Principal Components Analysis:

Hamid Dehghani
School of Computer Science
Birmingham
Feb 2024

UNIVERSITY OF BIRMINGHAM

Outline

- By the end of these series you will
  - Learn about relationship between multi-dimensional data
  - Understand Principal Components Analysis
  - Apply data compression

Covariance

- Variance and Covariance
  - measure of the "spread" of a set of points around their centre of mass (mean)

  - Variance
    - measure of the deviation from the mean for points in one dimension e.g. heights
  - Covariance
    - measure of how much each of the dimensions vary from the mean with respect to each other

Covariance

- Covariance
  - measured between 2 dimensions to see if there is a relationship between the 2 dimensions e.g. number of hours studied & marks obtained

$$\mathrm{cov}(x_1, x_2) = \frac{\sum_{i=1}^{n} (x_1^i - \bar{x}_1)(x_2^i - \bar{x}_2)}{n-1}$$

UNIVERSITY OF BIRMINGHAM

Covariance

- For a 3-dimensional data set (x,y,z)
  - measure the covariance between
    - x and y dimensions,
    - y and z dimensions.
    - x and z dimensions.

  - Measuring the covariance between
    - x and x
    - y and y
    - z and z
  - Gives you the variance of the x , y and z dimensions respectively

UNIVERSITY OF BIRMINGHAM

Covariance Matrix

- Representing Covariance between dimensions as a matrix e.g. for 3 dimensions:

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix}$$

Variances

- Diagonal is the variances of x, y and z
- cov(x,y) = cov(y,x) hence matrix is symmetrical about the diagonal
- N-dimensional data will result in NxN covariance matrix

Covariance

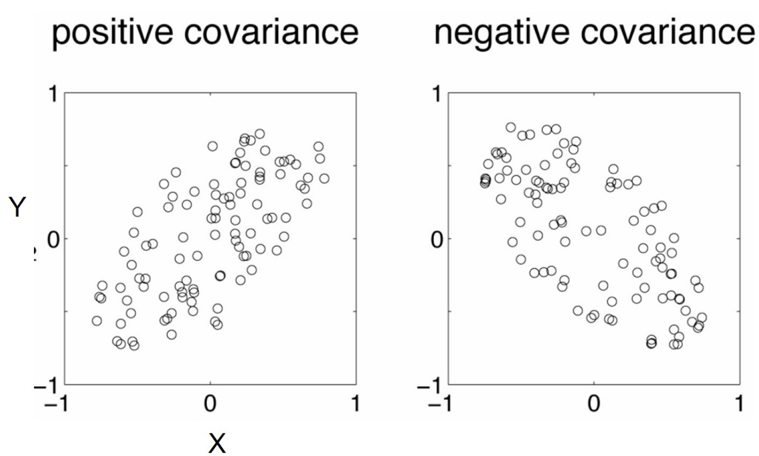
- What is the interpretation of covariance calculations?
  - e.g.: 2 dimensional data set
    - x: number of hours studied for a subject
    - y: marks obtained in that subject
  - covariance value is say: 104.53

- What does this value mean?

- Exact value is not as important as its sign.

Covariance examples

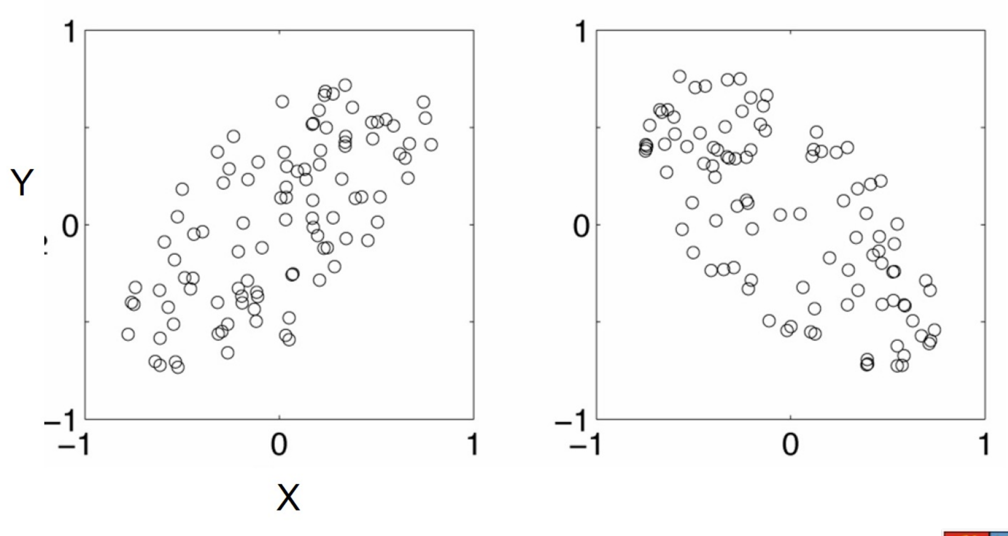
- A positive value of covariance
  - Both dimensions increase or decrease together e.g. as the number of hours studied increases, the marks in that subject increase.
- A negative value
  - while one increases the other decreases, or vice-versa e.g. Hours awake vs performance in Final Assessment.
- If covariance is zero:
  - the two dimensions are independent of each other e.g. heights of students vs the marks obtained in the Quiz



UNIVERSITY OF BIRMINGHAM

Why is it interesting

- Why bother with calculating covariance when we could just plot the 2 values to see their relationship?
  - Covariance calculations are used to find relationships between dimensions in high dimensional data sets (usually greater than 3) where visualization is difficult.



UNIVERSITY OF BIRMINGHAM

Principal components analysis (PCA)

- PCA is a technique that can be used to simplify a dataset
  - A linear transformation that chooses a new coordinate system for the data set such that:
  - greatest variance by any projection of the data set comes to lie on the first axis (then called the first principal component),
  - the second greatest variance on the second axis,
  - and so on.
- PCA can be used for reducing dimensionality by eliminating the later (none substantial principal components.

PCA: Simple example

- Consider the following 3D points in space (x,y,z)
  - P1 = [1 2 3]
  - P2 = [2 4 6]
  - P3 = [4 8 12]
  - P4 = [3 6 9]
  - P5 = [5 10 15]
  - P6 = [6 12 18]
- To store each point in the memory, we need
  - 18 = 3 x 6 bytes

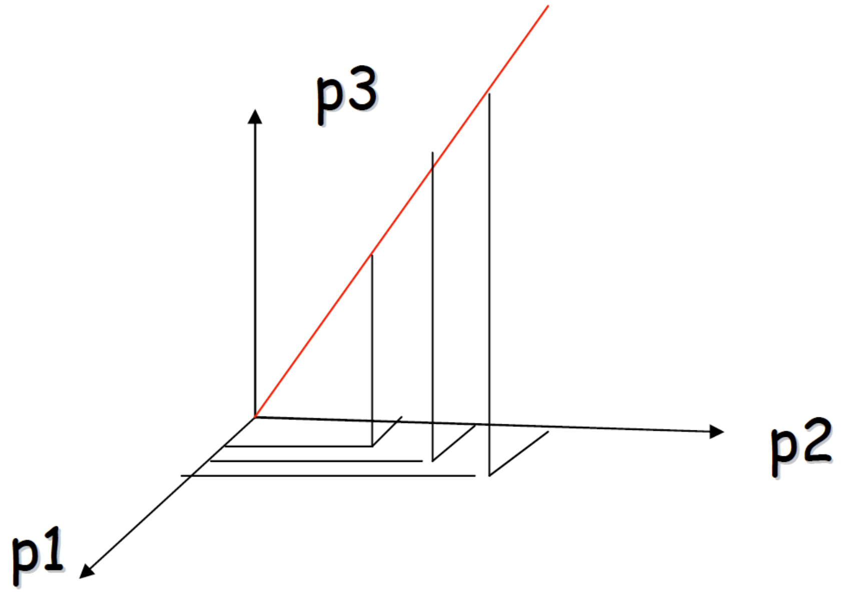
PCA: Simple example

- But
  - P1 = [1 2 3] = P1 × 1
  - P2 = [2 4 6] = P1 × 2
  - P3 = [4 8 12] = P1 × 4
  - P4 = [3 6 9] = P1 × 3
  - P5 = [5 10 15] = P1 × 5
  - P6 = [6 12 18] = P1 × 6
- All the points are related geometrically: they are all the same point, scaled by a factor
- They can be stored using only 9 bytes
  - Store one point (3 bytes) + the multiplying constants (6 bytes)

UNIVERSITY OF BIRMINGHAM

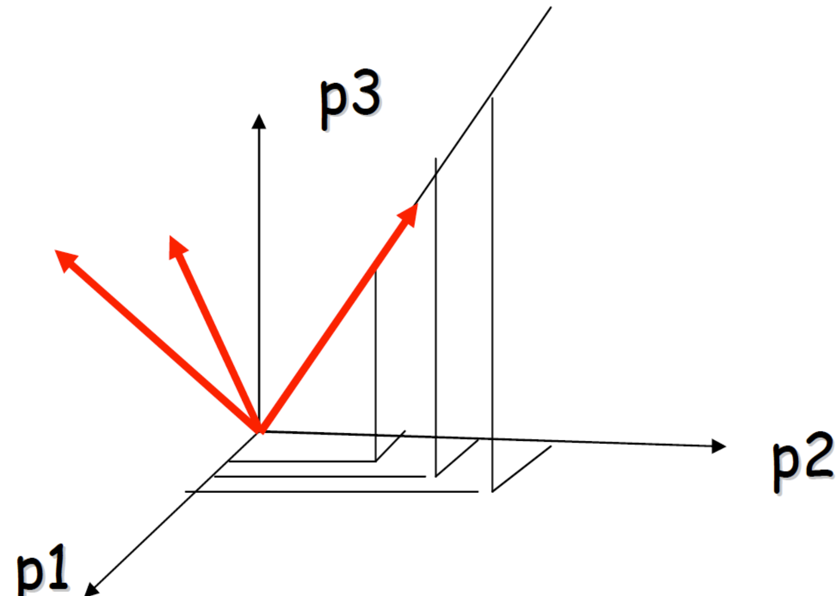
PCA: Simple example

- Viewing the points in 3D
- this example
  - all the points happen to belong to a line:
  - a 1D subspace of the original 3D space



PCA: Simple example

- Viewing the points in 3D
  - Consider a new coordinat system where one of the axes is along the direction of the line
  - In this coordinate system, every point has only one non-zero coordinate: we only need to store the direction of the line and the nonzero coordinate for each of the points.



UNIVERSITY OF BIRMINGHAM

Principal Component Analysis (PCA)

- Given a set of points, how do we know if they can be compressed like in the previous example?
- The answer is to look into the correlation between the points

- The tool for doing this is called PCA

PCA

- What is the principal component.
  - By finding the eigenvalues and eigenvectors of the covariance matrix, we find that the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset.
- PCA is a useful statistical technique that has found application in:
  - fields such as face recognition and image compression
  - finding patterns in data of high dimension.

Basic Theory

- Let x_i be a set of 'n' N × 1 numbers

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iN} \end{bmatrix} \qquad \bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{i=n} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iN} \end{bmatrix}$$

UNIVERSITY OF BIRMINGHAM

Basic Theory

- Let X be the [N $\times$ n] matrix with rows

$$X = \left[\begin{array}{cccc} \mathbf{x}_1 - \bar{\mathbf{x}} & \mathbf{x}_2 - \bar{\mathbf{x}} & \cdots & \mathbf{x}_n - \bar{\mathbf{x}} \end{array} \right]$$

- Subtracting the mean is equivalent to translating the coordinate system to the location of the mean.

UNIVERSITY OF BIRMINGHAM

Basic Theory

- Let $Q = XX^T$ be the [N × N] matrix:

$$Q = XX^T = \left[\begin{array}{cccc} \mathbf{x}_1 - \bar{\mathbf{x}} & \mathbf{x}_2 - \bar{\mathbf{x}} & \cdots & \mathbf{x}_n - \bar{\mathbf{x}} \end{array} \right] \left[\begin{array}{c} (\mathbf{x}_1 - \bar{\mathbf{x}})^T \\ (\mathbf{x}_2 - \bar{\mathbf{x}})^T \\ \vdots \\ (\mathbf{x}_n - \bar{\mathbf{x}})^T \end{array} \right]$$

- Q is square
- Q is symmetric
- Q is the covariance matrix [aka scatter matrix]
- Q can be very large (in vision, N is often the number of pixels in an image!)

UNIVERSITY OF BIRMINGHAM

PCA Theorem

- Each x_j can be written as:

$$\mathbf{x}_j = \bar{\mathbf{x}} + \sum_{i=1}^{i=n} g_{ji}\mathbf{e}_i$$

- where e_i are the n eigenvectors of Q with non-zero eigenvalues
- Remember
  - The eigenvectors $e_1\ e_2\ \ldots\ e_n$ span an eigenspace
  - $e_1\ e_2\ \ldots\ e_n$ are N $\times$ 1 orthonormal vectors (directions in N-Dimensional space)
  - The scalars g_{ji} are the coordinates of x_j in the space.

$$g_{ji} = (\mathbf{x}_j - \bar{\mathbf{x}}).\mathbf{e}_i$$

UNIVERSITY OF BIRMINGHAM

Using PCA to Compress Data

- Expressing x in terms of e_1 e_2 ... e_n has not changed the size of the data
- BUT:
  - if the points are highly correlated many of the coordinates of x will be zero or closed to zero.
  - This means they lie in a lower-dimensional linear subspace
- Sort the eigenvectors e_i according to their eigenvalue:

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$$

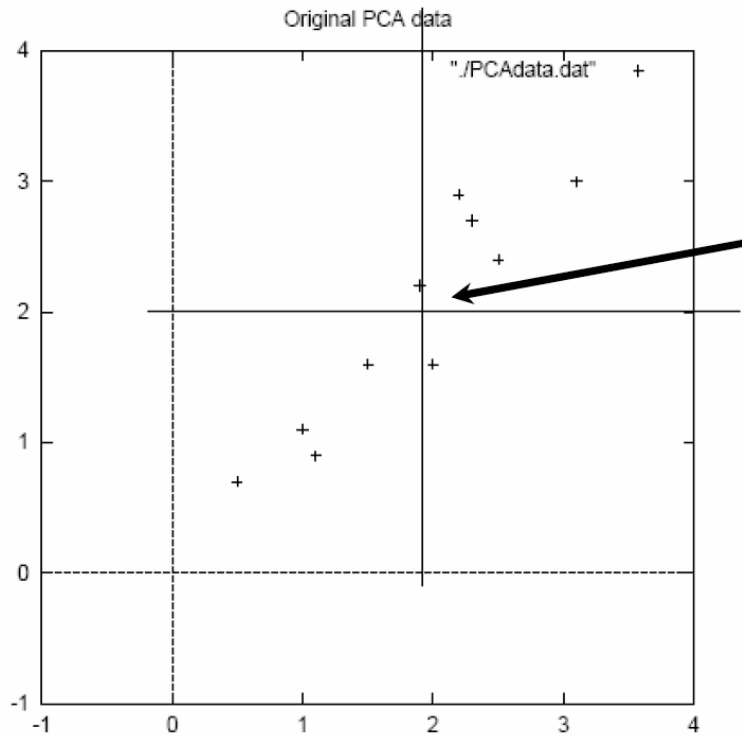
- Assuming that: $\lambda_i \approx 0 \text{ if } i > k$
- Then: $$\mathbf{x}_j \approx \bar{\mathbf{x}} + \sum_{i=1}^{i=k} g_{ji}\mathbf{e}_i$$

UNIVERSITY OF BIRMINGHAM

Example –STEP 1

DATA:

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9



mean

this becomes the new origin of the data from now on

UNIVERSITY OF BIRMINGHAM

Example –STEP 2

- Calculate the covariance matrix

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

- Since the non-diagonal elements in this covariance matrix are positive, we should expect that both the x and y variable increase together.
- Calculate the eigenvectors and eigenvalues of the covariance matrix

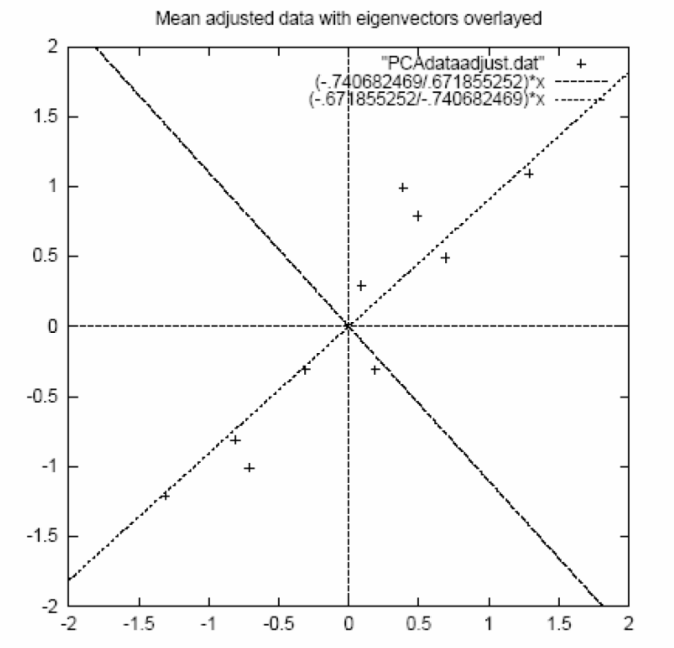
$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -735178656 \end{pmatrix}$$

UNIVERSITY OF BIRMINGHAM

Example –STEP 3

- eigenvectors are plotted as diagonal dotted lines on the plot.
- Note they are perpendicular to each other.
- Note one of the eigenvectors goes through the middle of the points, like drawing a line of best fit.
- The second eigenvector gives us the other, less important, pattern in the data, that all the points follow the main line, but are off to the side of the main line by some amount.



UNIVERSITY OF BIRMINGHAM

Example –STEP 4

- Feature Vector
  – Feature Vector = $(eig_1\ eig_2\ eig_3\ \ldots\ eig_n)$
- We can either form a feature vector with both of the eigenvectors:

$$\begin{pmatrix} -.677873399 & -.735178656 \\ -.735178656 & .677873399 \end{pmatrix}$$

- or, we can choose to leave out the smaller, less significant component and only have a single column:

$$\begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix}$$

Example –STEP 5

- Deriving new data coordinates
  - FinalData = RowZeroMeanData $\times$ RowFeatureVector
    - RowFeatureVector is the matrix with the eigenvectors with the most significant eigenvector first
    - RowZeroMeanData is the mean-adjusted data, i.e. the data items are in each column, with each row holding a separate dimension.
- Note: This is essentially rotating the coordinate axes so higher-variance axes come first

Event Code:

UNIVERSITY OF
BIRMINGHAM

Implementing PCA

- Finding the 'first k eigenvectors of Q:

$$Q = XX^T = \left[\begin{array}{cccc} \mathbf{x}_1 - \bar{\mathbf{x}} & \mathbf{x}_2 - \bar{\mathbf{x}} & \cdots & \mathbf{x}_n - \bar{\mathbf{x}} \end{array}\right] \left[\begin{array}{c} (\mathbf{x}_1 - \bar{\mathbf{x}})^T \\ (\mathbf{x}_2 - \bar{\mathbf{x}})^T \\ \vdots \\ (\mathbf{x}_n - \bar{\mathbf{x}})^T \end{array}\right]$$

- Q is N $\times$ N (N could be the number of pixels in an image. For a 256 x 256 image, N = 65536 !!)
- Don't want to explicitly compute Q!!!!

UNIVERSITY OF BIRMINGHAM

Singular Value Decomposition (SVD)

- Any $m \times n$ matrix X can be written as the product of 3 matrices:

$$X = UDV^T$$

- where:
  - U is m x m and its columns are orthonormal vectors
  - V is n x n and its columns are orthonormal vectors
  - D is m x n diagonal and its diagonal elements are called the singular values of X, and are such that:
  - s1>s2>s3>…>0

SVD Properties

$$X = UDV^T$$

- The columns of U are the eigenvectors of XX^T
- The columns of V are the eigenvectors of X^TX
- The diagonal elements of D are the eigenvalues of XX^T and X^TX

UNIVERSITY OF BIRMINGHAM