalue or characteristic value

Example

$$Show\ \ x = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \ is\ an\ eigenvector\ for\ A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}$$

$$Solution: Ax = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2\times 2 - 4\times 1 = 0 \\ 3\times 2 - 6\times 1 = 0 \end{bmatrix}$$

$$But\ \ for\ \ \lambda = 0,\ \ \lambda x = 0\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Thus, x\ is\ an\ eigenvector\ of\ A, and\ \lambda = 0\ is\ an\ eigenvalue.$$

EIGENVECTOR AND EIGENVALUE

$Ax = \lambda x$ → $Ax - \lambda x = 0$

$(A - \lambda I)x = 0$

If we define a new matrix B: → $\boxed{B = A - \lambda I}$

$Bx = 0$

If B has an inverse: → $x = B^{-1}0 = 0$ ✗ **BUT! an eigenvector cannot be zero!!**

x will be an eigenvector of A if and only if B does not have an inverse, or equivalently det(B)=0 :

$$\boxed{\det(A - \lambda I) = 0}$$

EIGENVECTOR AND EIGENVALUE

Example 1: Find the eigenvalues of

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$$

$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

$$|\lambda I - A| = \begin{vmatrix} \lambda - 2 & 12 \\ -1 & \lambda + 5 \end{vmatrix} = (\lambda - 2)(\lambda + 5) + 12$$

$$= \lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2)$$

Q.E

two eigenvalues: $-1, -2$

Note: The roots of the characteristic equation can be repeated. That is, $\lambda_1 = \lambda_2 = \ldots = \lambda_k$. If that happens, the eigenvalue is said to be of multiplicity k.

Example 2: Find the eigenvalues of

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$|\lambda I - A| = \begin{vmatrix} \lambda - 2 & -1 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & \lambda - 2 \end{vmatrix} = (\lambda - 2)^3 = 0$$

$\lambda = 2$ is an eigenvector of multiplicity 3.

PRINCIPAL COMPONENT ANALYSIS

Input: $\mathbf{x} \in \mathbb{R}^D: \ \mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$

Set of basis vectors: $\mathbf{u}_1, \ldots, \mathbf{u}_K$

Summarize a *D* dimensional vector *X* with *K* dimensional feature vector *h(x)*

$$h(\mathbf{x}) = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{x} \\ \mathbf{u}_2 \cdot \mathbf{x} \\ \cdots \\ \mathbf{u}_K \cdot \mathbf{x} \end{bmatrix}$$

PRINCIPAL COMPONENT ANALYSIS

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$$

Basis vectors are orthonormal

$$\mathbf{u}_i^T \mathbf{u}_j = 0$$

$$||\mathbf{u}_j|| = 1$$

New data representation *h(x)*

$$z_j = \mathbf{u}_j \cdot \mathbf{x}$$

$$h(\mathbf{x}) = [z_1, \ldots, z_K]^T$$

PRINCIPAL COMPONENT ANALYSIS

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$$

New data representation *h(x)*

$$h(\mathbf{x}) = \mathbf{U}^T \mathbf{x}$$

$$h(\mathbf{x}) = \mathbf{U}^T (\mathbf{x} - \mu_0)$$

Empirical mean of the data

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

PRACTICE PROBLEMS BASED ON PCA

Given data = { 2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8 }.

Compute the principal component using PCA Algorithm.

OR

Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8).

Compute the principal component using PCA Algorithm.

OR

Compute the principal component of following data-

CLASS 1

X = 2 , 3 , 4

Y = 1 , 5 , 3

CLASS 2

X = 5 , 6 , 7

Y = 6 , 7 , 8

PRACTICE PROBLEMS BASED ON PCA

PCA Algorithm-

The steps involved in PCA Algorithm are as follows-

Step-01: Get data. —— n

Step-02: Compute the mean vector (µ).

Step-03: Subtract mean from the given data.

Step-04: Calculate the covariance matrix.

Step-05: Calculate the Eigen vectors and Eigen values of the covariance matrix.

Step-06: Choosing components and forming a feature vector.

Step-07: Deriving the new data set. —— k

$k < n$

PRACTICE PROBLEMS BASED ON PCA

Step-01:

Get data.

The given feature vectors are-

$x_1 = (2, 1)$

$x_2 = (3, 5)$

$x_3 = (4, 3)$

$x_4 = (5, 6)$

$x_5 = (6, 7)$

$x_6 = (7, 8)$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 6 \\ 7 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

PRACTICE PROBLEMS BASED ON PCA

Step-02:

Calculate the mean vector (μ). μ

Mean vector (μ)

= ((2 + 3 + 4 + 5 + 6 + 7) / 6, (1 + 5 + 3 + 6 + 7 + 8) / 6)

= (4.5, 5)

Thus,

$$\text{Mean vector } (\mu) = \begin{bmatrix} 4.5 \\ 5 \end{bmatrix}$$

PRACTICE PROBLEMS BASED ON PCA

Step-03:

Subtract mean vector (μ) from the given feature vectors.

$x_1 - \mu = (2 - 4.5, 1 - 5) = (-2.5, -4)$

$x_2 - \mu = (3 - 4.5, 5 - 5) = (-1.5, 0)$

$x_3 - \mu = (4 - 4.5, 3 - 5) = (-0.5, -2)$

$x_4 - \mu = (5 - 4.5, 6 - 5) = (0.5, 1)$

$x_5 - \mu = (6 - 4.5, 7 - 5) = (1.5, 2)$

$x_6 - \mu = (7 - 4.5, 8 - 5) = (2.5, 3)$

Feature vectors (x_i) after subtracting mean vector (μ) are- $\begin{bmatrix} -2.5 \\ -4 \end{bmatrix}$ $\begin{bmatrix} -1.5 \\ 0 \end{bmatrix}$ $\begin{bmatrix} -0.5 \\ -2 \end{bmatrix}$ $\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$ $\begin{bmatrix} 1.5 \\ 2 \end{bmatrix}$ $\begin{bmatrix} 2.5 \\ 3 \end{bmatrix}$

PRACTICE PROBLEMS BASED ON PCA

<u>Step-04:</u>

Calculate the covariance matrix.

Covariance matrix is given by-

$$\text{Covariance Matrix} = \frac{\sum (x_i - \mu)(x_i - \mu)^t}{n}$$

PRACTICE PROBLEMS BASED ON PCA

$m_1 =$

$$m_1 = (x_1 - \mu)(x_1 - \mu)^t = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \begin{bmatrix} -2.5 & -4 \end{bmatrix} = \begin{bmatrix} 6.25 & 10 \\ 10 & 16 \end{bmatrix}$$ ✓

$m_2 =$

$$m_2 = (x_2 - \mu)(x_2 - \mu)^t = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -1.5 & 0 \end{bmatrix} = \begin{bmatrix} 2.25 & 0 \\ 0 & 0 \end{bmatrix}$$ ✓

$$m_3 = (x_3 - \mu)(x_3 - \mu)^t = \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} -0.5 & -2 \end{bmatrix} = \begin{bmatrix} 0.25 & 1 \\ 1 & 4 \end{bmatrix}$$ ✓

$$m_4 = (x_4 - \mu)(x_4 - \mu)^t = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$ ✓

$$m_5 = (x_5 - \mu)(x_5 - \mu)^t = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 1.5 & 2 \end{bmatrix} = \begin{bmatrix} 2.25 & 3 \\ 3 & 4 \end{bmatrix}$$ ✓

$m_6 =$

$$m_6 = (x_6 - \mu)(x_6 - \mu)^t = \begin{bmatrix} 2.5 \\ 3 \end{bmatrix} \begin{bmatrix} 2.5 & 3 \end{bmatrix} = \begin{bmatrix} 6.25 & 7.5 \\ 7.5 & 9 \end{bmatrix}$$ ✓

PRACTICE PROBLEMS BASED ON PCA

Now,

$$\frac{\sum (x_1-\mu)(x_1-\mu)^T}{n} \quad n=6$$

Covariance matrix

$= (m_1 + m_2 + m_3 + m_4 + m_5 + m_6) / 6$

On adding the above matrices and dividing by 6, we get-

$$\text{Covariance Matrix} = \frac{1}{6}\begin{bmatrix} 17.5 & 22 \\ 22 & 34 \end{bmatrix}$$

$$\text{Covariance Matrix} = \begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix}$$

PRACTICE PROBLEMS BASED ON PCA

Step-05:

Calculate the eigen values and eigen vectors of the covariance matrix.

λ is an eigen value for a matrix M if it is a solution of the characteristic equation $|M - \lambda I| = 0$.

So, we have-

$$\begin{vmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{vmatrix} - \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} = 0$$

$$\begin{vmatrix} 2.92 - \lambda & 3.67 \\ 3.67 & 5.67 - \lambda \end{vmatrix} = 0$$

$(2.92 - \lambda)(5.67 - \lambda) - 3.67 \times 3.67 = 0$

PRACTICE PROBLEMS BASED ON PCA

From here,

$(2.92 - \lambda)(5.67 - \lambda) - (3.67 \times 3.67) = 0$

$16.56 - 2.92\lambda - 5.67\lambda + \lambda^2 - 13.47 = 0$

$\lambda^2 - 8.59\lambda + 3.09 = 0$

Solving this quadratic equation, we get $\lambda = 8.22, 0.38$

Thus, two eigen values are $\lambda_1 = 8.22$ and $\lambda_2 = \boxed{0.38}$. → Remove

Clearly, the second eigen value is very small compared to the first eigen value.

So, the second eigen vector can be left out.

Eigen vector corresponding to the greatest eigen value is the principal component for the given data set.

So. we find the eigen vector corresponding to eigen value λ_1.

PRACTICE PROBLEMS BASED ON PCA

We use the following equation to find the eigen vector-

$MX = \lambda X$

Where, M = Covariance Matrix, X = Eigen vector, λ = Eigen value

Substituting the values in the above equation, we get-

$$\begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \end{bmatrix} = \boxed{8.22} \begin{bmatrix} X1 \\ X2 \end{bmatrix}$$

$2.92x_1 + 3.67x_2 = 8.22x_1$

$3.67x_1 + 5.67x_2 = 8.22x_2$

PRACTICE PROBLEMS BASED ON PCA

Solving these, we get-

$$\left.\begin{aligned}2.92X_1 + 3.67X_2 &= 8.22X_1\\ 3.67X_1 + 5.67X_2 &= 8.22X_2\end{aligned}\right\}$$

On simplification, we get-

$\underline{5.3X_1 = 3.67X_2}$ ....(1)

$3.67X_1 = 2.55X_2$ ....(2)

From (1) and (2), $\boxed{\mathbf{X_1 = 0.69X_2}}$

From (2), the eigen vector is-

Eigen Vector : $\begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$

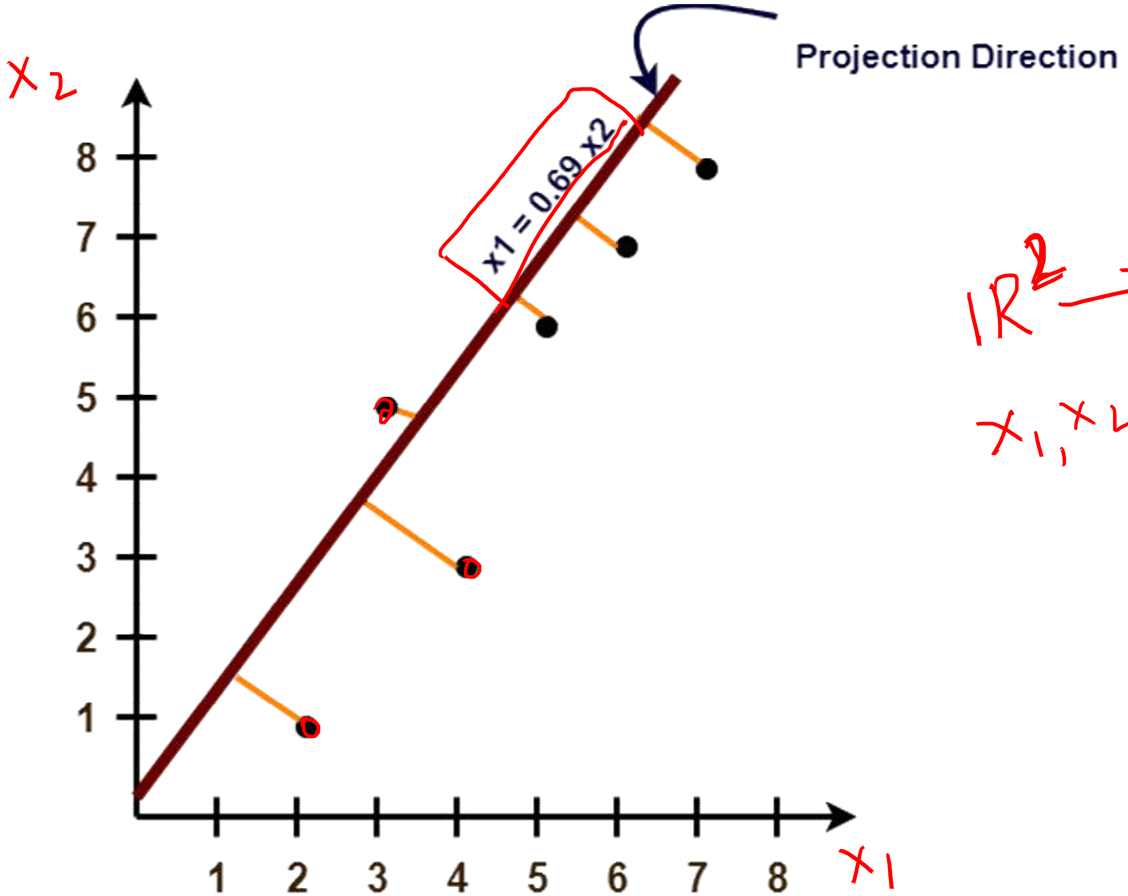
PRACTICE PROBLEMS BASED ON PCA

Thus, principal component for the given data set is-

$$\textbf{Principal Component :} \begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$$

PRACTICE PROBLEMS BASED ON PCA

Lastly, we project the data points onto the new subspace as-



$\mathbb{R}^2 \rightarrow \mathbb{R}^1$

$x_1, x_2 \rightarrow z_1$

PRACTICE PROBLEMS BASED ON PCA

Use PCA Algorithm to transform the pattern (2, 1) onto the eigen vector in the previous question.

The given feature vector is (2, 1).

Given Feature Vector : $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

PRACTICE PROBLEMS BASED ON PCA

The feature vector gets transformed to

= Transpose of Eigen vector x (Feature Vector – Mean Vector)

$$= \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}^T \times \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 4.5 \\ 5 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 2.55 & 3.67 \end{bmatrix} \times \begin{bmatrix} -2.5 \\ -4 \end{bmatrix}$$

$$\boxed{= -21.055}$$

PCA

The space of all face images

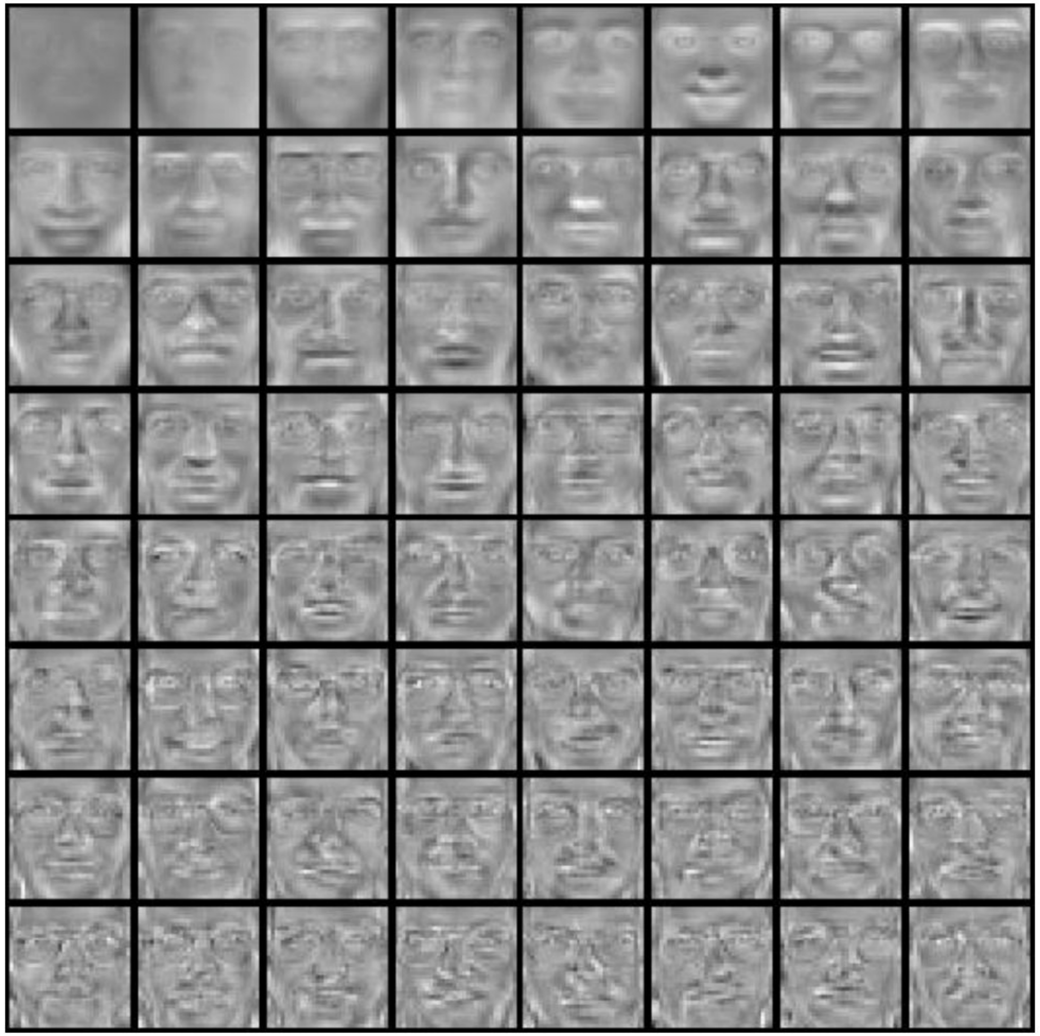
- When viewed as vectors of pixel values, face images are extremely high-dimensional
  - 100x100 image = 10,000 dimensions
  - Slow and lots of storage
- But very few 10,000-dimensional vectors are valid face images
- We want to effectively model the subspace of face images

slide by Derek Hoiem

Eigenfaces example

Mean: μ

Top eigenvectors: $u_1,\ldots u_k$



slide by Derek Hoiem

Representation and reconstruction

- Face **x** in "face space" coordinates:

$$\mathbf{x} \rightarrow [\mathbf{u}_1^{\mathrm{T}}(\mathbf{x} - \mu), \ldots, \mathbf{u}_k^{\mathrm{T}}(\mathbf{x} - \mu)]$$
$$= \quad w_1, \ldots, w_k$$

- Reconstruction:

 =  + 

$\hat{x} = \mu + w_1u_1 + w_2u_2 + w_3u_3 + w_4u_4 + \ldots$

slide by Derek Hoiem

Reconstruction

P = 4

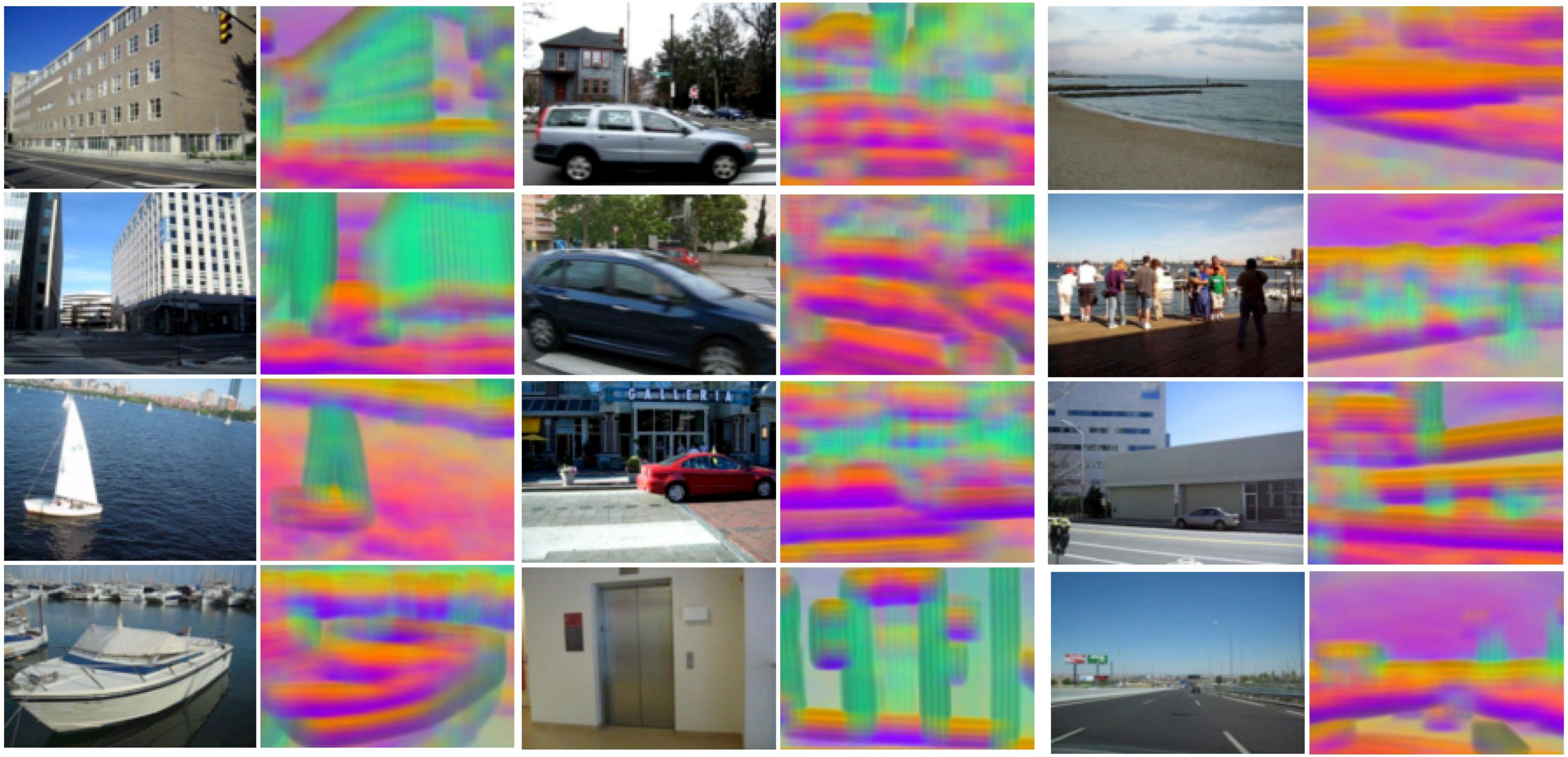
P = 200

P = 400

After computing eigenfaces using 400 face images from ORL face database

slide by Derek Hoiem

SIFT feature visualization



- The top three principal components of SIFT descriptors from a set of images are computed
- Map these principal components to the principal components of the RGB space
- pixels with similar colors share similar structures

Application: Image compression

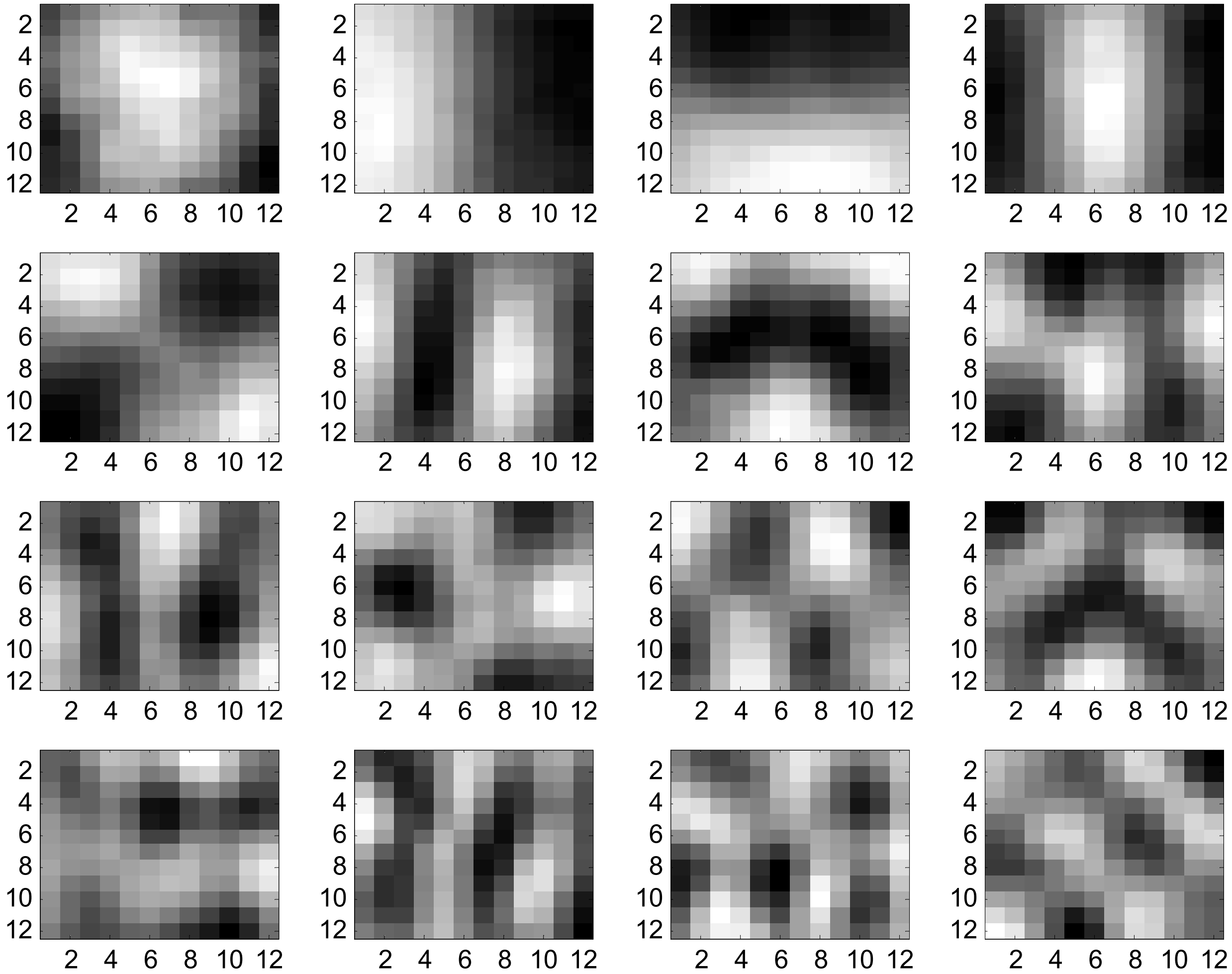
ORIGINAL IMAGE

- 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

PCA COMPRESSION: 144D → 60D

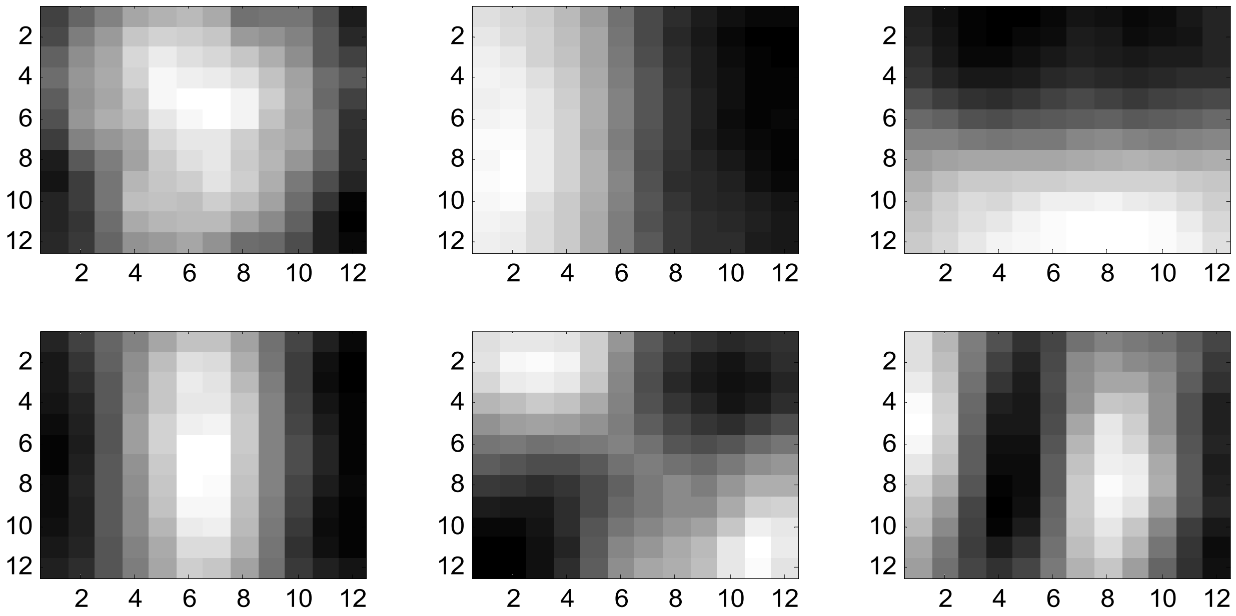
PCA COMPRESSION: 144D → 16D

16 MOST IMPORTANT EIGENVECTORS



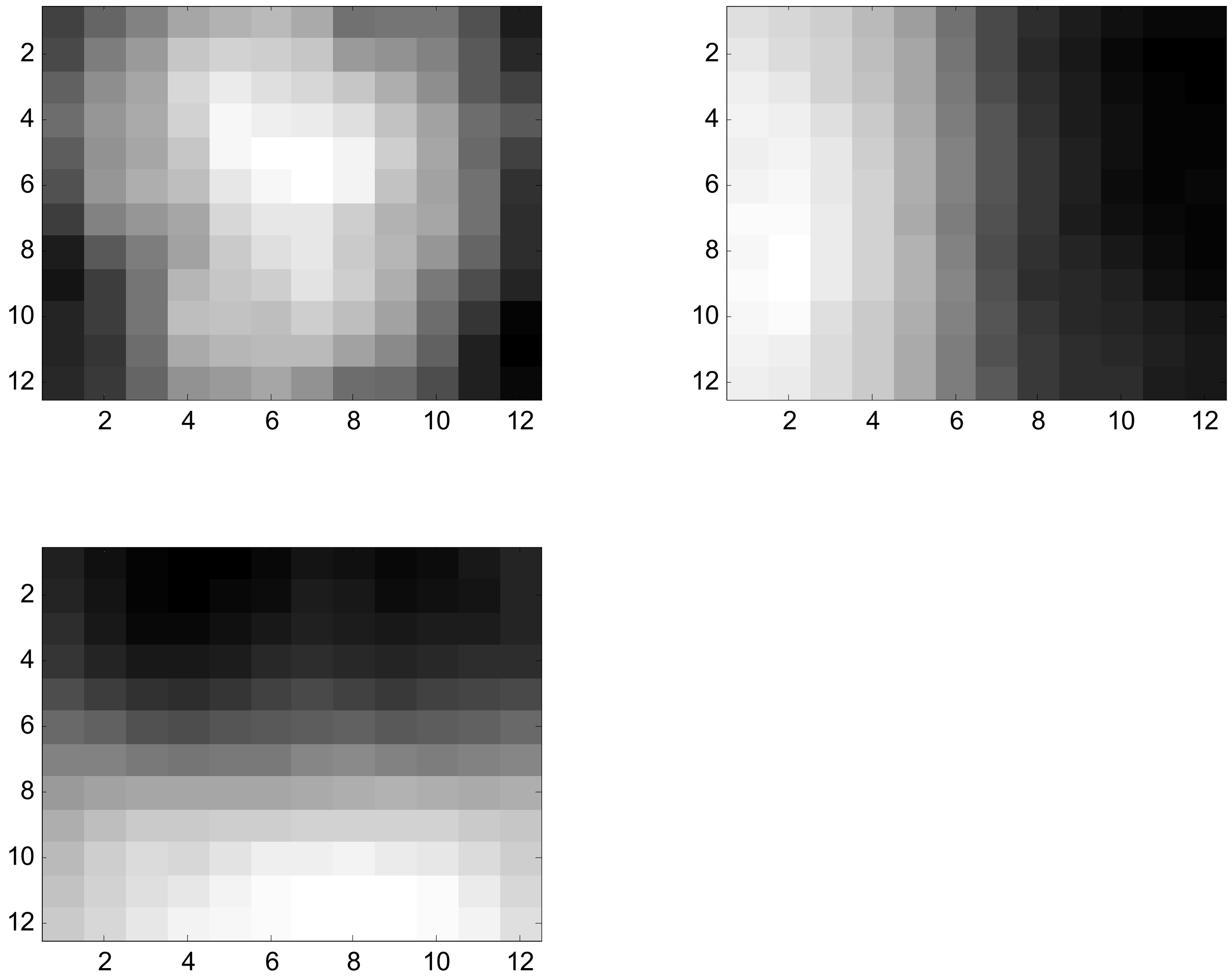
PCA COMPRESSION: 144D) 6D

6 MOST IMPORTANT EIGENVECTORS

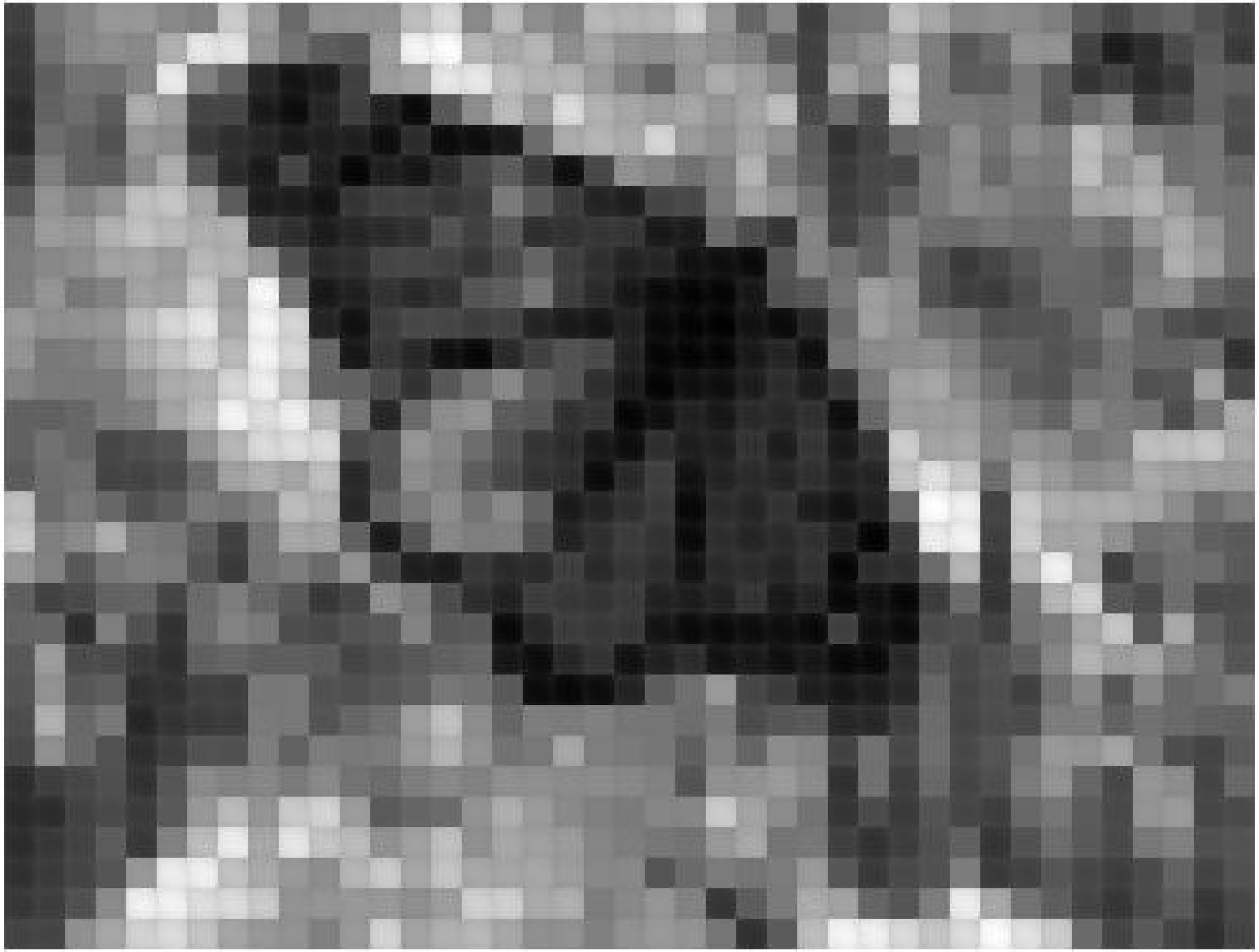


PCA COMPRESSION: 144D → 3D

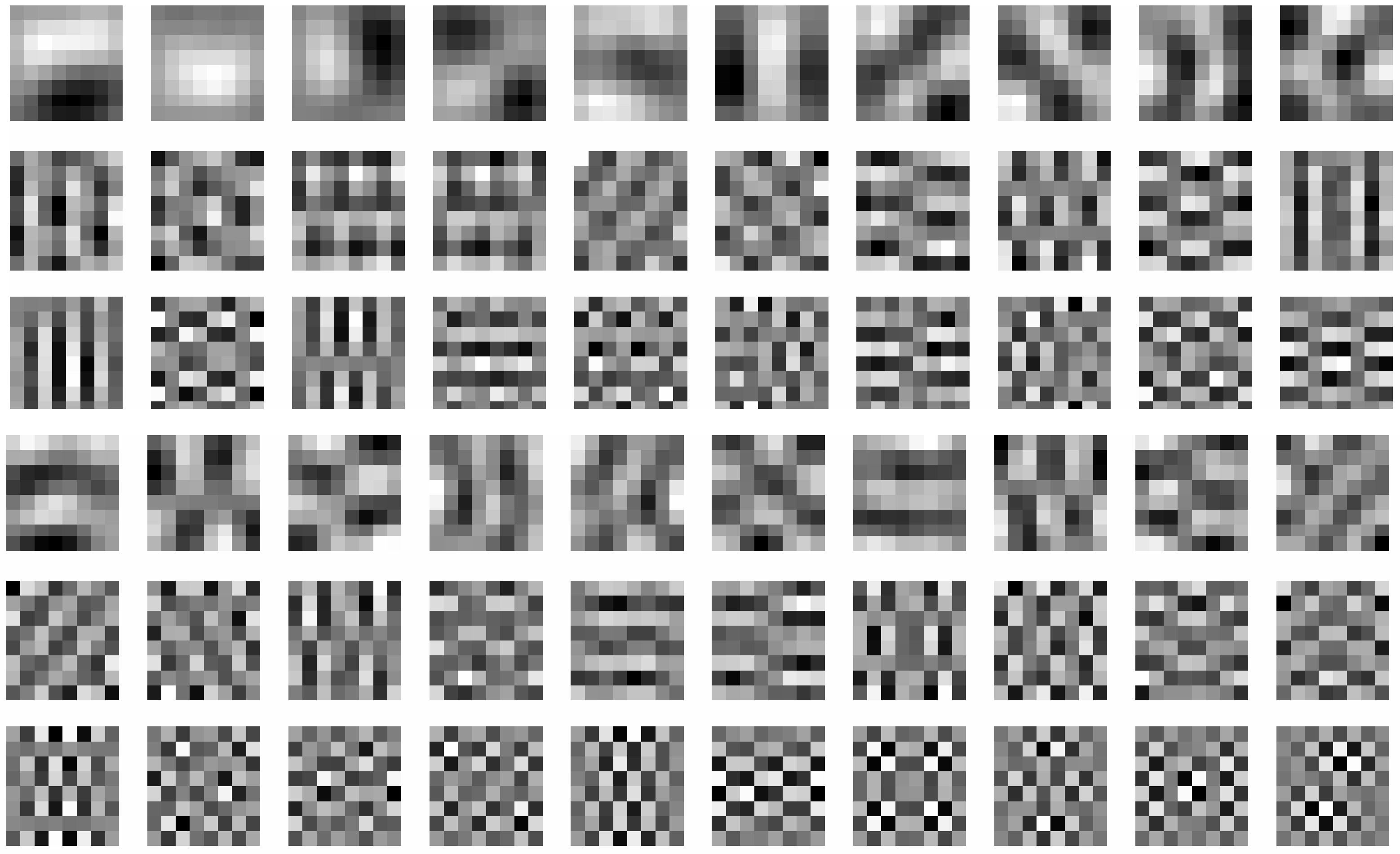
3 most important eigenvectors



PCA COMPRESSION: 144D → 1D

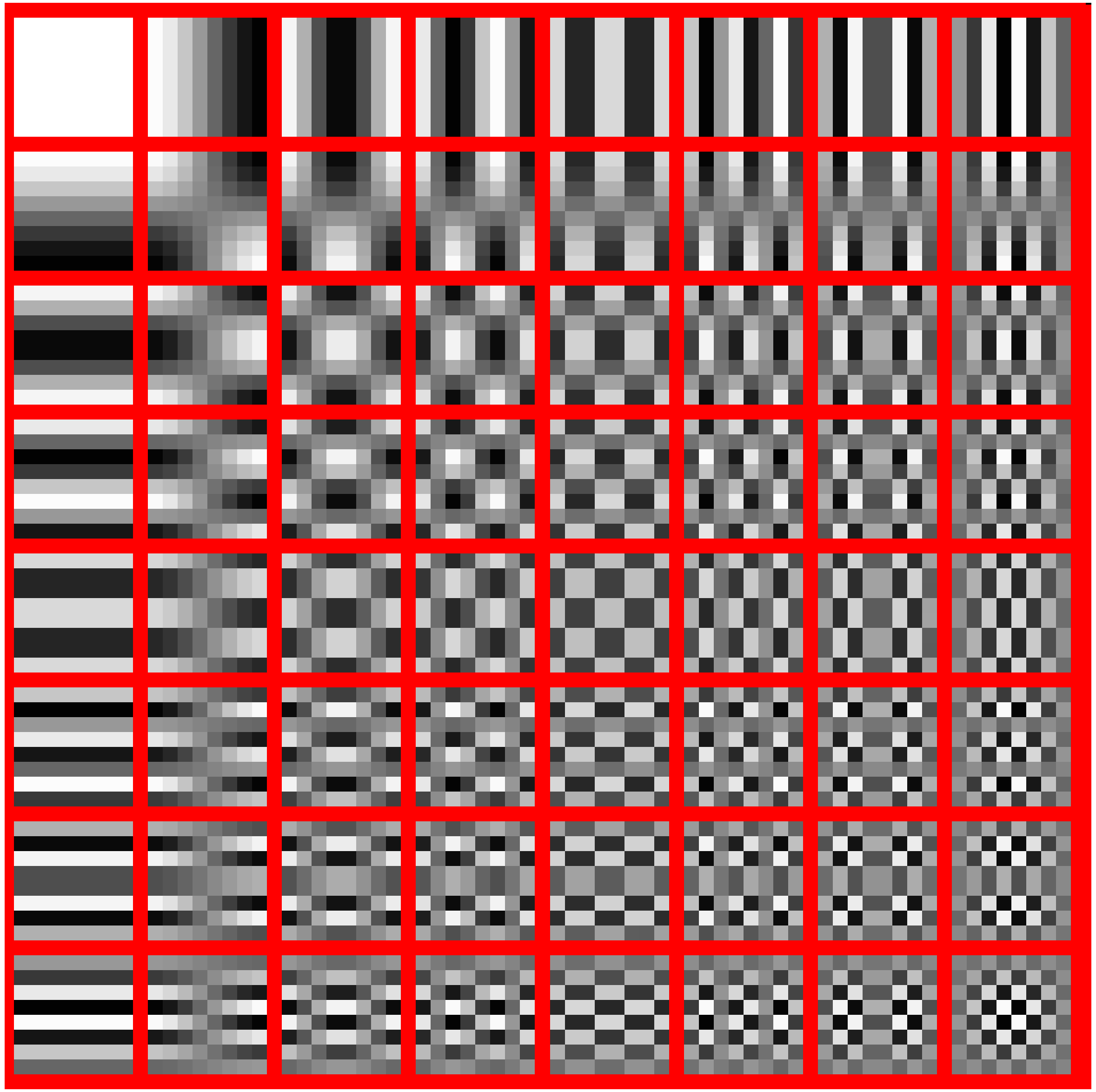


60 most important eigenvectors



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

2D DISCRETE COSINE BASIS



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

DIMENSIONALITY REDUCTION

PCA (Principal Component Analysis):

- Find projection that maximize the variance

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

…

…

THANKS


Keep Learning
Keep Growing


Dr. Neeraj Gupta
Assistant Professor, Dept. of CEA
neeraj.gupta@gla.ac.in