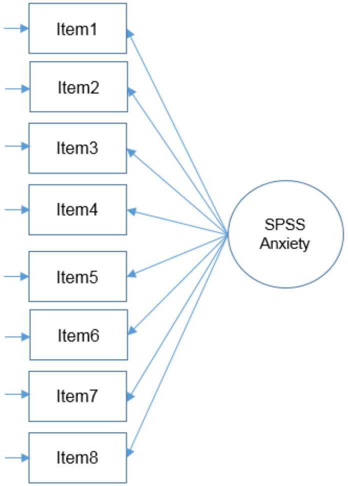
Factor analysis vs ICA

Irene Hofmann, Tore Gude, Johannes Djupesland, Bjørnar Ørjansen Kaarevik

Multivariate data analysis

20.11.2024

Factor analysis



- Used for **data reduction** for simplifying complex data sets
- Identifying **underlying relationships** between variables to create a **construct**

- **Basic assumption**: for a collection of observed variables there are a set of *underlying* variables called **factors** (smaller than the observed variables), that can explain the interrelationships among those variables.

- "Factor" refers to an underlying, unobserved variable or **latent construct** that represents a common source of **variation** among a set of observed or measured **variables**.

- **Types**:
  - **Exploratory Factor Analysis (EFA)**: Used when the researcher does not have a preconceived notion of how many factors are there or what they might be, exploring the data to find patterns.
  - **Confirmatory Factor Analysis (CFA)**: Used when the researcher has specific hypotheses about the number of factors and the loadings of observed variables based on theory or previous studies.

Factor analysis

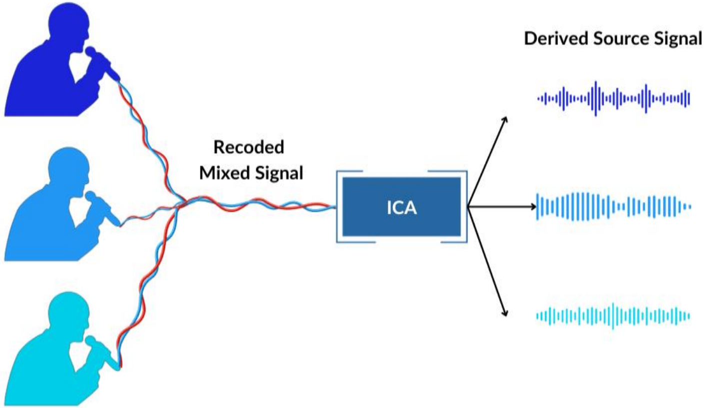
- **Key assumptions:**
  - linearity, absence of multicollinearity, inclusion of relevant variables, and a true correlation between variables and factors

- **Process**:
  - **Extraction of Factors**:
    - **Principal Component Analysis (PCA):** common variances takes up all of total variance
    - **Common Factor Analysis or Principal Axis Factoring (PAF):** total variance can be partitioned into common and unique variance
    - **Image Factoring**: uses ordinary least squares regression to predict factors
    - **Maximum Likelihood Method**: correlation matrix to derive factors
  - **Factor rotation**:
    - Interpret factor loadings
    - Different techniques depending if factors are correlated or not (Varimax or Promax) are applied to make the output more interpretable by simplifying the factor structure.
  - **Compute factor scores**
    - estimated scores of each observation for the factors and are used for further analysis

Examples: A Practical Introduction to Factor Analysis: Exploratory Factor Analysis, Confirmatory Factor Analysis (CFA) in R with lavaan

What is ICA?

- Independent Component Analysis
- Objective: Decompose a multivariate signal into independent components
- Key application areas:
  - Signal processing
  - Image analysis
  - EEG



Assumptions of ICA

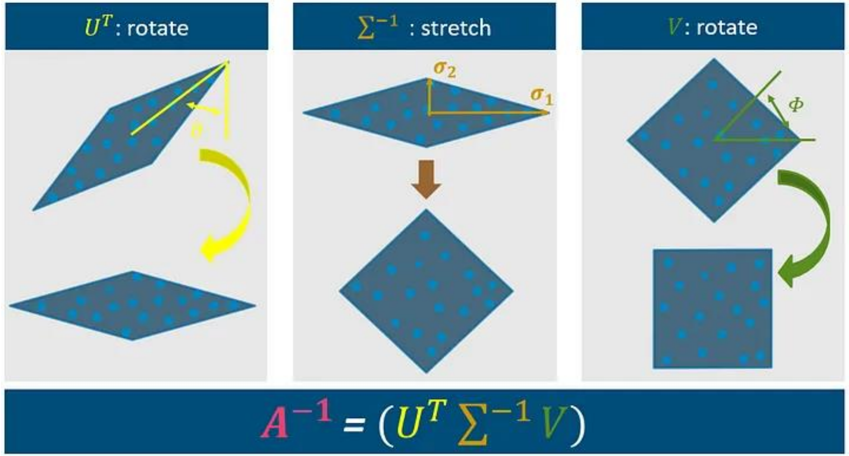
The source signals are independent of each other

Each source signal exhibits non-Gaussian distributions

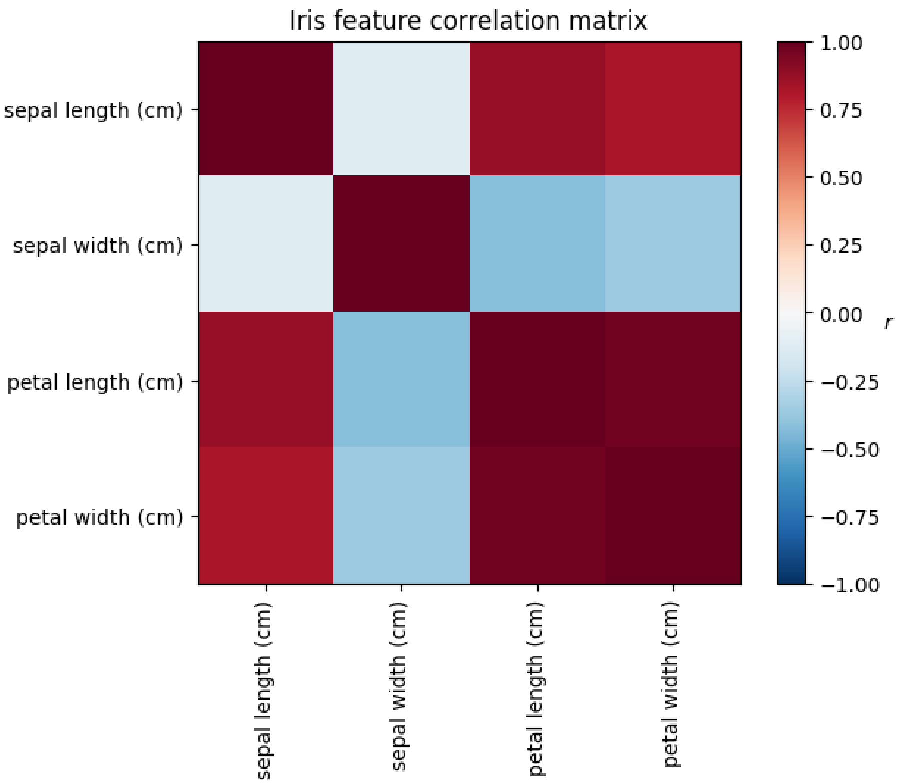
How does ICA work?

- Observed data X = AS
  - A: Mixing matrix
  - S: Independent source signals
- Goal: Estimate S and A using statistical independence

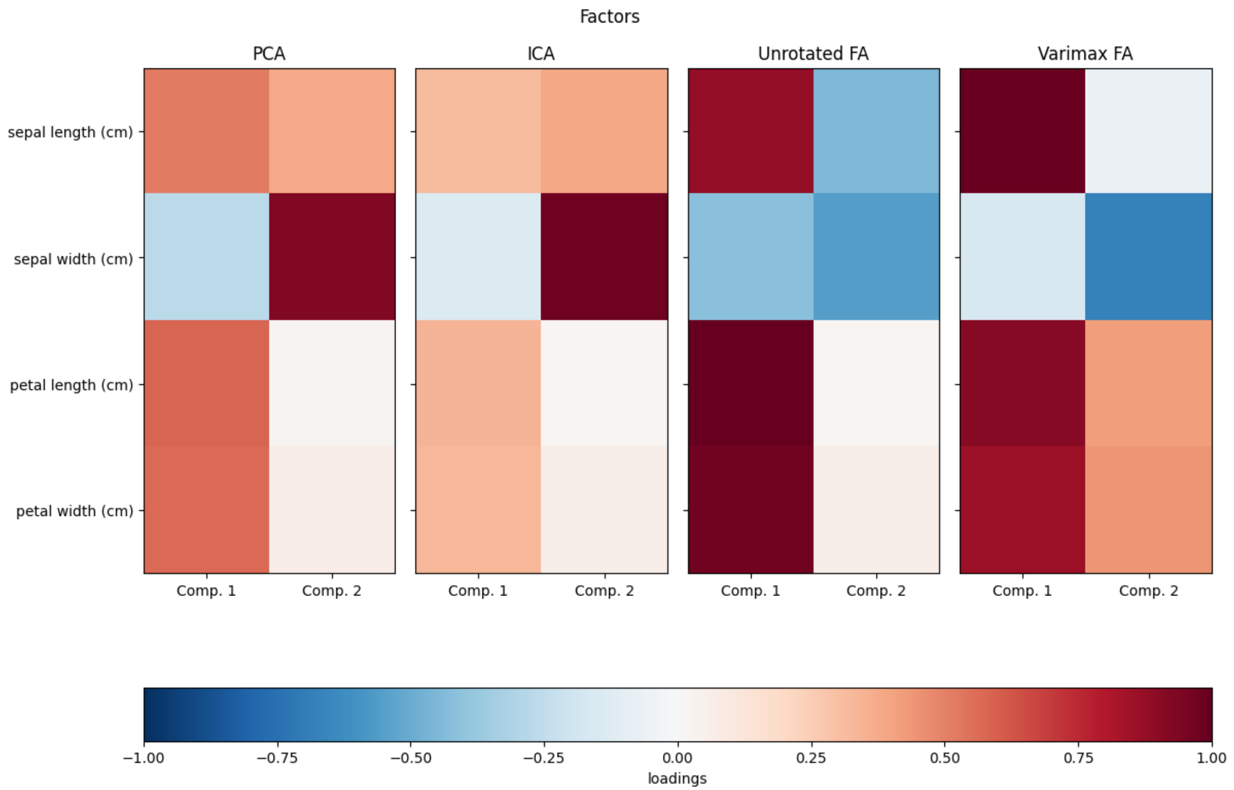
1. Find the angle with maximum variance to rotate
2. Find the scaling of the principal components
3. Independence assumptions for rotation



Code example – IRIS dataset

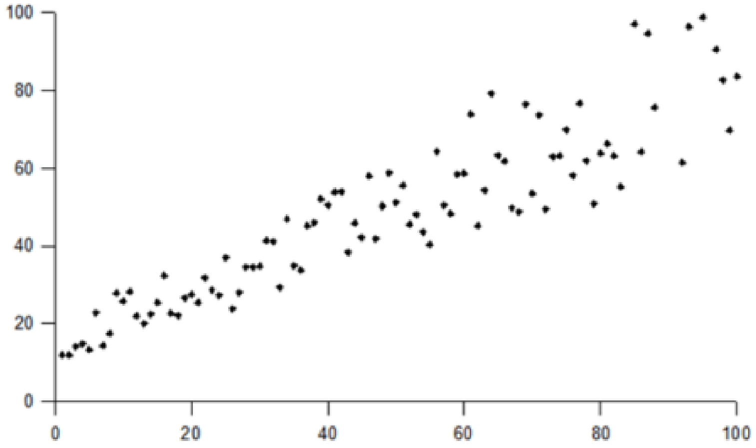


FA components are more purely indicating

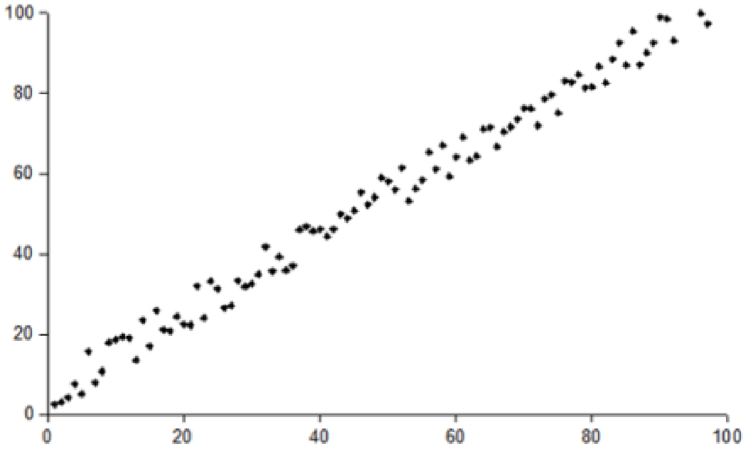


Code example – different noise processes

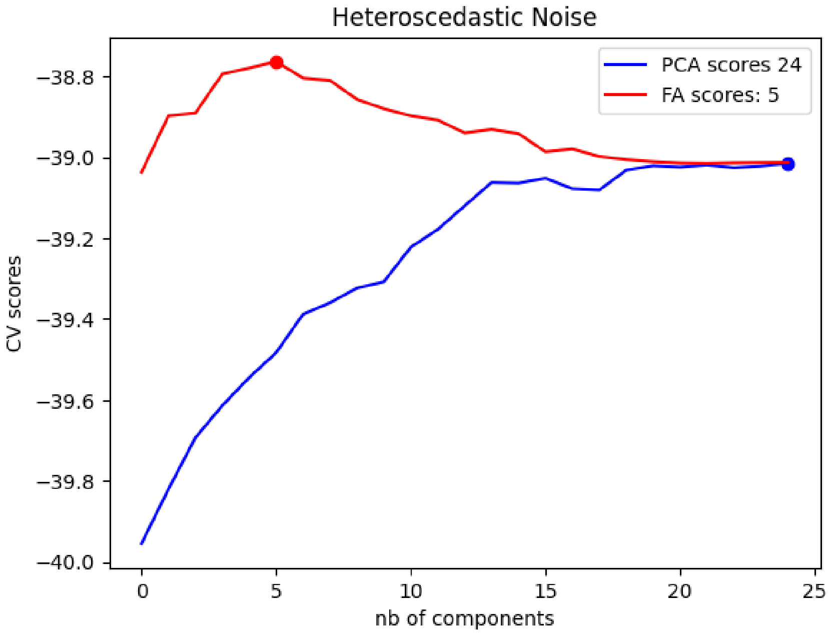
Heteroskedasticity

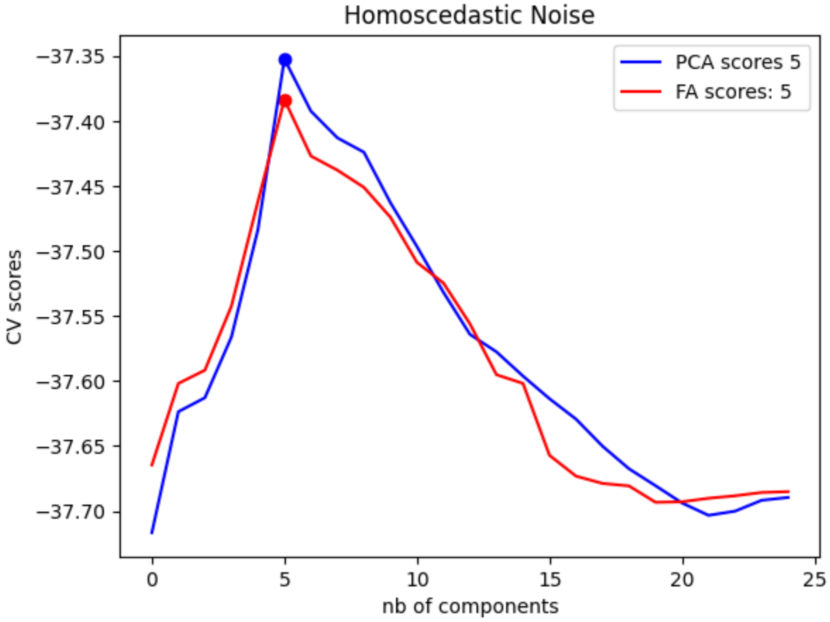


Homoskedasticity



Cross validation score computed by `cross_val_score` from `sklearn.model_selection`





Comparison

Factor Analysis

- **Objective:**
  - Identify latent factors and explain shared variance

- **Assumptions:**
  - Latent factors are uncorrolated; noise is Gaussian

ICA

- **Objective:**
  - Extract independent signals

- **Assumptions:**
  - Components are independent and non-Gaussian (more than one)

Comparison

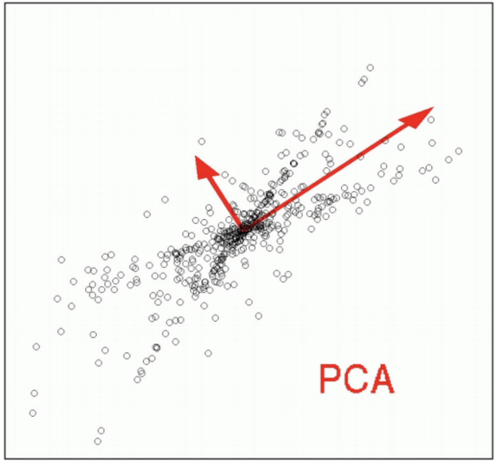
Factor analysis

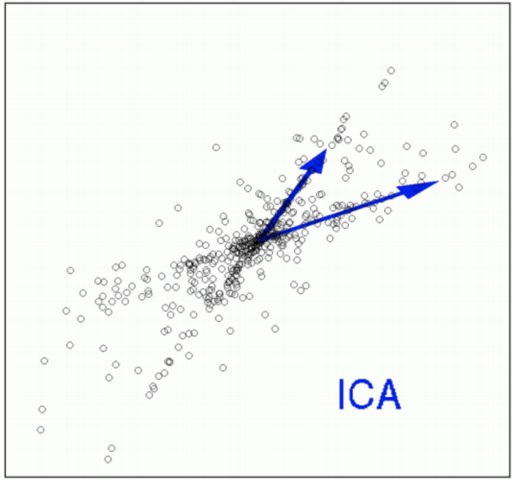
- **Noise Handling:**
  - Explicitly modeled
- **Independence Criterion:**
  - Based on correlation structure

ICA

- **Noise Handling:**
  - Not necessarily modeled
- **Independence Criterion:**
  - Based on statistical independence

Comparison - Example





Conclusion

- Most appropriate method depends on:
  - Objective
  - Structure of observed data
  - Structure of noise