Introducing Principal Components Analysis

CS 510

Lecture #10

February 18th, 2013

Colorado State University

Overview: Goal

- Assume you have a gallery (database) of images, and a "probe" (test) image.
- The goal is to find the database image that is most similar to the probe image.
- "Similar" defined according to any measure
  - e.g. correlation

Example: Cats

Probe image, registered to gallery

Registered Gallery of Images

Registration

- Whole image matching presumes alignment
- Images are points in N-dimensional space
  - Dimensions meaningless unless points correspond
- Comparisons undefined if sizes differ
- Faces, often eye's map to same positions.
  - Specifies rotation, translation and scale.

Multiple Images of One Object

- Another reason for matching a probe against a gallery
  - Sample possible object variants, e.g.
  - Object seen from all (standard) viewpoints
  - Object seen under all (standard) illuminations
- Goal: find pose or illumination condition
- Bit brute force,
  - but strong if variants are present.

Alternate Example

Example Probe image

Five of 71 gallery images (COIL)

But Wait, This is Expensive!

- It is very costly to compare whole images.
- How can we save effort …
  - A lot of effort!
- Just how much variation is there in …
  - Faces of cats
- Is there a systematic way to measure
  - … and then work with less data?
- Yes, which takes us next to covariance.

Background Concepts: Variance

- Variance - the central tendency,
  - variance is defined as:

$$\frac{\sum\left(x-\bar{x}\right)^{2}}{N}$$

- Square root of variance is the standard deviation

Background Concepts: Covariance

- Covariance measures if two signals vary together:

$$\Omega = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{N}$$

- How does this differ from correlation?
- The range of the covariance of two signals?
- Note that the covariance of two signals is a scalar

Covariance Matrices (I)

- What if I want to know the relation between the ith element of x and the jth element of y?

$$\frac{1}{N}\Sigma = \begin{bmatrix} \sigma_{x_1,y_1} & \sigma_{x_1,y_2} & \cdots \\ \sigma_{x_2,y_1} & \sigma_{x_2,y_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$\sigma_{x_i,y_j} = \sum \left(x_i - \bar{x}\right)\left(y_j - \bar{y}\right)$$

Background Concepts: Outer Products

- Remember outer products :

$$\begin{bmatrix} ad & ae & af \\ bd & be & bf \\ cd & ce & cf \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} d & e & f \end{bmatrix}$$

- Why?
- Because if I have two vectors, their covariance is their outer product

Covariance Matrices (II)

- The covariance between two vectors isn't too interesting, but…
- What if I have two sets of vectors:
  - Let X = {(a1, b1, c1), (a2, b2, c2)}
  - Let Y = {(d1, e1, f1), (d2, e2, f2)}
- Assume the vectors are ***centered***
  - Meaning that the average X vector is subtracted from the X set, and the average Y is subtracted from the Y set
- What is the covariance between the sets of vectors?

Covariance Matrices (III)

- The covariance matrix is the outer product:

$$\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} d_1 & e_1 & f_1 \\ d_2 & e_2 & f_2 \end{bmatrix} = \begin{bmatrix} a_1 d_1 + a_2 d_2 & a_1 e_1 + a_2 e_2 & a_1 f_1 + a_2 f_2 \\ b_1 d_1 + b_2 d_2 & b_1 e_1 + b_2 e_2 & b_1 f_1 + b_2 f_2 \\ c_1 d_1 + c_2 d_2 & c_1 e_1 + c_2 e_2 & c_1 f_1 + c_2 f_2 \end{bmatrix}$$

- $\Omega_{i,j}$ is the covariance of position i in set X with position j in set Y,
- assumes pair-wise matches

Covariance Matrices (IV)

- It is interesting & meaningful to look at the covariance of a set with itself:

$$\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2 \end{bmatrix}\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} = \begin{bmatrix} a_1a_1 + a_2a_2 & a_1b_1 + a_2b_2 & a_1c_1 + a_2c_2 \\ b_1a_1 + b_2a_2 & b_1b_1 + b_2b_2 & b_1c_1 + b_2c_2 \\ c_1a_1 + c_2a_2 & c_1b_1 + c_2b_2 & c_1c_1 + c_2c_2 \end{bmatrix}$$

- Now how do you interpret $\Omega_{i,j}$?
- Does Σ have any special properties?

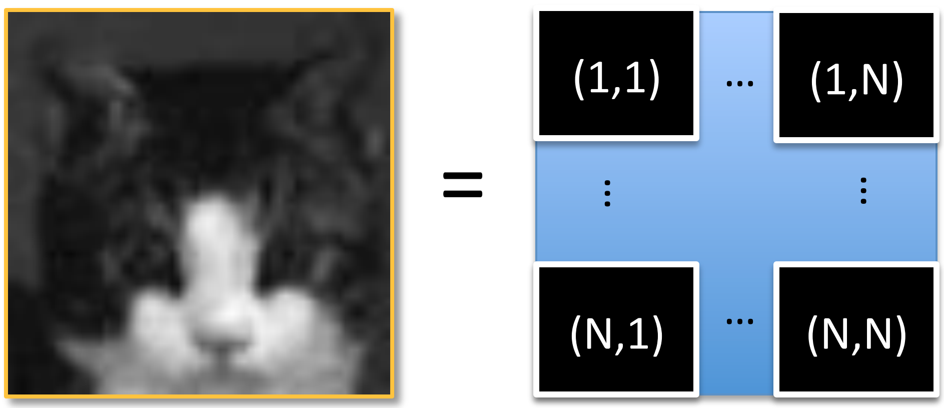
Principal Component Analysis

- PCA $\equiv$ SVD(Cov(X)) = SVD($XX^T/(n-1)$)
- PCA: $XX^T = R\Lambda R^{-1}$
  - R is a rotation matrix (the Eigenvector matrix)
  - Λ is a diagonal matrix (diagonal values are the Eigenvalues)
- The Eigenvalues capture how much the dimensions in X co-vary
- The Eigenvectors show which combinations of dimensions tend to vary together

PCA (II)

- The Eigenvector with the largest Eigenvalue is the direction of maximum variance
- The Eigenvector with the 2^{nd} largest Eigenvalue is orthogonal to the 1^{st} vector and has the next greatest variance.
- And so on…
- The Eigenvalues describe the amount of variance along the Eigenvectors

Review: Cookbook
Step 1: Image as Vector



$$\begin{bmatrix} I(1,1) \\ \vdots \\ I(1,N) \\ I(2,1) \\ \vdots \\ I(N,N) \end{bmatrix}$$

Step 2 (optional): Normalize

- Normalize each vector:
  - Compute mean value of vector
  - Subtract mean value

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \overline{X} = \sum_{i=0}^{N} x_i, X - \overline{X} = \begin{bmatrix} x_1 - \overline{x} \\ \vdots \\ x_N - \overline{x} \end{bmatrix}$$

Step 3 (optional) : mean-center data set

- Form data matrix
  - Images (samples) as columns

$$\begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ I_1 & I_2 & \cdots & I_K \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}$$

  - Subtract mean image from all columns

Step 4: Covariance of a Data Set

$$\begin{bmatrix} \omega_{1,1} & \cdots & \omega_{1,N} \\ \vdots & \vdots & \vdots \\ \omega_{N,1} & \cdots & \omega_{N,N} \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \cdots & \vdots \\ I_1 & I_2 & \cdots & I_K \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} \cdots & I_1 & \cdots \\ \cdots & I_2 & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & I_K & \cdots \end{bmatrix}$$

$$\omega_{i,j} = \sum_{k}^{K} \left(x^k{}_i - \bar{x}^k \right)\left(x^k{}_j - \bar{x}^k \right)$$

Step 5: PCA

- Let $I_1,\ldots, I_N$ be normalized images

$$X \equiv \begin{bmatrix} \vdots & \cdots & \vdots \\ I_1 & \cdots & I_N \\ \vdots & \cdots & \vdots \end{bmatrix}$$

$$Cov(X) = XX^T$$

$$PCA(X) = SVD\left(XX^T\right) = R^T \Lambda R$$

PCA: where are we?

- Done
  - Mechanics & algorithms
  - Motivation as maximizing variance
- To do
  - Motivation as Gaussian Random Process
  - Image space interpretation

Multivariate Normal Random Variables

- The equation of a 1D Gaussian

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

  - μ is the mean and σ is the standard deviation
- Covariance generalizes variance to n-dimensions.
- The N-dimensional Gaussian is defined as

$$f(\vec{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{\frac{-1}{2}(\vec{x}-\mu)\Sigma^{-1}(\vec{x}-\mu)}$$

Multivariate Normal (II)

- Consider the case of a 2D Gaussian.

$$f(x,y) = \frac{1}{2\pi \left\| \begin{matrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{matrix} \right\|^{1/2}} e^{-\frac{1}{2} \left| \begin{matrix} x-\mu_x \\ y-\mu_y \end{matrix} \right|^T \left| \begin{matrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{matrix} \right|^{-1} \left| \begin{matrix} x-\mu_x \\ y-\mu_y \end{matrix} \right|}$$

Covariance Matrix

Special Case: Axis Aligned

$$\sigma_{xx} = \sigma_x^{\,2} \quad \sigma_{yy} = \sigma_y^{\,2} \quad \sigma_{xy} = 0$$

$$f(\vec{x}) = \frac{1}{2\pi\left(\sigma_x^{\,2}\sigma_y^{\,2}\right)^{1/2}} \exp\left[-\frac{1}{2} \begin{vmatrix} x-\mu_x \\ y-\mu_y \end{vmatrix}^T \begin{vmatrix} \frac{1}{\sigma_x^{\,2}} & 0 \\ 0 & \frac{1}{\sigma_y^{\,2}} \end{vmatrix} \begin{vmatrix} x-\mu_x \\ y-\mu_y \end{vmatrix} \right]$$

$$= \frac{1}{2\pi\left(\sigma_x\sigma_y\right)} \exp\left[-\frac{1}{2}\left(\frac{\left(x-\mu_x\right)^2}{\sigma_x^{\,2}} + \frac{\left(y-\mu_y\right)^2}{\sigma_y^{\,2}} \right) \right]$$

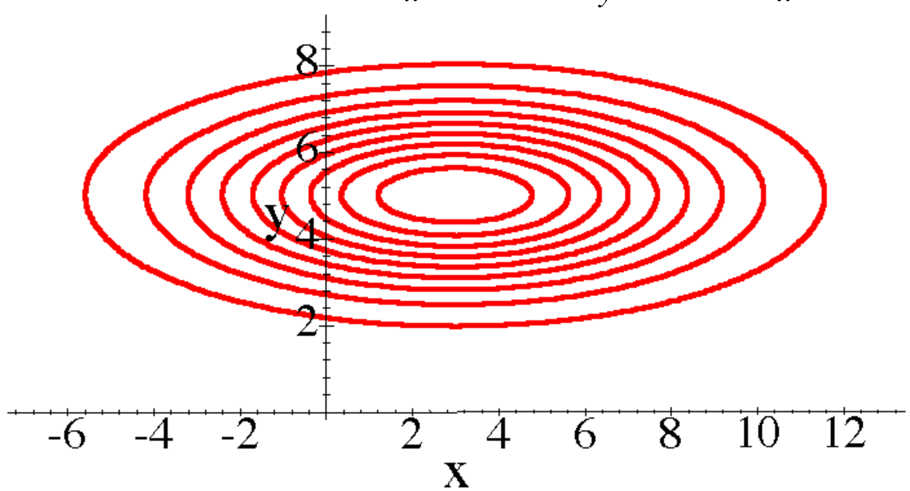
$$= \left(\frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2} \frac{\left(x-\mu_x\right)^2}{\sigma_x^{\,2}} \right] \right) \left(\frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2} \frac{\left(y-\mu_y\right)^2}{\sigma_y^{\,2}} \right] \right)$$

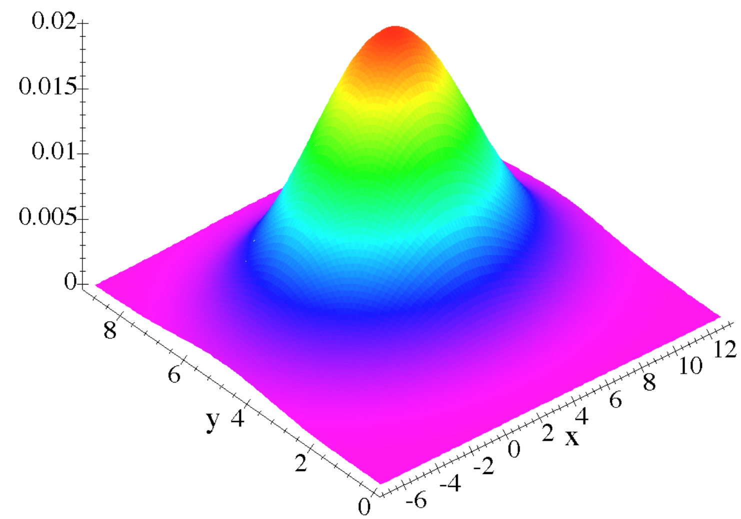
Probability Level Curves

- Consider the following axis-aligned Gaussian:

$$f(x,y)=\frac{1}{2\pi(\sigma_x\sigma_y)}\exp\left[-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{{\sigma_x}^2}+\frac{(y-\mu_y)^2}{{\sigma_y}^2}\right)\right]$$

$$\sigma_x=4,\quad \sigma_y=2,\ \mu_x=3,\ \mu_y=5$$





Quadratic Forms

- Look at the exponent of the centered (μ=0) 2D Gaussian, it has the form:

$$f(x,y) = V^T M V = \begin{vmatrix} x & y \end{vmatrix} \begin{vmatrix} a & b \\ b & c \end{vmatrix} \begin{vmatrix} x \\ y \end{vmatrix}$$

$$= ax^2 + 2bxy + cy^2$$

- Singular value decomposition tells us that:

$$M = R \Delta R^{-1}$$

$$= \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{vmatrix} \begin{vmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{vmatrix}$$

- R rotates coordinates so M is diagonal.

Quadratic Forms Rotated

We may specify any quadratic form as being rotated from an axis aligned equivalent.

$$f(u,v) = V^T D V \qquad f(u,v) = [u \quad v] \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$V = R X \qquad \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

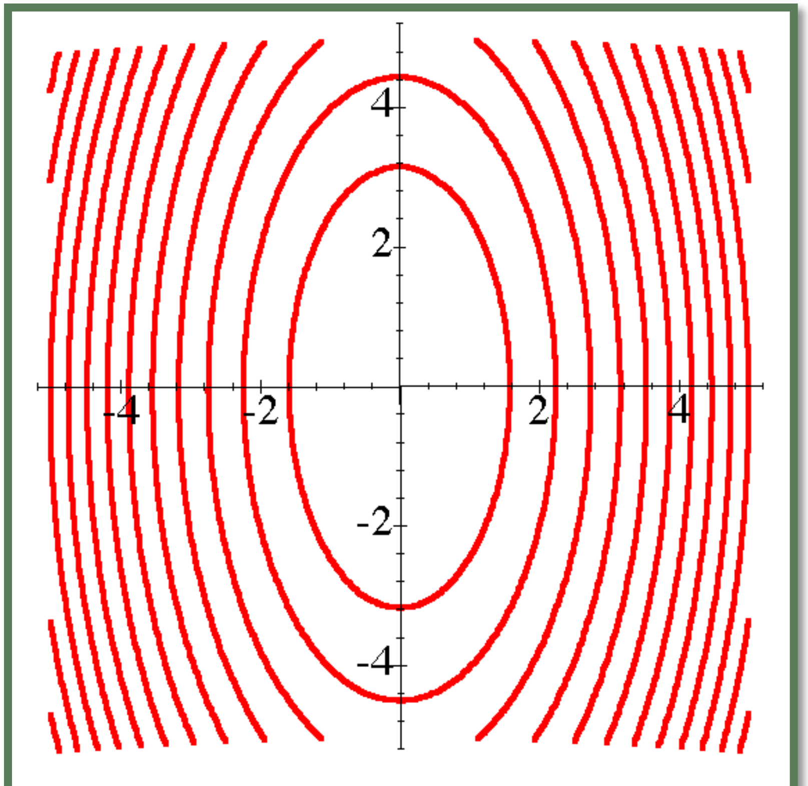
$$V^T = (R X)^T \qquad [u \quad v] = [x \quad y] \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

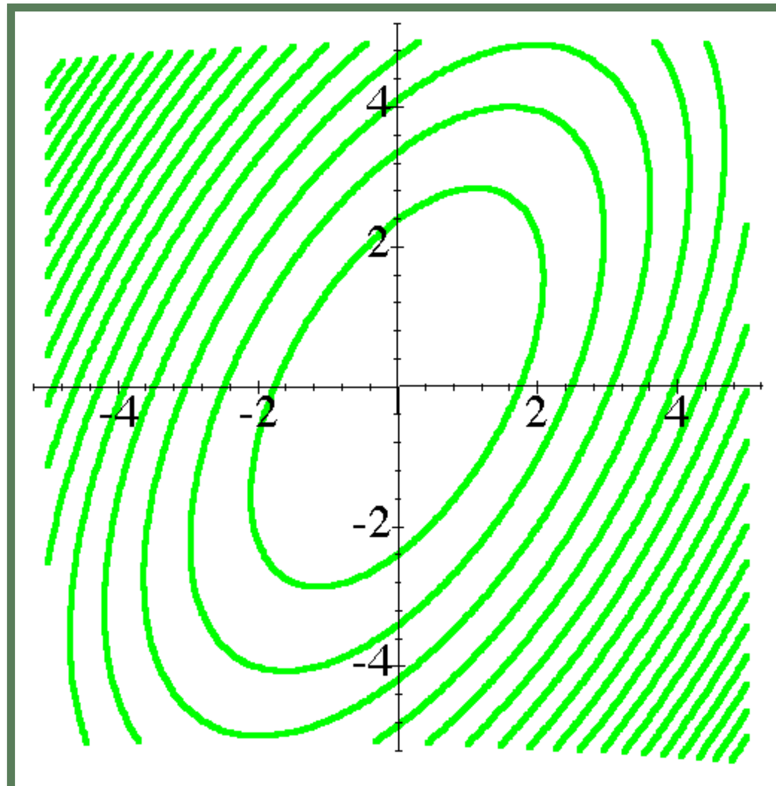
$$f(x,y) = X^T R^T D R X = X^T M X$$

Rotated Example

$$R\left(\frac{1}{6}\pi\right) = \begin{bmatrix} .865 & -.500 \\ .500 & .865 \end{bmatrix}$$

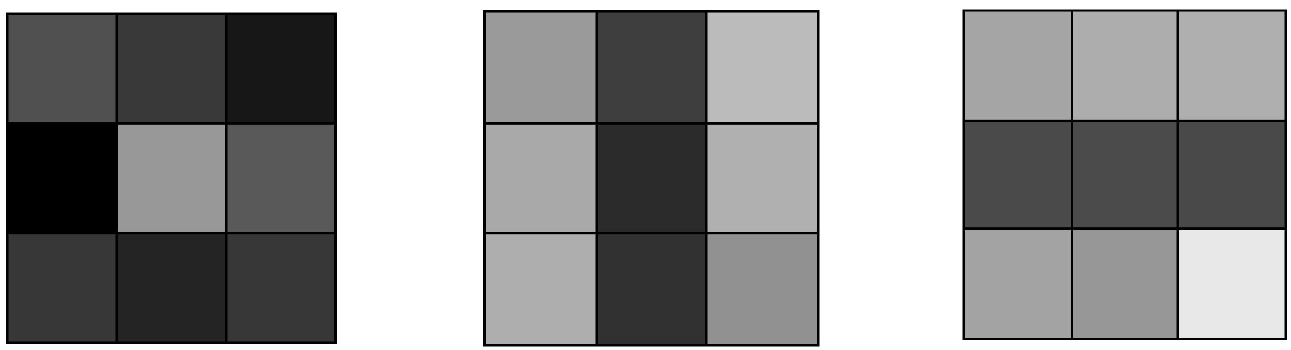
$$f_1(u, v) = 8\,x^2 + 2\,y^2 \qquad f_2(x, y) = 6.49\,x^2 - 5.20\,x\,y + 3.50\,y^2$$





A 9 dimensional example (via prototypes)

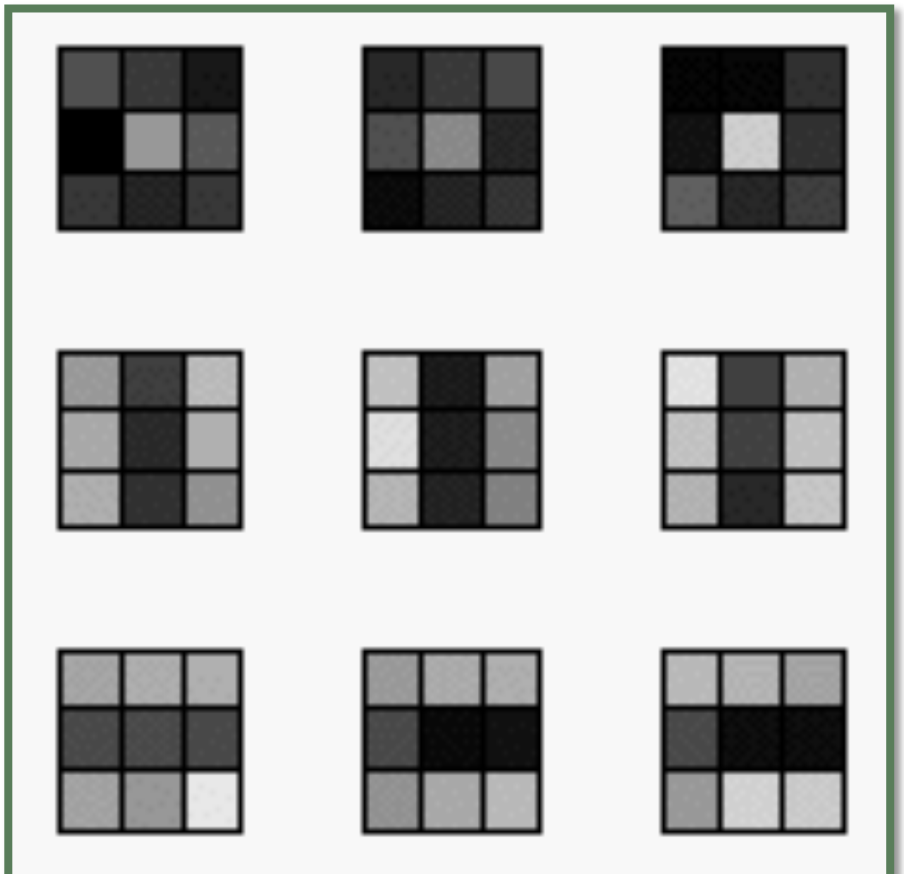
- Class 1 is dark around the edges and bright in the middle.
- Class 2 is light with dark vertical bars.
- Class 3 is light with dark horizontal bars.
- All classes initially use 2 for low value, 7 for high value.
- Each instance is corrupted by sigma=1 Guassian Noise.



Class 1 Class 2 Class 3

Eigenspace Example 1

- Consider 3 examples from the 3 classes.



Class 1

Class 2

Class 3

The Image Matrices

- Here they are as matrices.

$$\left[\begin{bmatrix}1.65 & 3.11 & 2.25\\ 3.22 & 5.79 & 3.09\\ 1.10 & 2.47 & 2.96\end{bmatrix}, \begin{bmatrix}1.55 & 3.29 & 1.62\\ 2.91 & 3.88 & .71\\ 2.35 & 3.60 & 2.46\end{bmatrix}, \begin{bmatrix}.80 & 2.43 & 2.04\\ 1.59 & 8.17 & .79\\ .69 & 1.96 & 4.34\end{bmatrix},\right.$$

$$\begin{bmatrix}6.36 & 2.39 & 9.36\\ 6.05 & .55 & 6.60\\ 5.97 & 3.49 & 7.33\end{bmatrix}, \begin{bmatrix}6.43 & 1.43 & 7.01\\ 7.66 & 3.20 & 6.66\\ 6.96 & 1.82 & 7.52\end{bmatrix}, \begin{bmatrix}6.52 & .89 & 7.74\\ 4.80 & 1.97 & 7.58\\ 5.75 & 1.06 & 7.24\end{bmatrix},$$

$$\left.\begin{bmatrix}8.11 & 8.94 & 5.85\\ 2.63 & 2.60 & 5.16\\ 7.20 & 6.09 & 6.12\end{bmatrix}, \begin{bmatrix}6.94 & 6.68 & 5.99\\ 3.63 & 3.15 & 1.37\\ 8.50 & 6.89 & 6.49\end{bmatrix}, \begin{bmatrix}7.02 & 7.73 & 7.08\\ 2.75 & 2.10 & 1.91\\ 5.92 & 6.85 & 7.16\end{bmatrix}\right]$$

Normalized Image Vectors

- Each as a 9x1 vector, an unrolled image.
- Each has zero mean and unit length.

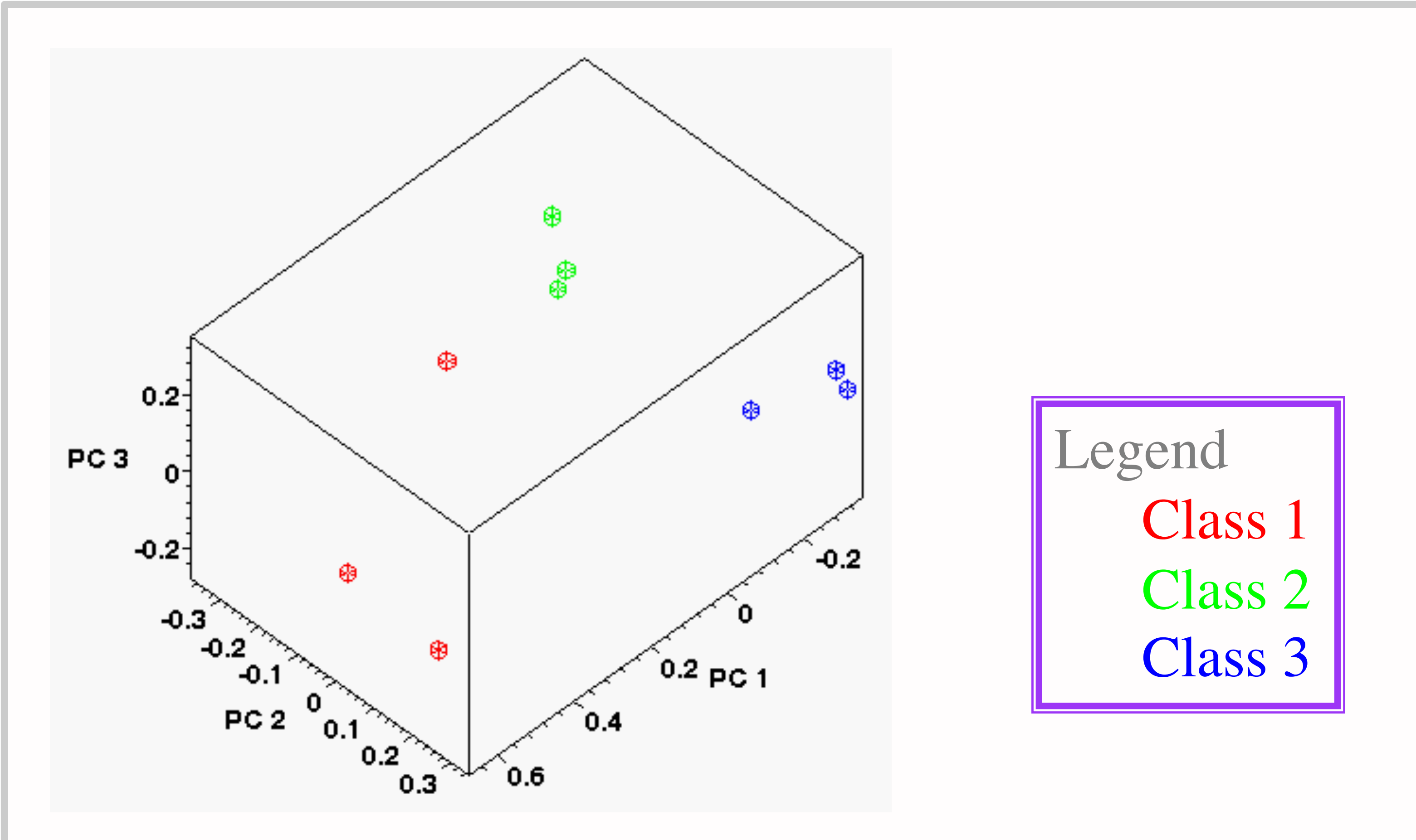
$$X = \left[\begin{bmatrix} -.133 \\ .0500 \\ -.102 \\ .0750 \\ .324 \\ .0910 \\ -.188 \\ .00100 \\ -.0630 \end{bmatrix}, \begin{bmatrix} -.117 \\ .127 \\ -.141 \\ .0930 \\ .186 \\ -.152 \\ -.0140 \\ .185 \\ -.0740 \end{bmatrix}, \begin{bmatrix} -.231 \\ -.0450 \\ -.143 \\ -.114 \\ .505 \\ -.163 \\ -.239 \\ -.0710 \\ .0450 \end{bmatrix}, \begin{bmatrix} .0480 \\ -.149 \\ .184 \\ .0710 \\ -.266 \\ .131 \\ .0300 \\ -.0670 \\ .0320 \end{bmatrix}, \begin{bmatrix} .0530 \\ -.202 \\ .0520 \\ .162 \\ -.116 \\ .135 \\ .0860 \\ -.161 \\ .0440 \end{bmatrix}, \begin{bmatrix} .0840 \\ -.229 \\ .124 \\ .0200 \\ -.178 \\ .217 \\ .0410 \\ -.200 \\ .0570 \end{bmatrix}, \begin{bmatrix} .126 \\ .197 \\ -.0290 \\ -.129 \\ -.157 \\ .0370 \\ .0800 \\ .0630 \\ -.0520 \end{bmatrix}, \begin{bmatrix} .0810 \\ .0930 \\ -.00600 \\ -.0660 \\ -.120 \\ -.163 \\ .172 \\ .124 \\ -.0150 \end{bmatrix}, \begin{bmatrix} .0900 \\ .157 \\ .0600 \\ -.113 \\ -.177 \\ -.132 \\ .0310 \\ .126 \\ .0270 \end{bmatrix} \right]$$

Singular Value Decomposition

- Actual values for this example.
- The Eigenvalues are 1.2, 0.55, 0.19 etc.
- The Eigenvectors are columns of U matrix.

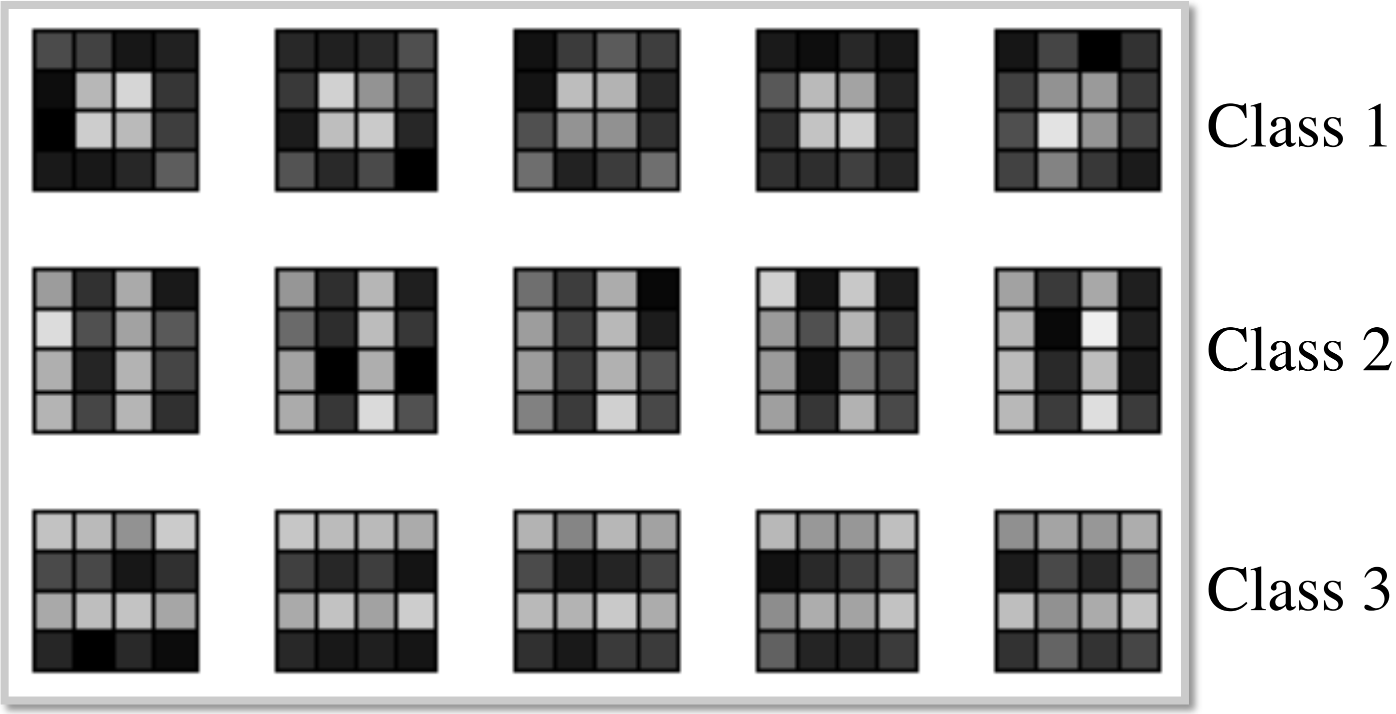
$$
\underbrace{\begin{bmatrix}
-.29 & -.080 & -.31 & -.51 & .40 & .30 & -.17 & -.15 & -.51 \\
-.14 & .46 & .10 & -.48 & -.27 & -.17 & -.46 & -.30 & .35 \\
-.23 & -.060 & .34 & .050 & -.49 & -.29 & -.18 & .16 & -.66 \\
-.23 & -.52 & .64 & -.060 & .12 & .16 & .050 & -.44 & .17 \\
.84 & .040 & .16 & -.10 & -.090 & .21 & -.12 & -.32 & -.30 \\
.080 & -.47 & -.42 & -.36 & -.45 & -.29 & .35 & -.22 & .080 \\
-.14 & -.10 & -.38 & .60 & -.080 & -.020 & -.40 & -.54 & -.050 \\
-.14 & .50 & .080 & .080 & .13 & -.21 & .62 & -.47 & -.23 \\
-.20 & .16 & -.050 & .050 & -.53 & .77 & .21 & 0 & .040
\end{bmatrix}}_{U}
\underbrace{\begin{bmatrix}
1.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & .55 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & .19 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & .12 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .030 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & .010 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}}_{D}
\underbrace{\begin{bmatrix}
-.29 & -.14 & -.23 & -.23 & .84 & .080 & -.14 & -.14 & -.20 \\
-.080 & .46 & -.060 & -.52 & .040 & -.47 & -.10 & .50 & .16 \\
-.31 & .10 & .34 & .64 & .16 & -.42 & -.38 & .080 & -.050 \\
-.51 & -.48 & .050 & -.060 & -.10 & -.36 & .60 & .080 & .050 \\
.40 & -.27 & -.49 & .12 & -.090 & -.45 & -.080 & .13 & -.53 \\
.30 & -.17 & -.29 & .16 & .21 & -.29 & -.020 & -.21 & .77 \\
-.17 & -.46 & -.18 & .050 & -.12 & .35 & -.40 & .62 & .21 \\
-.15 & -.30 & .16 & -.44 & -.32 & -.22 & -.54 & -.47 & 0 \\
-.51 & .35 & -.66 & .17 & -.30 & .080 & -.050 & -.23 & .040
\end{bmatrix}}_{U^T}
$$

Subspace Projection Pictures



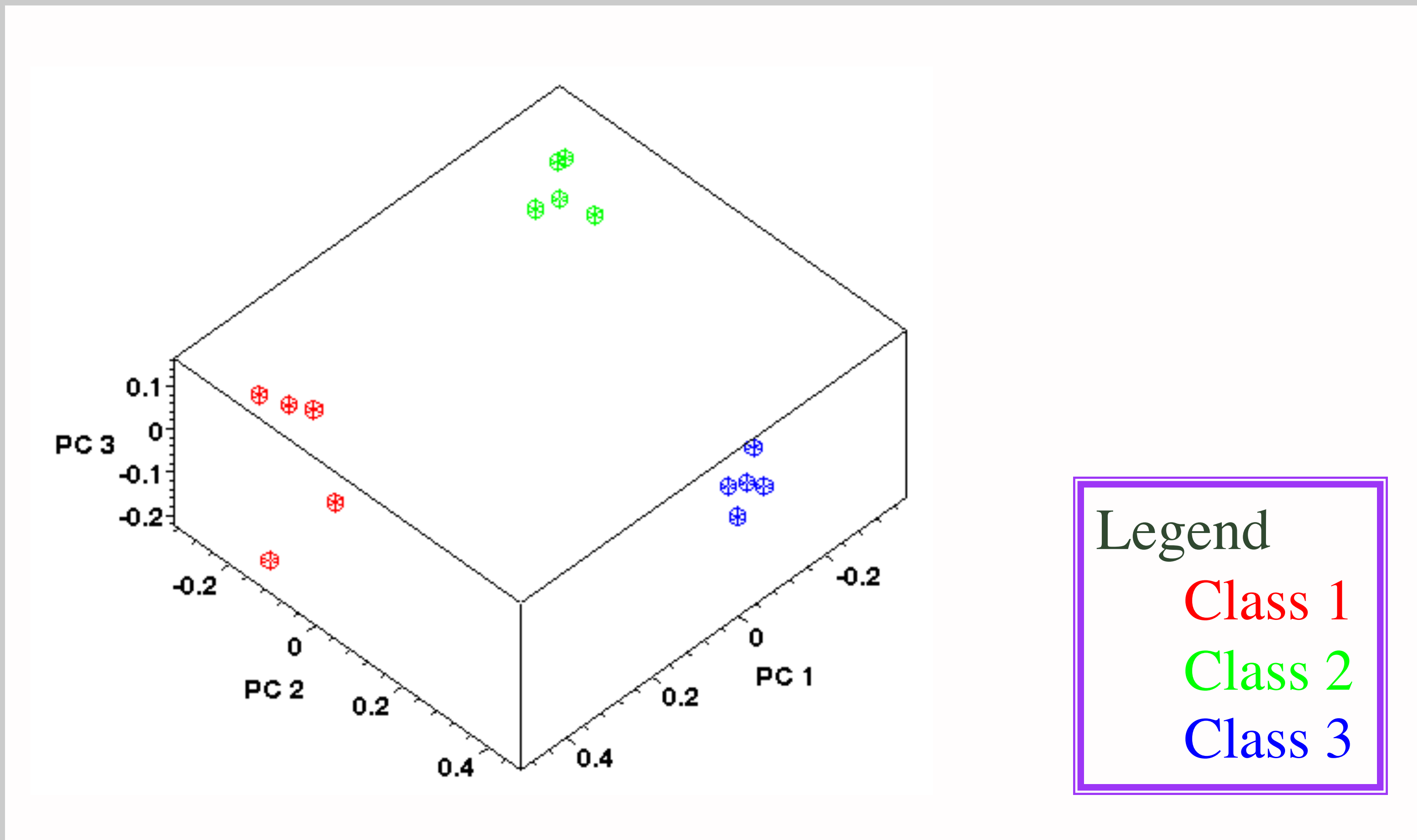
Eigenspace Example 2

- Consider 12 4x4 images.

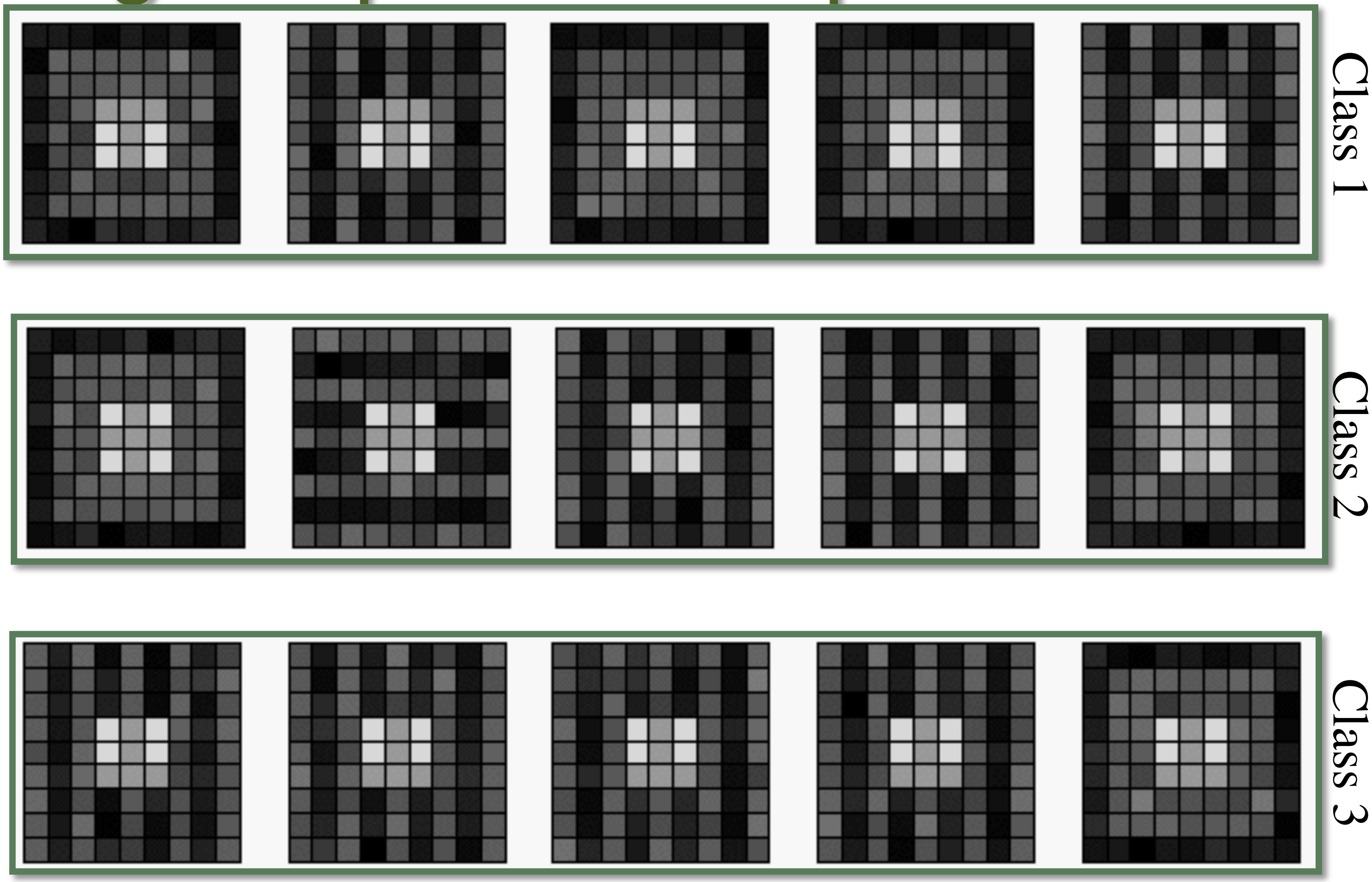


➢ Low value is 2, high is 7, noise sigma 1.0

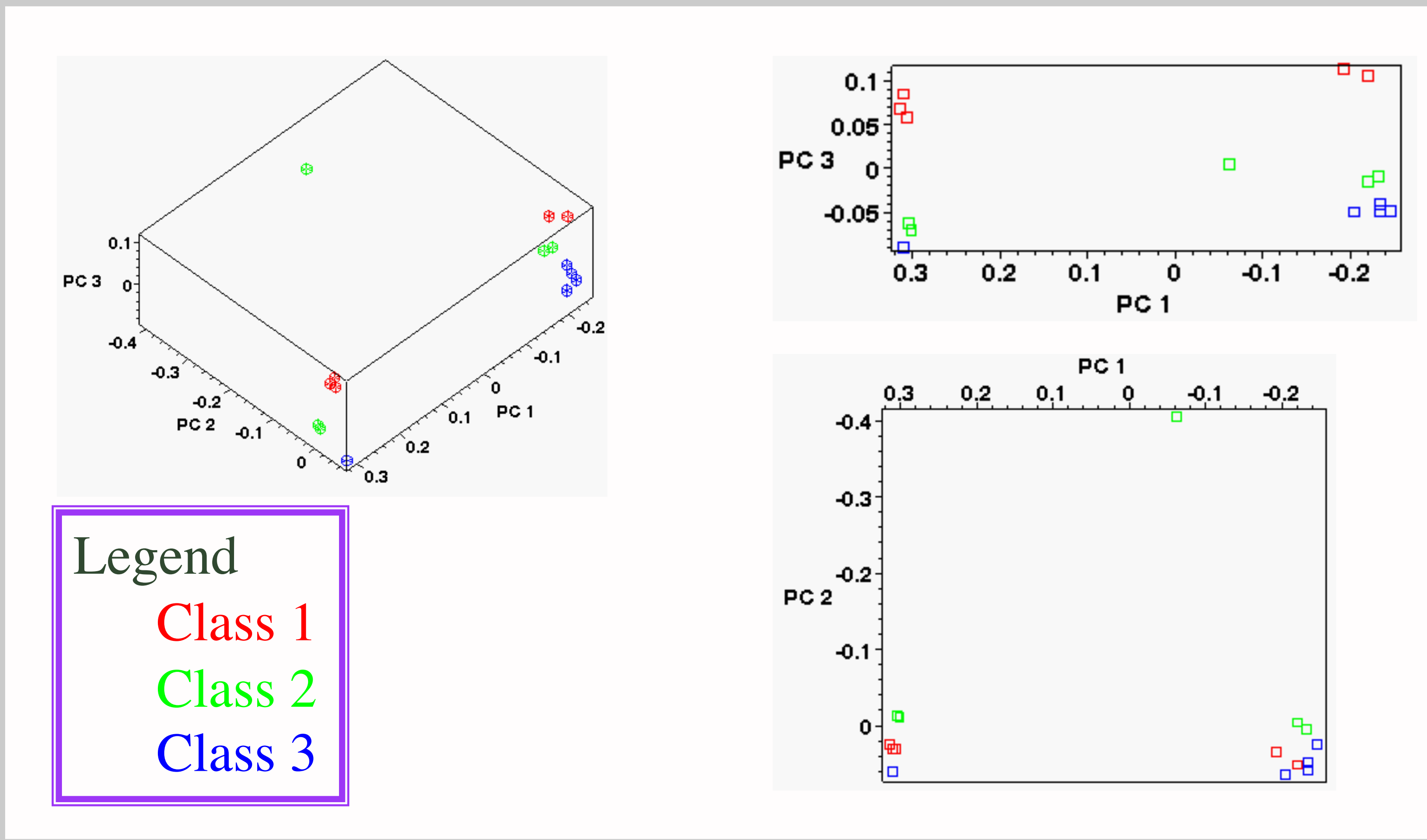
Example 2 Subspace 3D



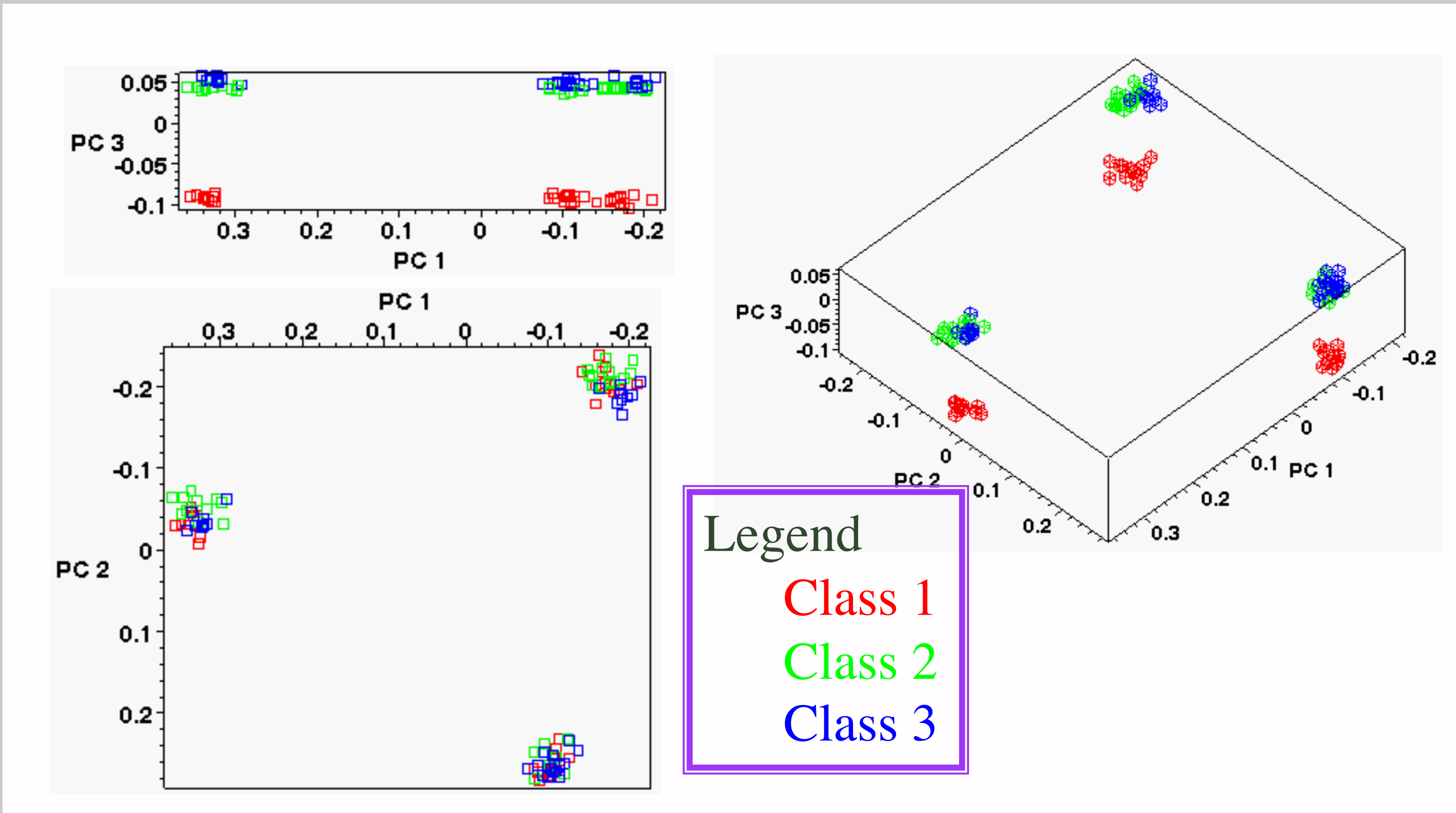
Eigen Space Example 3



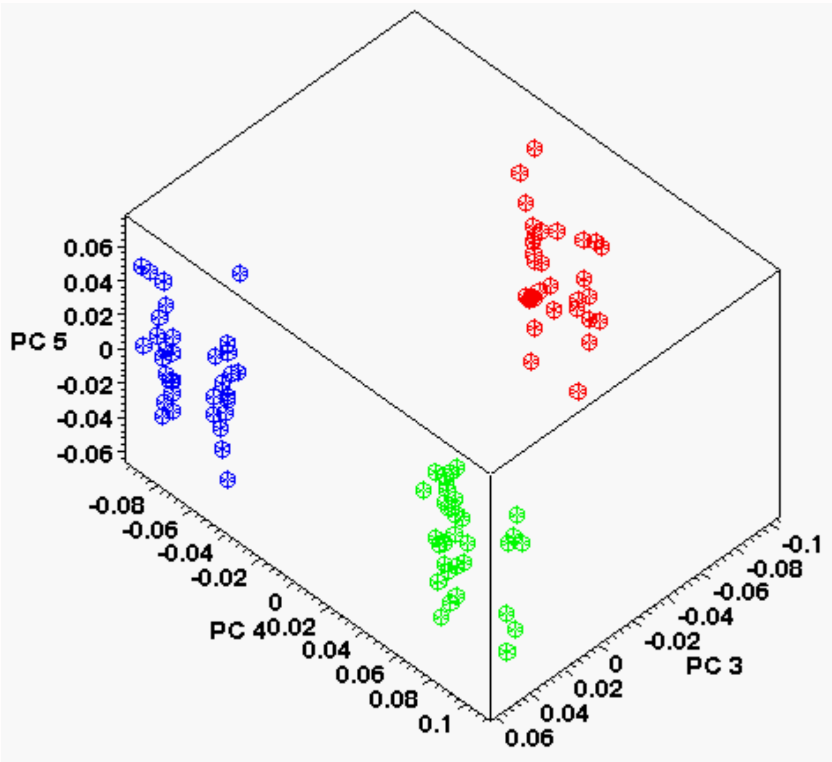
Example 3 Subspace

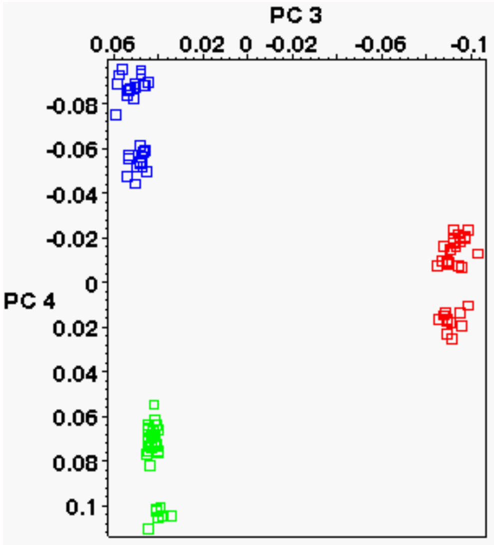


Example 3 Subspace: 99 Samples



Example 3: Dimensions 3, 4 & 5





Legend
Class 1
Class 2
Class 3

Example 3 Observations

- The first two Principle Components carry no information with respect to image class.
- However, Principle Components 3 and 4 carry all the information necessary for a nearest neighbors classifier

[4.775, 3.846, .4245, .3970, .07430]

1 2 3 4 5

The Eigenvalues, which record variance along each axis, show higher PC's have more variance:

Summary

- PCA can be applied to any set of registered images
- It extracts the dimensions of maximum co-variance
  - In some sense, the "structure" of the domain
- Dimensions of co-variance may or may not be related to classification