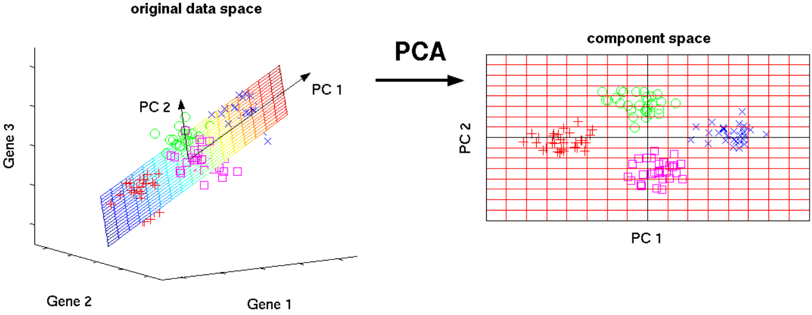
Rozpoznávanie obrazcov - 4. cvičenie
Redukcia dimenzionality

Viktor Kocur
viktor.kocur@fmph.uniba.sk

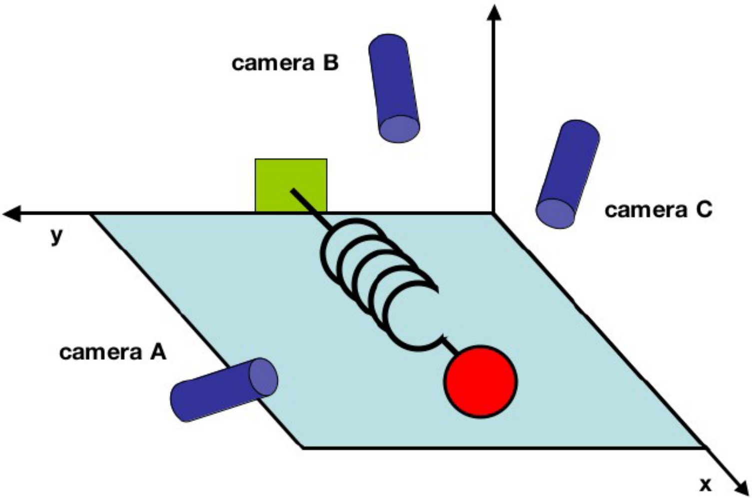
DAI FMFI UK

19.3.2019

PCA princíp


original data space
PC 1
PC 2
Gene 3
Gene 2
Gene 1
PCA
component space
PC 2
PC 1

PCA motivácia



PCA matematika

Eigenvectors and eigenvalues

Let $\mathbb{A}$ be a matrix $n \times n$, then a nonzero vector $\vec{v} \in \mathbb{R}^n$ is an eigenvector of $\mathbb{A}$ with eigenvalue $\lambda \in \mathbb{C}$ if: $\mathbb{A}\vec{v} = \lambda\vec{v}$.

Hermitian Matrix

Hermitian matrices have only real eigenvalues. It is also possible to find an eigenvector for each eigenvalue.

PCA theory

Covariance

$cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}$

Covariance matrix

$COV(X)_{i,j} = cov(X_i, X_j)$

Covariance matrix

Covariance matrix is Hermitian and positively semi-definite.

PCA theory

Matrix of eigenvectors

Z normalizovaných vlastných čísel $v_1 ... v_n$ zostavíme maticu $(v_1, v_2, ..., v_n)$. Značíme $\mathbb{W}$.

PCA

PCA is based on the fact that eigenvectors of covariance matrix represent an orthogonal base in which the covariance matrix is diagonal.

PCA process

We start by centering our data as $\vec{x}' = \vec{x} - \overline{x}$. Then we compute $\mathbb{W}$. The new data are obtained by matrix multiplication $y = \vec{x}'\mathbb{W}^T$, if $\vec{x}$ is a row vector. Eigenvalues correspond to the portion of the variance explained by given direction.

Matlab

Loading data

load data.mat

Exercise

Display 2D data in a plot.

Matlab functions

cov

cov(A) - returns covariance matrix of A

eig

[W, vals] $=$ eig(A) - return matrix W with eigenvectors W and eigenvalues on the diagonals of matrix vals for matrix A.

Exercise

Apply PCA on the data and plot the transformed data. Show how much variance is explained by each direction.

Solution

```
load data.mat
plot(data(:,1), data(:,2), 'r*');
centered = data - mean(data);
[W, eigenvals] = eig(cov(centered))
newdata = centered * W'
plot(newdata(:,1), newdata(:,2), 'r*');
ylim([-2 2]);
xlim([-2 2]);
disp(diag(eigenvals)/sum(diag(eigenvals)))
```

Matlab - PCA function

cov

[coeff,score, , ,explained,mu] = pca(X) - returns transformation matrix coeff (we denote as $\mathbb{W}^T$), transformed data in score, percentages of variance explained by different directions and mean values of X.

Transformation

score == (X - mu) * coeff

Inverse transformation

X == score * coeff' + mu

Exercise

Test this function on data.mat.

Matlab - PCA

Data

load ovariancancer

gscatter

gscatter(obs(:,1), obs(:,2), grp) - displays points from the first two columns of the dataset and shows the color based on grp class.

Exercise

Determine how many features (dimensions after transformation) do you need to keep at least 95 percent of variance of the dataset. What does it mean if some eigenvalues are zeros?

Exercise

Show the plot for 2 most significant dimensions of the data after PCA using gscatter.

PCA - exercises

Exercise

For data from data.mat show the directions to which we project the data in the original coordinate system.

Exercise

Create a function, which takes as an input a color image and takes values of RGB channels as features then performs PCA on these data and sets the last values for the (two) least significant transformed feature to zero and converts the image back to RGB.

Exercise

For the PCA in the second exercise show an image containing all of the colors, which can be used in this representation.

LDA

LDA.m

[Y, W, lambdas] = LDA(X, T) - for data X and classes T retursn new values Y after transformation with matrix W and values of eigenvalues in lambdas.

Note

Y == X * W

Exercise

Load the Fisher database (load fisheriris). Compare the data after using PCA and LDA.

Exercise

Plot the directions into which the data are projected by LDA for every pair of original columns.

LDA

LDA.m

MdlLinear = fitcdiscr(X,T) - returns a linear classifier, which uses LDA. If you apply LDA in your project you can use this.

Tutorial

https://www.mathworks.com/help/stats/create-and-visualize-discriminant-analysis-classifier.html