

# TDA 231

## Dimension Reduction: PCA

Devdatt Dubhashi  
dubhashi@chalmers.se

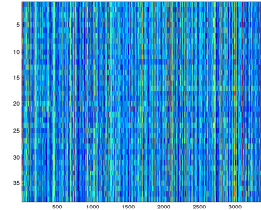
Department of Computer Science and Engg.  
Chalmers University

March 3, 2017

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## A problem - too many features

- ▶ Aim: To build a classifier that can diagnose leukaemia using Gene expression data.
- ▶ Data: 27 healthy samples, 11 leukaemia samples ( $N = 38$ ). Each sample is the expression (activity) level for 3751 genes. (Also have an independent test set)



- ▶ In general, the number of parameters will increase with the number of **features** –  $D = 3751$ .
  - ▶ e.g. Logistic regression –  $\mathbf{w}$  would have length 3751!
- ▶ Fitting lots of parameters is hard – imagine Metropolis-Hastings in 3751 dimensions rather than 2!

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Features

- ▶ For visualisation, most examples we've seen have had only 2 features  $\mathbf{x} = [x_1, x_2]^T$ .
- ▶ We sometimes **created** more:  $\mathbf{x} = [1, x_1 x_1^2, x_1^3, \dots]^T$ .
- ▶ Now, we've been given lots (3751) to start with.
- ▶ We need to reduce this number.
- ▶ 2 general schemes:
  - ▶ Use a **subset** of the originals.
  - ▶ Make new ones by **combining** the originals.

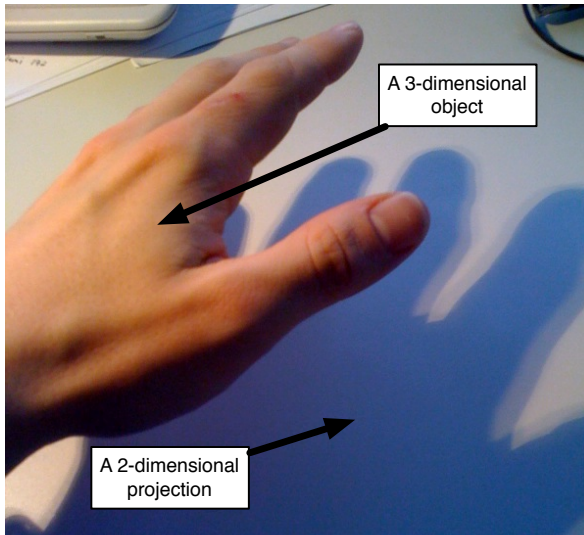
- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Making new features

- ▶ An alternative to choosing features is making new ones.
- ▶ Cluster:
  - ▶ Cluster the features (turn our clustering problem around)
  - ▶ If we use say K-means, our new features will be the  $K$  mean vectors.
- ▶ Projection/combination
  - ▶ Reduce the number of features by projecting into a lower dimensional space.
  - ▶ Do this by making new features that are combinations (linear) of the old ones.

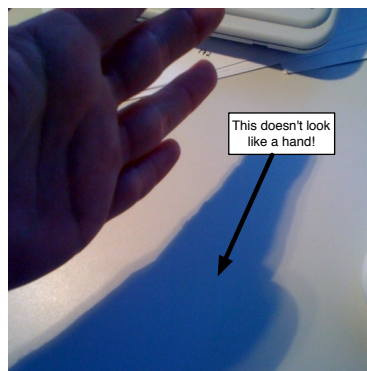
- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Projection



## Choosing $\mathbf{W}$

- ▶ Different  $\mathbf{W}$  will give us different projections (imagine moving the light).
- ▶ Which should we use?
- ▶ Not all will represent our data well...



Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

## Projection

- ▶ We can project data ( $D$  dimensions) into a lower number of dimensions ( $M$ ).
- ▶  $\mathbf{Z} = \mathbf{XW}$ 
  - ▶  $\mathbf{X}$  is  $N \times D$
  - ▶  $\mathbf{W}$  is  $D \times M$
- ▶  $\mathbf{Z}$  is  $N \times M$  – an  $M$ -dimensional representation of our  $N$  objects.
- ▶  $\mathbf{W}$  defines the projection
  - ▶ Changing  $\mathbf{W}$  is like changing where the light is coming from for the shadow (or rotating the hand).
  - ▶ ( $\mathbf{X}$  is the hand,  $\mathbf{Z}$  is the shadow)
- ▶ Once we've chosen  $\mathbf{W}$  we can project test data into this new space too:  $\mathbf{Z}_{\text{new}} = \mathbf{X}_{\text{new}}\mathbf{W}$

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

## Principal Components Analysis

- ▶ Principal Components Analysis (PCA) is a method for choosing  $\mathbf{W}$ .
- ▶ It finds the columns of  $\mathbf{W}$  one at a time (define the  $m$ th column as  $\mathbf{w}_m$ ).
  - ▶ Each  $D \times 1$  column defines one new dimension.
- ▶ Consider one of the new dimensions (columns of  $\mathbf{Z}$ ):

$$\mathbf{z}_m = \mathbf{X}\mathbf{w}_m$$

- ▶ PCA chooses  $\mathbf{w}_m$  to maximise the variance of  $\mathbf{z}_m$

$$\frac{1}{N} \sum_{n=1}^N (z_{mn} - \mu_m)^2, \quad \mu_m = \frac{1}{N} \sum_{n=1}^N z_{mn}$$

- ▶ Once the first one has been found, the  $\mathbf{w}_2$  is found that maximises the variance and is **orthogonal** to the first one etc etc.

Introduction

D. Dubhashi

Introduction

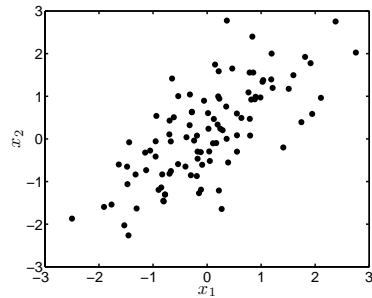
Features

Projections

PCA

ICA

## PCA – a visualisation



- ▶ Original data in 2-dimensions.
- ▶ We'd like a 1-dimensional projection.

Introduction

D. Dubhashi

Introduction

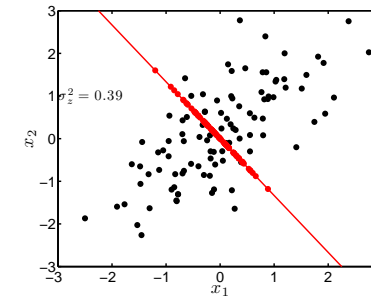
Features

Projections

PCA

ICA

## PCA – a visualisation



- ▶ Pick some arbitrary  $\mathbf{w}$ .
- ▶ Project the data onto it.
- ▶ Compute the variance (on the line).
- ▶ The position on the line is our 1 dimensional representation.

Introduction

D. Dubhashi

Introduction

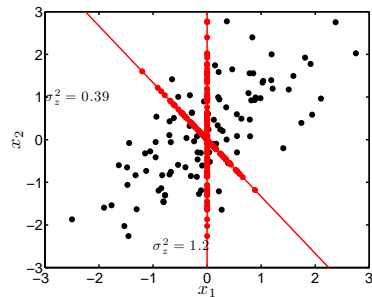
Features

Projections

PCA

ICA

## PCA – a visualisation



- ▶ Pick some arbitrary  $\mathbf{w}$ .
- ▶ Project the data onto it.
- ▶ Compute the variance (on the line).
- ▶ The position on the line is our 1 dimensional representation.

Introduction

D. Dubhashi

Introduction

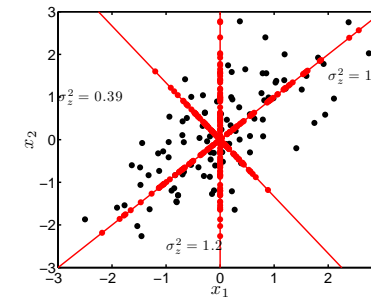
Features

Projections

PCA

ICA

## PCA – a visualisation



- ▶ Pick some arbitrary  $\mathbf{w}$ .
- ▶ Project the data onto it.
- ▶ Compute the variance (on the line).
- ▶ The position on the line is our 1 dimensional representation.

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

## PCA – analytic solution

- ▶ Could search for  $\mathbf{w}_1, \dots, \mathbf{w}_M$
- ▶ But, analytic solution is available.
- ▶  $\mathbf{w}$  are the **eigenvectors** of the covariance matrix of  $\mathbf{X}$ .
- ▶ Matlab: `princomp(x)`

Introduction

D. Dubhashi

Introduction

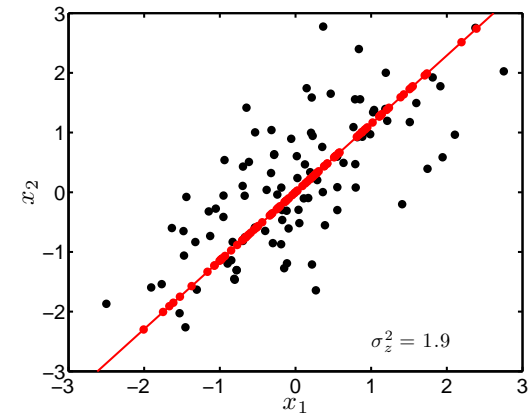
Features

Projections

PCA

ICA

## PCA – analytic solution



- ▶ What would be the second component?

Introduction

D. Dubhashi

Introduction

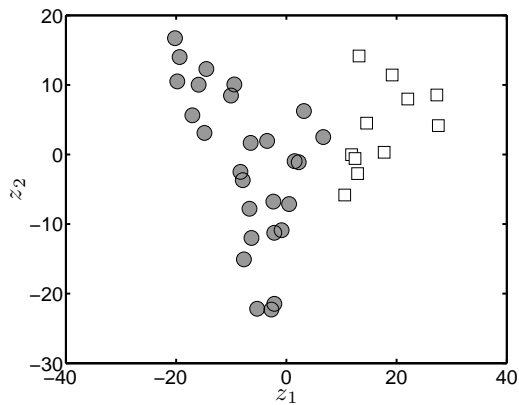
Features

Projections

PCA

ICA

## PCA – leukaemia data



First two principal components in our leukaemia data (points labeled by class).

Introduction

D. Dubhashi

Introduction

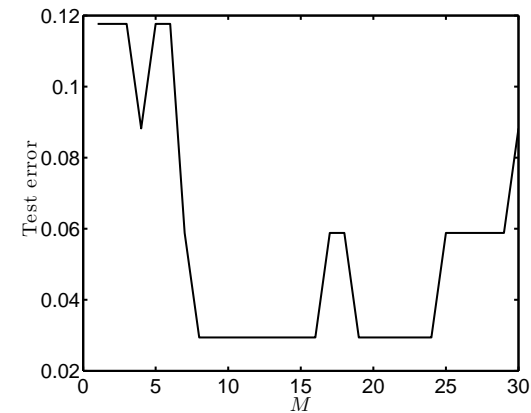
Features

Projections

PCA

ICA

## PCA – leukaemia data



Test error as more and more components are used.

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

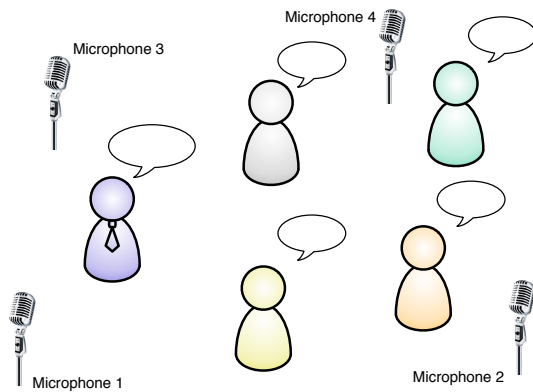
ICA

## Summary

- ▶ Sometimes we have too much data (too many dimensions).
- ▶ Features can be dimensions that already exist.
- ▶ Or we can make new ones.

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## The cocktail party problem



- ▶ Each microphone will record a combination of all speakers.
- ▶ Can we separate them back out again?

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Part 2: ICA (the cocktail party problem)

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Demo

- ▶ Online:
  - ▶ [http://www.cis.hut.fi/projects/ica/cocktail/cocktail\\_en.cgi](http://www.cis.hut.fi/projects/ica/cocktail/cocktail_en.cgi)
- ▶ Matlab:
  - ▶ Available on course webpage
  - ▶ To run:
    - ▶ `load ica_demo.mat`
    - ▶ `ica_image`

- Introduction
- D. Dubhashi
- Introduction
- Features
- Projections
- PCA
- ICA

## Independent components analysis – how it works...

- ▶ Corrupted data (images/sounds) is a vector of  $D$  numbers. i.e.  $n$ th image:

$$\mathbf{x}_n$$

- ▶ We have  $\mathbf{N}$  images – stack them up into an  $N \times D$  matrix:

$$\mathbf{X}$$

- ▶ Assume that this is the result of the following corrupting process:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E}$$

- ▶  $\mathbf{A}$  is mixing matrix.  $\mathbf{E}$  is noise. ( $\mathbf{S}$  is  $N \times D$ ).

$$e_{nd} \sim \mathcal{N}(0, \sigma^2)$$

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

## Inference

- ▶ From Bayes' (look back...)

$$p(\mathbf{S}|\mathbf{X}, \mathbf{A}, \sigma^2) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \sigma^2)p(\mathbf{S})$$

- ▶ In our demo, we found values of  $\mathbf{S}$ ,  $\mathbf{A}$  and  $\sigma^2$  that maximised the log posterior.
- ▶ MAP solution...
- ▶ There is some further reading on the webpage if you want to know more...

Introduction

D. Dubhashi

Introduction

Features

Projections

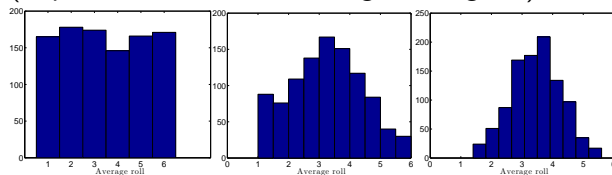
PCA

ICA

## Aside – ICA and the central limit theorem

- ▶ Central limit theorem (paraphrased):
  - ▶ If we keep adding the outcomes of independent random variables together, we eventually get something that looks Gaussian.

- ▶ Example: Roll a die  $m$  times and take the average. (Repeat this lots of times to get histogram)



- ▶ From left to right:  $m = 1$ ,  $m = 2$ ,  $m = 5$ . Looking more Gaussian as  $m$  increases.

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

## Aside – ICA and the central limit theorem

- ▶ Sometimes ICA is performed by **reversing** this theorem:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E}$$

- ▶  $\mathbf{X}$  is some random variables added together.
- ▶ It will be more 'Gaussian' than  $\mathbf{S}$
- ▶ Find  $\mathbf{S}$  that is as non-Gaussian as possible.

- ▶ More resource:

- ▶ <http://www.cis.hut.fi/projects/ica/icademo/>
- ▶ <http://www.cis.hut.fi/projects/ica/>

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA

# Summary

- ▶ PCA and ICA are both examples of projection techniques.
- ▶ Both assume a linear transformation
  - ▶ ICA:  $\mathbf{X} = \mathbf{AS} + \mathbf{E}$
  - ▶ PCA:  $\mathbf{Z} = \mathbf{XW}$
- ▶ PCA can be used for Data pre-processing or visualisation.
- ▶ ICA can be used to separate sources that have been mixed together.
- ▶ Also looked at PCA as a feature selection method.

Introduction

D. Dubhashi

Introduction

Features

Projections

PCA

ICA