Principal Component Analysis

Principal Component Analysis

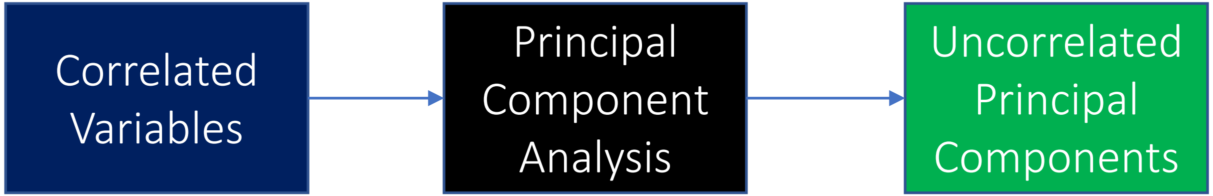
Introduction

- Problem
  - Data with many variables – curse of dimensionality
  - Finding patterns between observations hard
  - Visualisation not possible
- Main objective
  - Project higher dimensional data to lower dimensional space (reduce number of variables)
  - Visualise principal components (2D or 3D)

- PCA
  - Unsupervised learning technique
  - Most popular
  - Similar to clustering

Principal Component Analysis

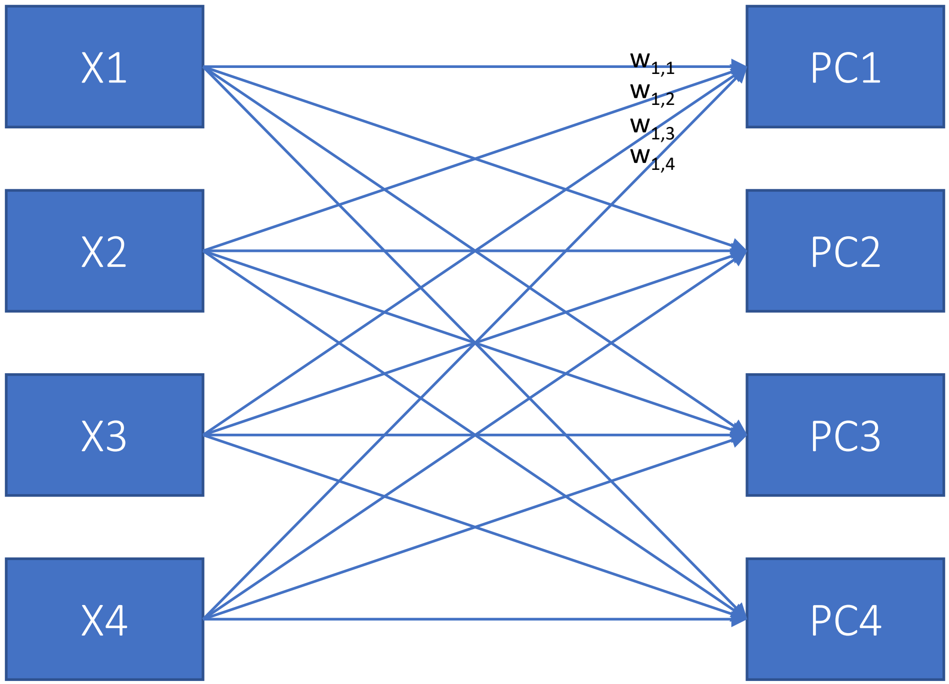
Principle



- Required: correlated variables
- PCA transforms correlated variables into uncorrelated principal components
- Principal components not interpretable

Principal Component Analysis

Principle



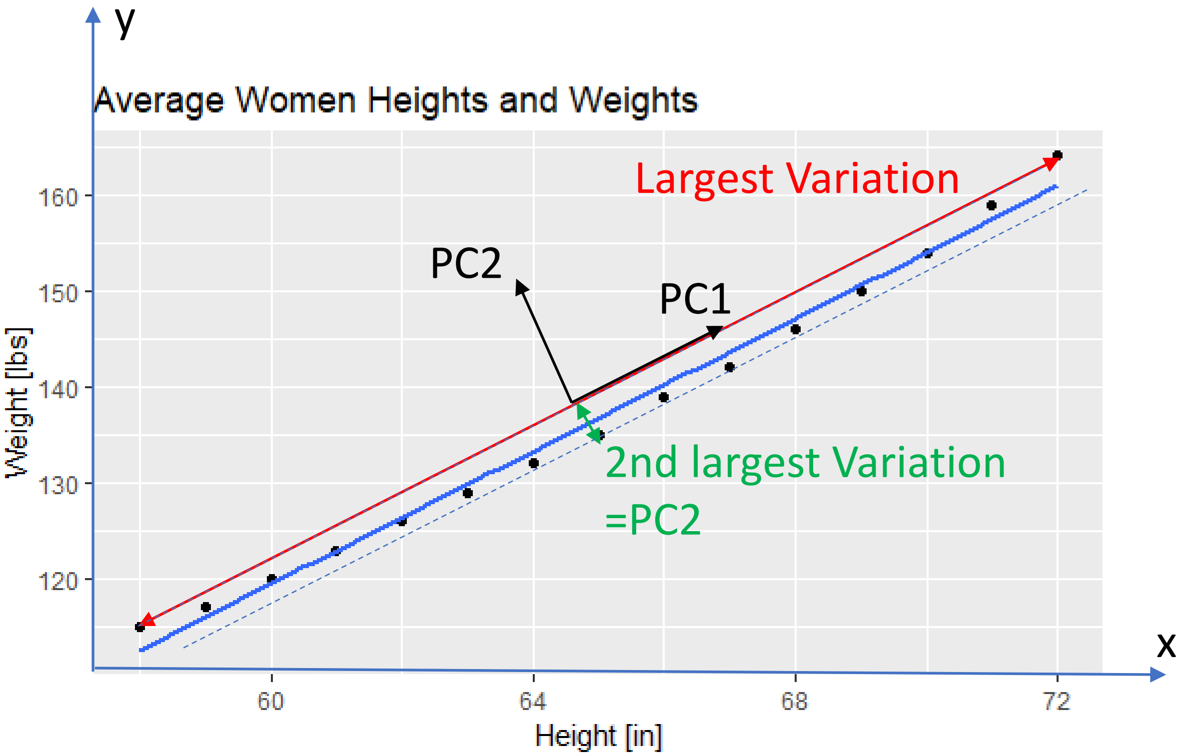
Principal Component Analysis

Example: 2D to 1D

Dataset: Women

- Highly correlated
- PC1 explains 99.8 % of total variance

→ Reduction from two to one dimensions possible



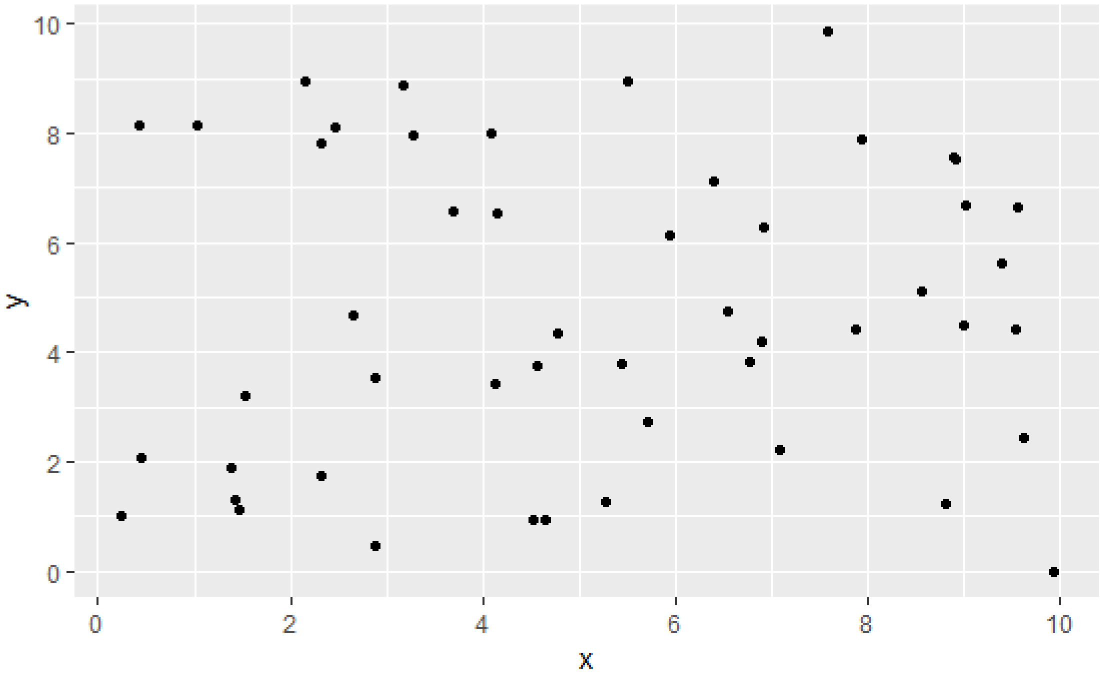
Principal Component Analysis

Example: 2D to 1D

Random Data Points

- uncorrelated
- PC1 and PC2 have similar proportion of variance

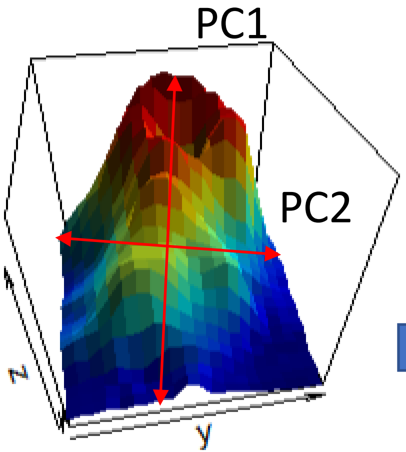
→ No reduction of dimensions possible

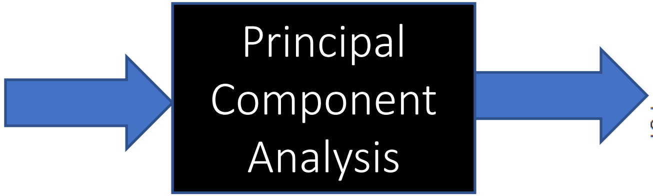


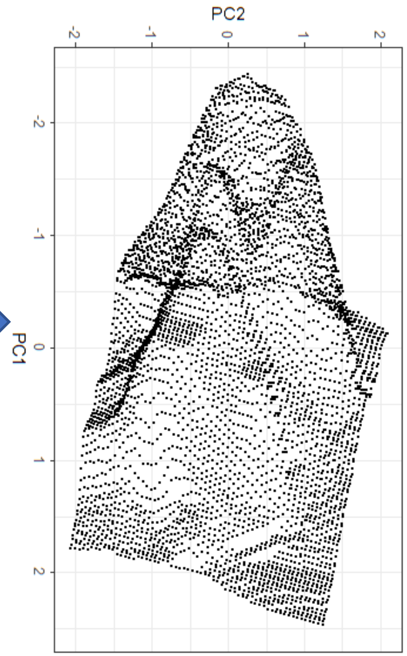
Principal Component Analysis

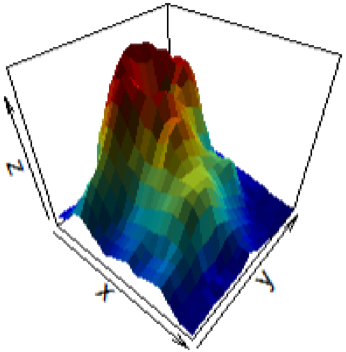
Example: 3D to 2D

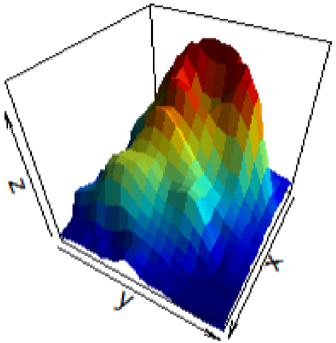
Volcano









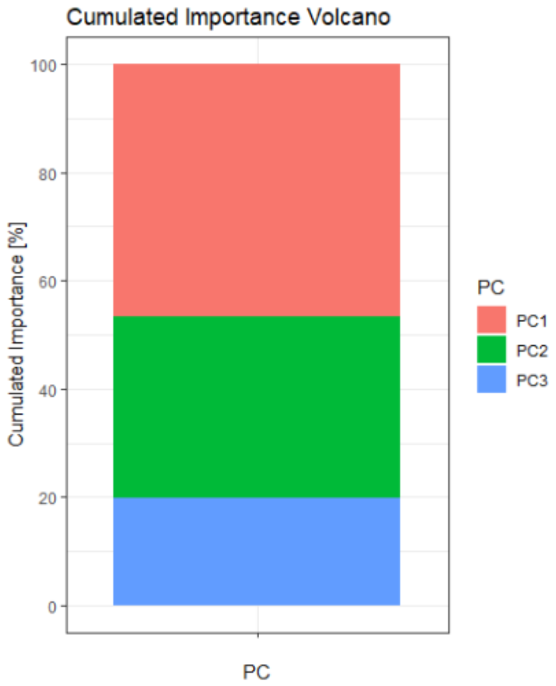


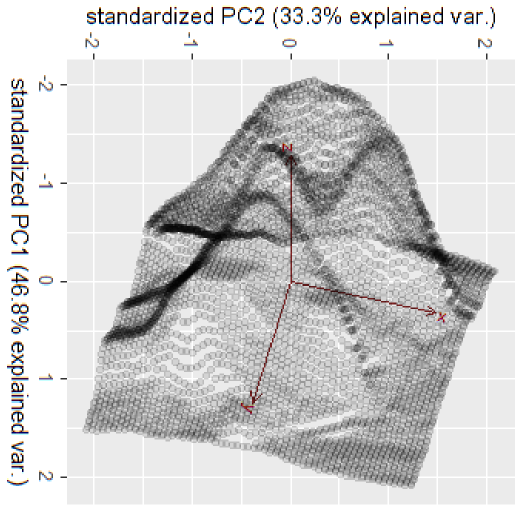
Principal Component Analysis

Example: 3D to 2D

Volcano

- PC1 and PC2 cumulative proportion of variance: 80%
- PC3 can be left out





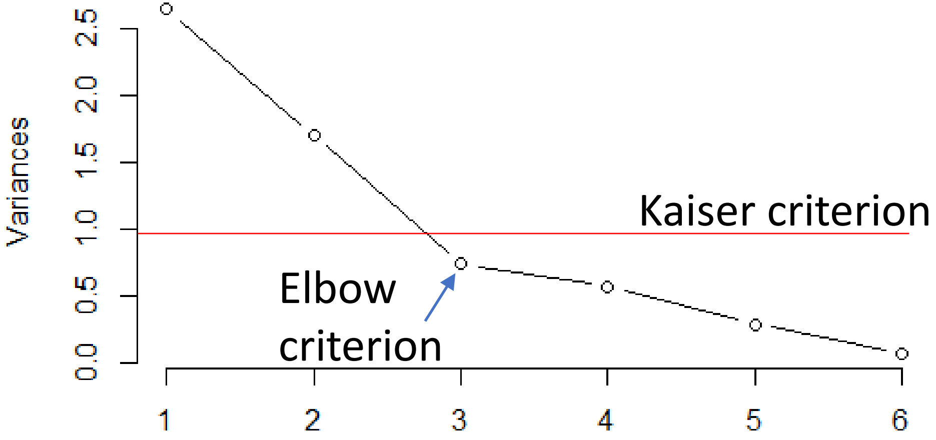
Principal Component Analysis

Number of Components

- Two or three for visualisation

- Scree plot
  - find „elbow“

- Kaiser criterion
  - Eigenvalues > 1



Principal Component Analysis

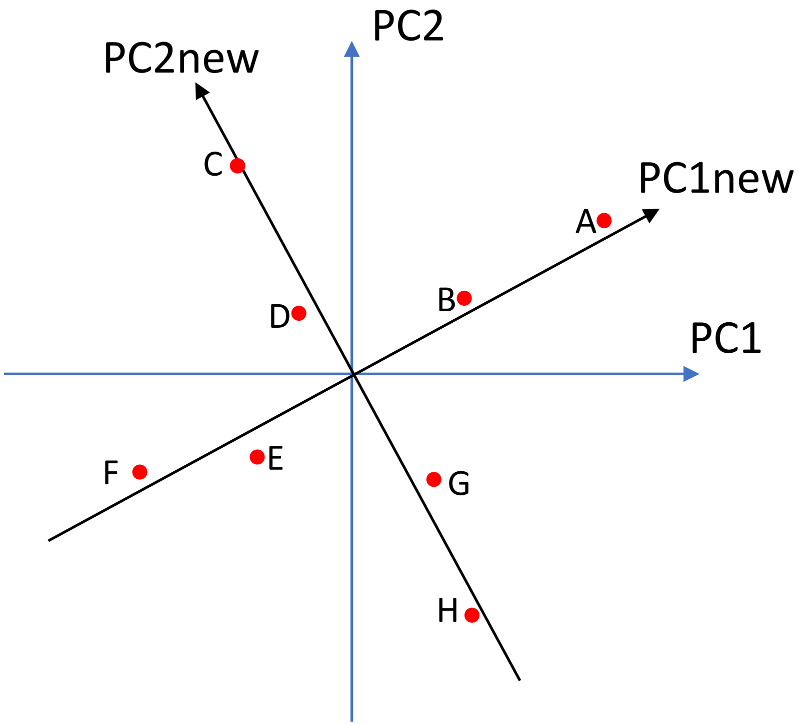
Rotation

Base point

- All observations load partially on PC1 and PC2
- Hard to interpret

After rotation

- Observations load on either PC1 or PC2
- Does not affect explained variance of components
- Most common rotation: varimax
- Helps to interpret the model



Principal Component Analysis

Advantages / Disadvantages

+

- Reduces higher dimensional data to a dimension that can be visualised
- Reduces noise
- Applicable for feature selection
- Well established

-

- Requires scaling of data!
- Relies on linear assumptions
- Very noisy variables can have too large impact