Introduction to Principal Component Analysis (PCA)

CS 510
Lecture #09
February 20, 2019
To Start – Correlation Matching

• 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: Finding Cats

“Probe” image -- image to be matched

Gallery of database images
Registration

• Whole image matching presumes alignment
• Images are points in N-dimensional space
  – Dimensions meaningless unless points correspond
• Comparisons undefined if sizes differ
• The same restrictions as (strict) correlation
• Faces, often eye’s map to same positions.
  – Specifies rotation, translation and scale.
Example: Aligned Cats

Probe image, registered to gallery

Registered Gallery of Images
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 from 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:
  \[
  \frac{\sum (x_i - \bar{x})^2}{N}
  \]

• Square root of variance is the standard deviation
Background Concepts: Covariance

- Covariance measures if two signals vary together:
  \[ \Omega = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N} \]

- How does this differ from correlation?
- The range of the covariance of two signals?
- Note the covariance of two signals is a scalar
Covariance Matrices (I)

• Let x and y be sets of vectors…

• What if I want to know the relation between the $i^{th}$ element of x and the $j^{th}$ element of y?

\[
\frac{1}{N} \sum \sigma_{x_i,y_j} = \begin{bmatrix}
\sigma_{x_1,y_1} & \sigma_{x_1,y_2} & \cdots \\
\sigma_{x_2,y_1} & \sigma_{x_2,y_2} & \cdots \\
\vdots & \vdots & \ddots 
\end{bmatrix}
\]

\[
\sigma_{x_i,y_j} = \sum (x_{i,k} - \bar{x}_i)(y_{j,k} - \bar{y}_j)
\]
Background Concepts: Outer Products

• Remember outer products:

\[
\begin{bmatrix}
ad & ae & af \\
bd & be & bf \\
\vdots & \vdots & \vdots \\
\end{bmatrix} = \begin{bmatrix}
a \\
b \\
c \\
\end{bmatrix}\begin{bmatrix}
d & e & f \\
\end{bmatrix}
\]

• Why?
• Because if I have two vectors, their covariance term is their outer product
Covariance Matrices (II)

• The covariance between two vectors isn’t too interesting, just a set of scalars, but…

• What if we have two sets of vectors:
  – Let $X = \{(a_1, b_1, c_1), (a_2, b_2, c_2)\}$
  – Let $Y = \{(d_1, e_1, f_1), (d_2, e_2, f_2)\}$

• And 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:

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

• \( \Omega_{ij} \) 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 \( \Omega_{i,j} \)?
Covariance Matrices (V)

Covariance matrices of 2D data sets (easy to draw)

What can you tell me about $\Omega_{x,y}$?

What can you tell me about $\Omega_{x,y}$?
Principal Component Analysis

- PCA ≡ SVD(Cov(X)) = SVD(XXᵀ/(n-1))
- SVD: XXᵀ = RΛR⁻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 2nd largest Eigenvalue is orthogonal to the 1st vector and has the next greatest variance.
- And so on...
- The Eigenvalues describe the amount of variance along the Eigenvectors.

We will now motivate and illustrate these ideas using Gaussian Random Variables …
The PCA Cookbook
Step 1: Image as Vector

\[
\begin{align*}
I(1,1) & \\
\vdots & \\
I(1,N) & \\
\vdots & \\
I(N,1) & \\
\vdots & \\
I(N,N) & \\
\end{align*}
\]
Step 2: Normalize

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

\[
X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad \overline{X} = \sum_{i=0}^{N} x_i, \quad X - \overline{X} = \begin{bmatrix} x_1 - \overline{x} \\ \vdots \\ x_N - \overline{x} \end{bmatrix}
\]
Step 3 (optional) : mean-center data set

• Form data matrix
  – Images (samples) as columns
    \[
    \begin{bmatrix}
    I_1 & I_2 & \ldots & I_K \\
    \vdots & \vdots & \ldots & \vdots \\
    \vdots & \vdots & \ldots & \vdots \\
    \end{bmatrix}
    \]
  – Subtract mean image from all columns
Step 4: Covariance of a Data Set

\[ \text{Cov} = XX^T \]

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

\[
\omega_{i,j} = \sum_{k}^{K} \left( x_{i}^{k} - \bar{x}^{k} \right) \left( x_{j}^{k} - \bar{x}^{k} \right)
\]
Step 5: PCA

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

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

$$\text{Cov}(X) = XX^T$$

$$\text{PCA}(X) = \text{SVD}(XX^T) = R^T \Lambda R$$
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
  \[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
  - \( \mu \) is the mean and \( \sigma \) 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.

\[ f(x, y) = \frac{1}{\sqrt{2\pi} \sigma_{xx}^{1/2}} e^{-\frac{1}{2} \begin{vmatrix} x-\mu_x \\ y-\mu_y \end{vmatrix}^T \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}^{-1} \begin{vmatrix} x-\mu_x \\ y-\mu_y \end{vmatrix}} \]
Special Case: Axis Aligned

\[
\sigma_{xx} = \sigma_x^2 \quad \sigma_{yy} = \sigma_y^2 \quad \sigma_{xy} = 0
\]

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

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

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

• Consider the following axis-aligned Gaussian:

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

\[
\sigma_x = 4, \quad \sigma_y = 2, \quad \mu_x = 3, \quad \mu_y = 5
\]
Quadratic Forms

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

\[
f(x, y) = V^T M V = \begin{vmatrix} x & y \\ b & c \end{vmatrix} \begin{vmatrix} a & b \\ b & c \end{vmatrix} = ax^2 + 2bxy + cy^2
\]

- Singular value decomposition tells us that:

\[
M = R\Delta R^{-1}
\]

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

- \(R\) rotates coordinates so \(M\) is diagonal.
Quadratic Forms Rotated

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

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

\[ V = R X \]

\[ \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \]

\[ V^T = (RX)^T \]

\[ \begin{bmatrix} u & v \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \]

\[ f(x, y) = X^T R^T D R X = X^T M X \]
Rotated Example

\[ f_1(u, v) = 8x^2 + 2y^2 \quad f_2(x, y) = 6.49x^2 - 5.20xy + 3.50y^2 \]

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