Intelligent Data Analysis: Principal Components Analysis (PCA)

Martin Russell

School of Computer Science, University of Birmingham

February 6, 2020

Overview

Covariance

$X = \{x^1, \dots, x^N\}$

The *sample covariance* for X between the i^{th} and j^{th} coordinates is given by

$$\sigma^{ij} = \frac{1}{N-1} \sum_{n=1}^{N} (\mu^i - x_n^i)(\mu^j - x_n^j) \tag{1}$$

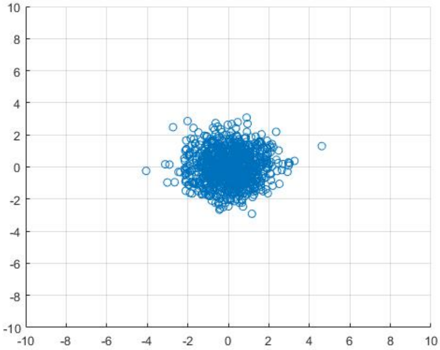
$i = j$. variance

$= \frac{1}{N-1} \sum (\mu^i - x_n^i)^2$

The sample covariance is represented naturally as a $D \times D$ **real symmetric** matrix Σ

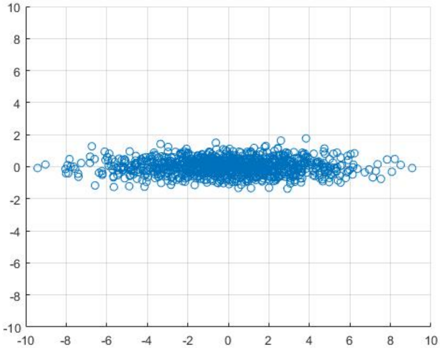
$$\Sigma = \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1D} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2D} \\ \vdots & \vdots & & \vdots \\ \sigma^{D1} & \sigma^{D2} & \cdots & \sigma^{DD} \end{bmatrix} \tag{2}$$

Example: 2D standard Gaussian



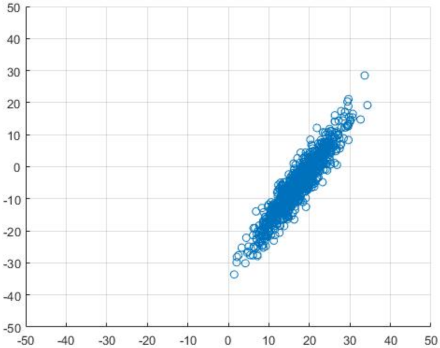
2D Gaussian, zero covariance, centre at origin $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Example: 2D elliptical



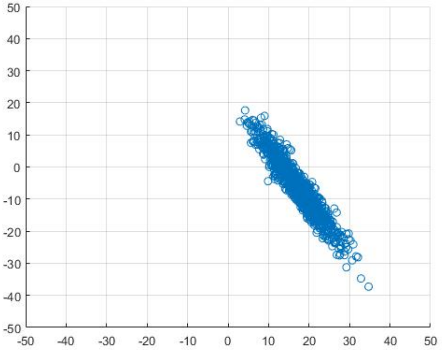
2D elliptical, zero covariance, centre at origin $\Sigma = \begin{bmatrix} 8.9 & 0 \\ 0 & 0.24 \end{bmatrix}$

Example: Positive covariance



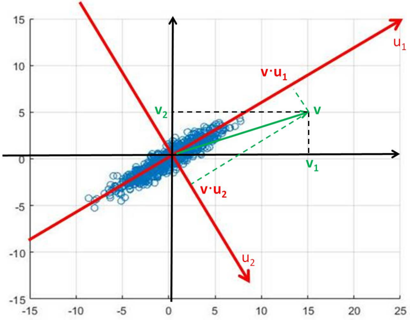
2D positive covariance, centre at $\begin{bmatrix} 17 \\ -5 \end{bmatrix}, \Sigma = \begin{bmatrix} 26.8 & 42.4 \\ 42.4 & 76.3 \end{bmatrix}$

Example: Negative covariance



2D negative covariance, centre at $\begin{bmatrix} 17 \\ -5 \end{bmatrix}, \Sigma = \begin{bmatrix} 27 & -43.8 \\ -43.8 & 78.34 \end{bmatrix}$

Motivation



- Positive covariance with respect to standard coordinates
- No covariance with respect to “red” coordinate system
- How can we identify the red coordinate system automatically?

Derivation of PCA

- Intuitively, we want direction of maximum variance
- For a potential first coordinate $\vec{u}$
  - Project data onto $\vec{u}$
  - Calculate variance in direction $\vec{u}$
  - Maximise with respect to $\vec{u}$ $\|\vec{u}\| = 1$
- Problem: $\vec{u}$ needs to be a **unit vector**
- Constrained optimization problem, soluble using Lagrange multipliers
- Solution: $\vec{u}$ is the eigenvector of covariance matrix corresponding to biggest eigenvalue
- This is the basis of **Principal Components Analysis (PCA)**

Principal Components Anlaysis (PCA)

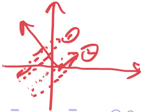
- PCA is a tool to reveal the structure of a data set $X \in \mathbb{R}^N$
- Apply **eigenvector decomposition** to covariance matrix C of X

$$C = UDU^T \tag{3}$$

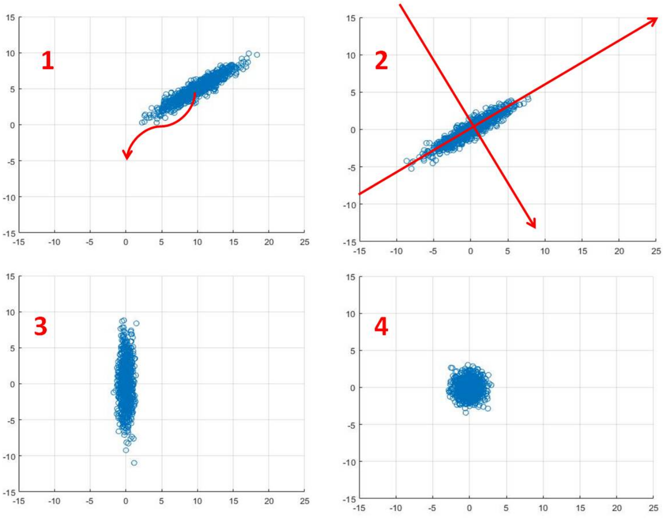
where D is a diagonal matrix with non-negative real values and U is an orthogonal matrix

  - Columns of U are a **new basis** $\{u_1, \cdots, u_N\}$ for $\mathbb{R}^N$. Basis vectors u_n point in directions of maximum variance of X
  - The ~~**eigenvector**~~ eigenvalue (element of D) d_n is the **variance** of the data in the direction u_n

often good for classification



Principal Components Anlaysis (PCA)

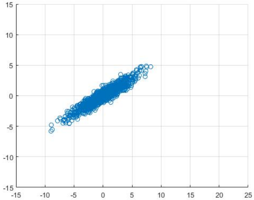


PCA Example - Step 1

- Calculate the sample mean m
- Subtract m from each data sample

$-\mu$

PCA Example - step 2

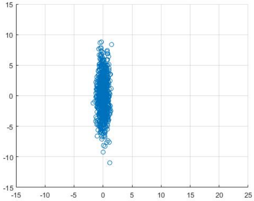


- Calculate the covariance matrix C
- Apply eigenvector decomposition to C

$$C = \begin{bmatrix} 6.56 & 3.64 \\ 3.64 & 2.35 \end{bmatrix}, C = UDU^T, \text{ where} \tag{4}$$

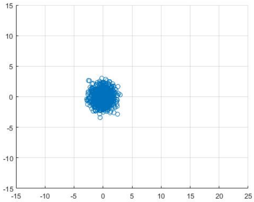
$$D = \begin{bmatrix} 0.25 & 0 \\ 0 & 8.67 \end{bmatrix}, U = \begin{bmatrix} 0.5 & -0.867 \\ -0.867 & -0.5 \end{bmatrix} \tag{5}$$

PCA Example - step 3



- Eigenvectors are new basis / coordinate system
- Apply orthogonal matrix U to change coordinate system
- (Alternatively, project data onto new basis vectors)
- Note orientation of the data - second (vertical) axis (second eigenvector) corresponds to biggest eigenvalue

PCA Example - step 4 (optional)



- Variance normalization (if required)
- Divide i^{th} coordinate of each data point by i^{th} eigenvalue
- Resulting data has covariance matrix I

Examples

- Example 1: Simple 2-dimensional data (can be done 'by hand')
- Example 2: 14-dimensional 'Boston' data from IDA(ext) assignment example - MATLAB
- Example 3: MEng Final Year Project - 90 dimensional dance data

Example 1: A simple example

- $X = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}$
- $m = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, C = cov(X) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
- To calculate eigenvalues, characteristic equation is $\lambda(\lambda - 2) = 0$. Hence $\lambda = 2$ or $\lambda = 0$ - just one eigenvalue
- For $\lambda = 2$ solve $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- Hence first unit basis vector is $u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$, and so $u_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

A simple example (continued)

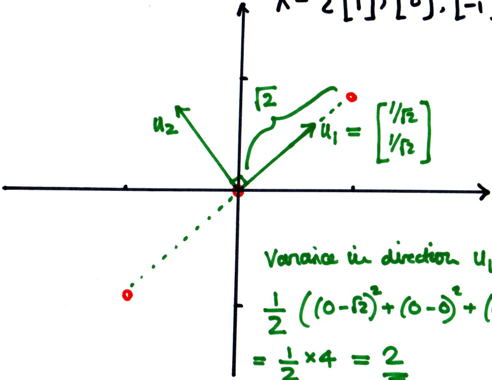
Therefore

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, U = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (6)$$

$$C = UDU^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (7)$$

A simple example (continued)

$$X = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}.$$



Variance in direction u_1 is

$$\frac{1}{2}\left((0-\sqrt{2})^2 + (0-0)^2 + (0--\sqrt{2})^2\right)$$

$$= \frac{1}{2} \times 4 = \underline{\underline{2}}.$$

Example 2: MATLAB example with 'boston' data

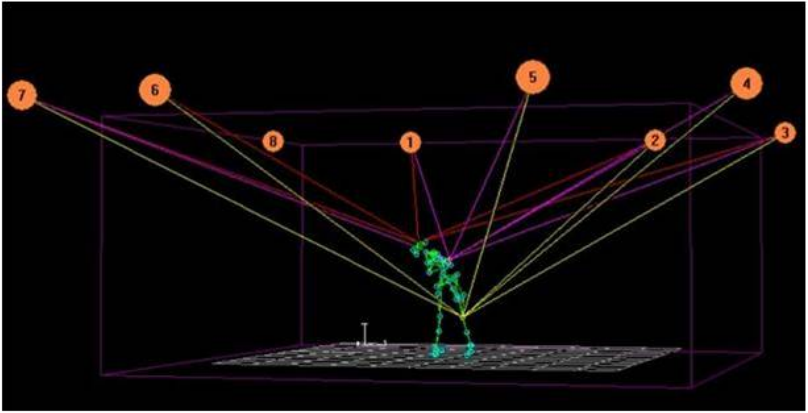
- 14-dimensional data set
- Social/demographic data for 506 Boston districts
- From MSc/MSci assignment example
- Focus on two parameters - 'price' and 'nox'

Example 3: Modelling 3D dance movement

- Results from MEng project 2003
- Analysis of 3 dimensional dance motion
- Pose represented as 90 dimensional vector - (x, y, z) coordinates of 30 critical body points
- Dance movement represented as a sequence of 90-dimensional vectors (poses)
- Intuitively , body motion representable with fewer parameters
- Redundancy in 90-dimensional representation?

Capturing pose - Qualysis 3D motion tracker

3D video motion tracking



Application of PCA

- Arrange data as a $T \times 90$ matrix X

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,90} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,90} \\ x_{3,1} & x_{3,2} & x_{3,3} & \cdots & x_{3,90} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{T,1} & x_{T,2} & x_{T,3} & \cdots & x_{T,90} \end{bmatrix} \tag{8}$$

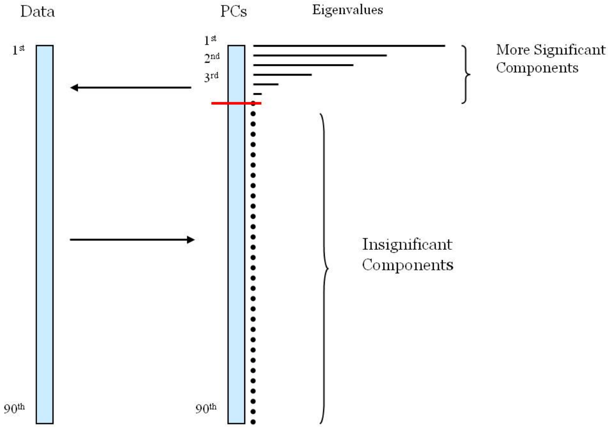
- Calculate the covariance matrix $C = cov(X)$ of X
- Calculate eigenvector decomposition $[U, D] = eig(C)$

$$C = UDU^T \tag{9}$$

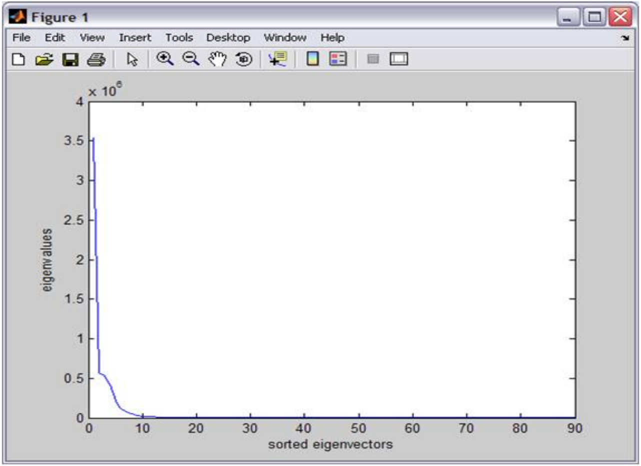
D is diagonal and U is an orthogonal (change of basis) matrix

Dimension reduction using PCA

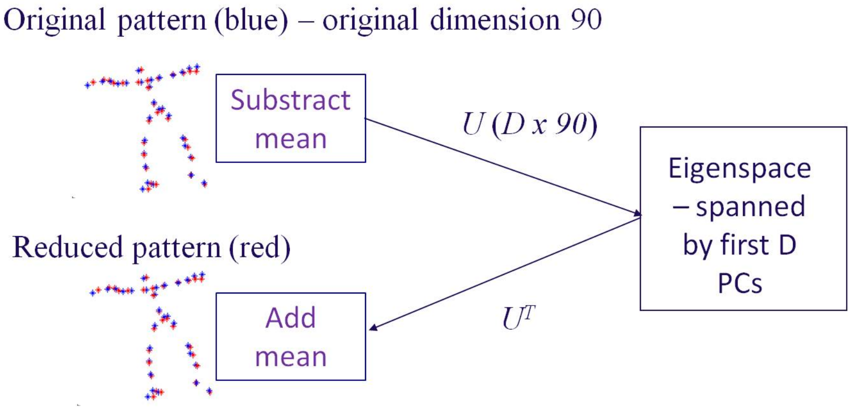
- Columns of U are new basis vectors (*principal vectors*)
- Eigenvalues are variances in directions of new basis vectors



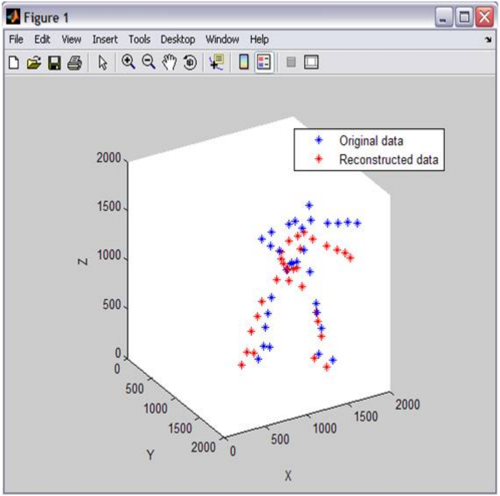
Eigenvalues



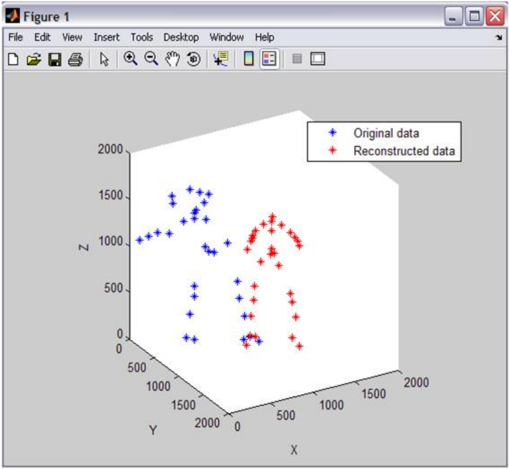
Visualizing effect of dimension reduction



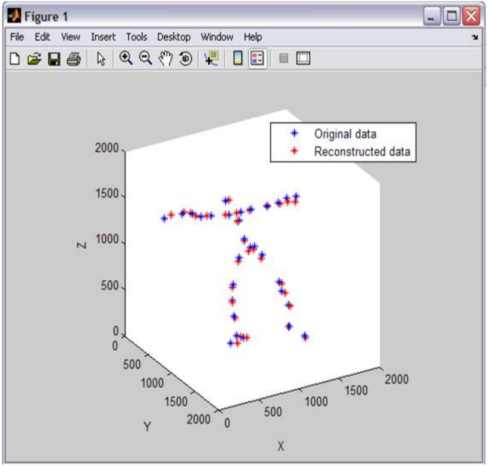
Visualizing PCA - 1D projection onto 1st PC



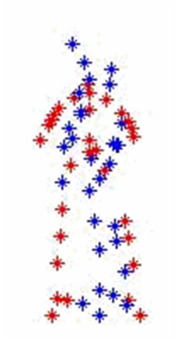
Visualizing PCA - 1D projection onto 10th PC



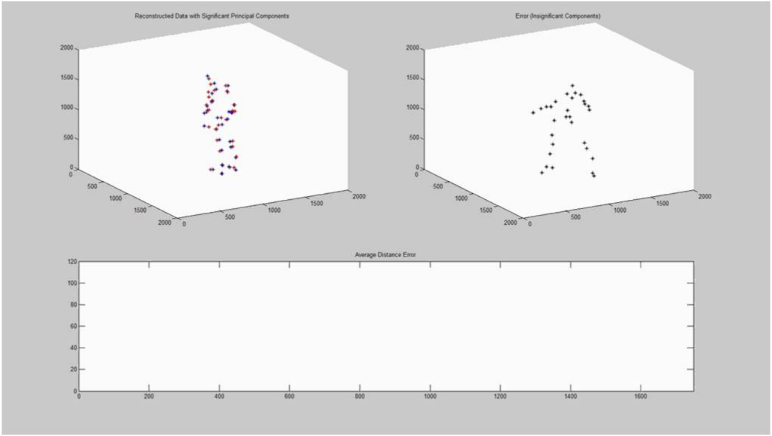
Visualizing PCA - 10D projection onto 1st-10th PCs



Visualizing PCA - 90D reduced to 1D - video



Visualizing PCA - 90D reduced to 10D - video



Applications of PCA

- Dimension reduction
- Visualization (reduction to 2 dimensions)
- De-correlation - diagonalization of the covariance matrix

Summary

- PCA doesn't change the data - only how we look at it
- For visualization, PCA finds the plane in which the variance of the data is maximized. This may not always be what we want.
- Suppose the data is in K **classes** and we want to separate them $\cdots$