Erasmus
School of
Economics

Generalising Invariant Coordinate Selection to a non-linear dimensionality reduction method

Master Thesis Defence for Econometrics and Management Science
Business Analytics & Quantitative Marketing
By Christopher Claassen

Contents

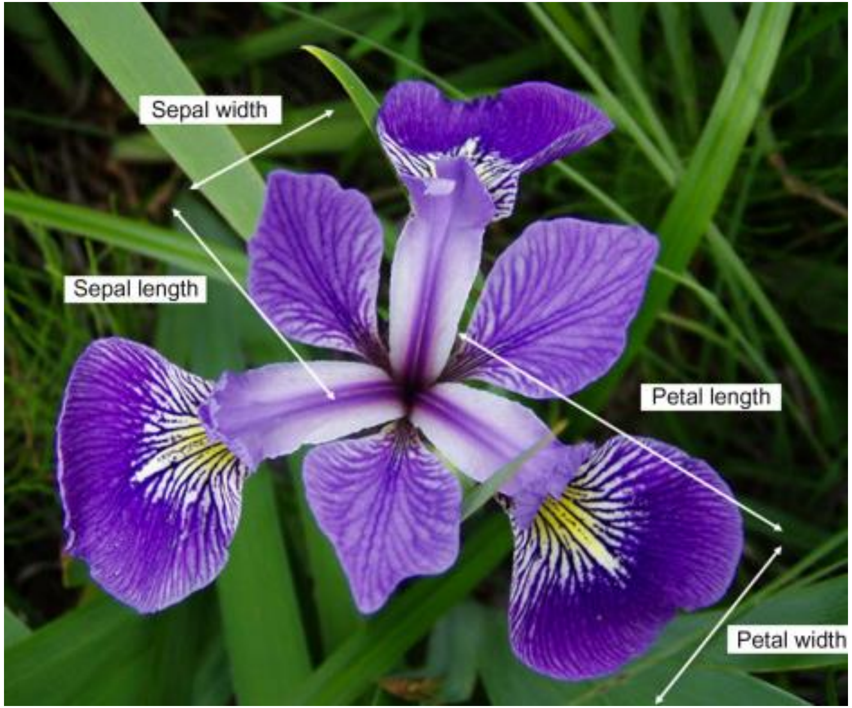
- Dimensions and orthogonality
- Dimensionality reduction methods
- Non-linear extensions via kernels
- Empirical applications of kernelised Invariant Coordinate Selection (ICS)
- Challenges and directions for future works

Setting the stage: what are dimensions?

- Dimensions are that what is required for describing everything residing in a space
  -> For example, the three physical dimensions (forwards, sideways and upwards) determine our position in this room

- Hyperspaces also exist and contain higher dimensional objects
  -> Finding a good representation in lower dimensions is challenging

From measurements to data matrices



Each type of measurement becomes a column, the number of columns gives the dimensionality

Sepal length	Sepal width	Petal length	Petal width
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.3
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5.0	3.4	1.5	0.2
4.4	2.9	1.4	0.2
4.9	3.1	1.5	0.1

Ideal dimensions versus reality

- Ideally, we would like for dimensions to be **orthogonal**
  -> This means we can move along one dimension without changing another

- In practice, measurements are often **correlated** and thus **not orthogonal**
  -> For the same flower species, observing a taller petal corresponds to a wider petal

- Nevertheless, we can often reorient the data such that it becomes orthogonal
  -> This turns out to be very useful for dimensionality reduction

Dimensionality reduction

- Dimensionality reduction entails using less dimensions than observed

- Possible advantages:
  -> Quicker computations
  -> Can reduce noise in the data

- Possible disadvantages:
  -> Requires additional computations
  -> Loss of relevant information

- Often done as a preprocessing step for subsequent statistical analysis

Dimensionality reduction

- Dimensionality reduction entails using less dimensions than observed

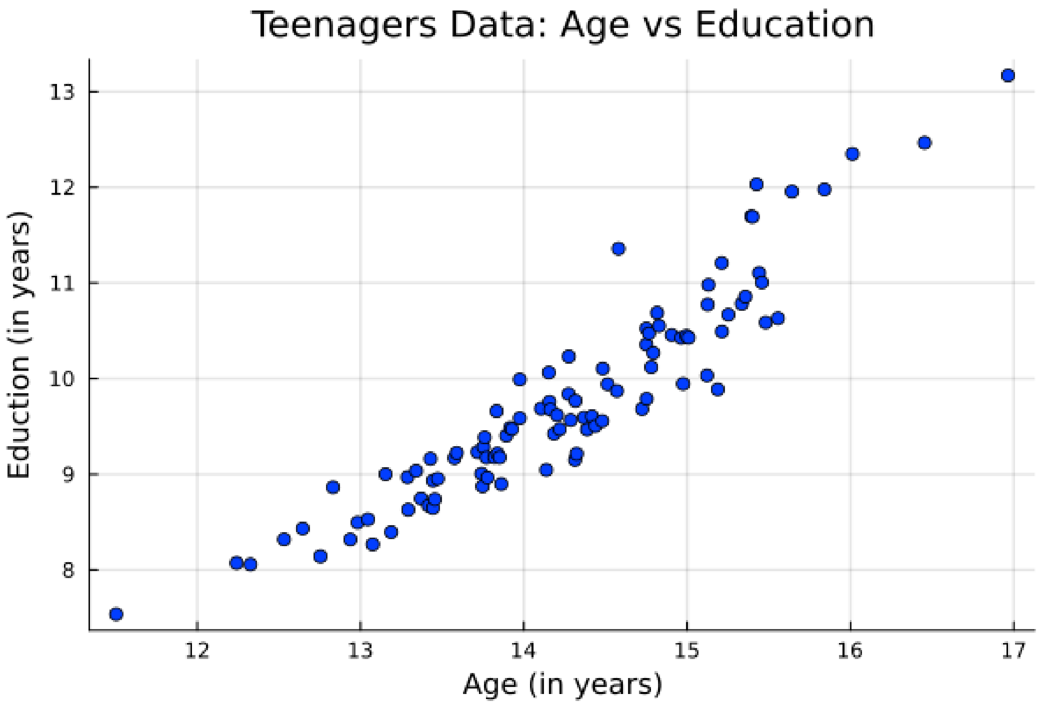
- Possible advantages:
  -> Quicker computations
  -> Can reduce noise in the data

Trade-off

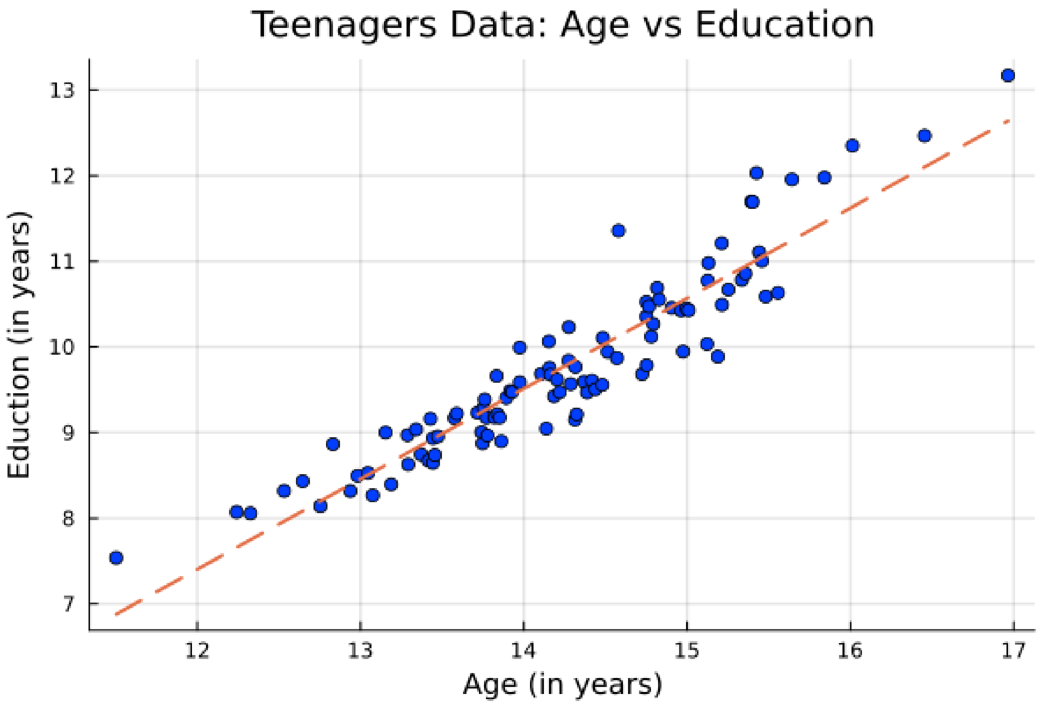
- Possible disadvantages:
  -> Requires additional computations
  -> Loss of relevant information

- Often done as a preprocessing step for a subsequent statistical method

Dimensionality reduction



Dimensionality reduction



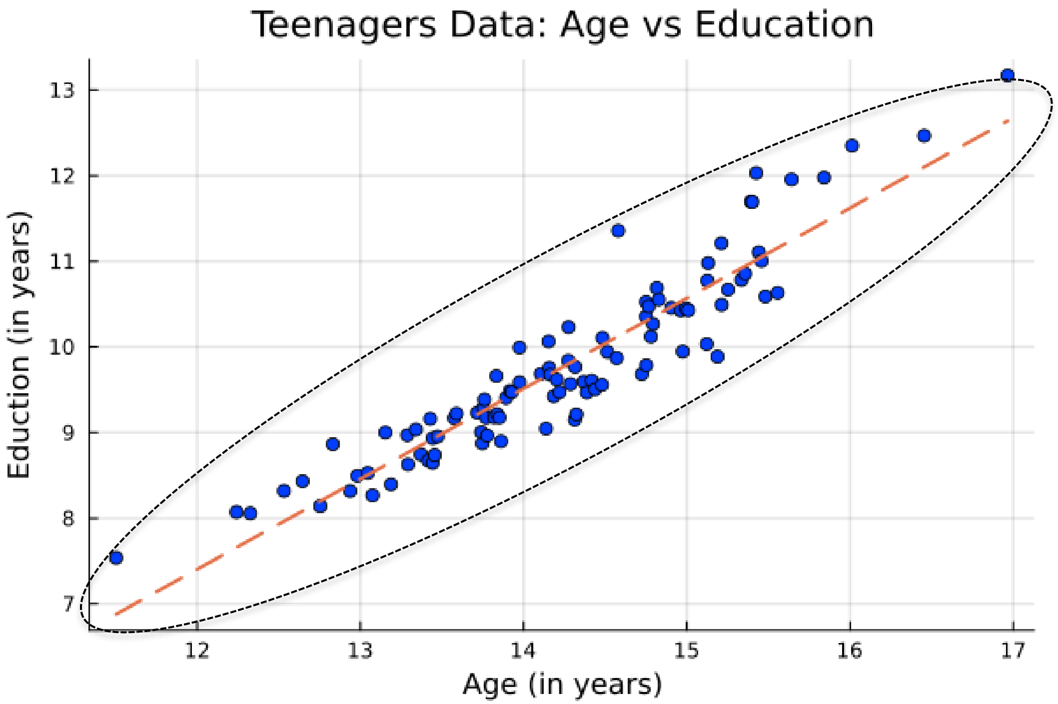
Principal Component Analysis (PCA)

- PCA is a method that reorients the data such that it becomes orthogonal
  -> In practice, this means rotating and stretching the data

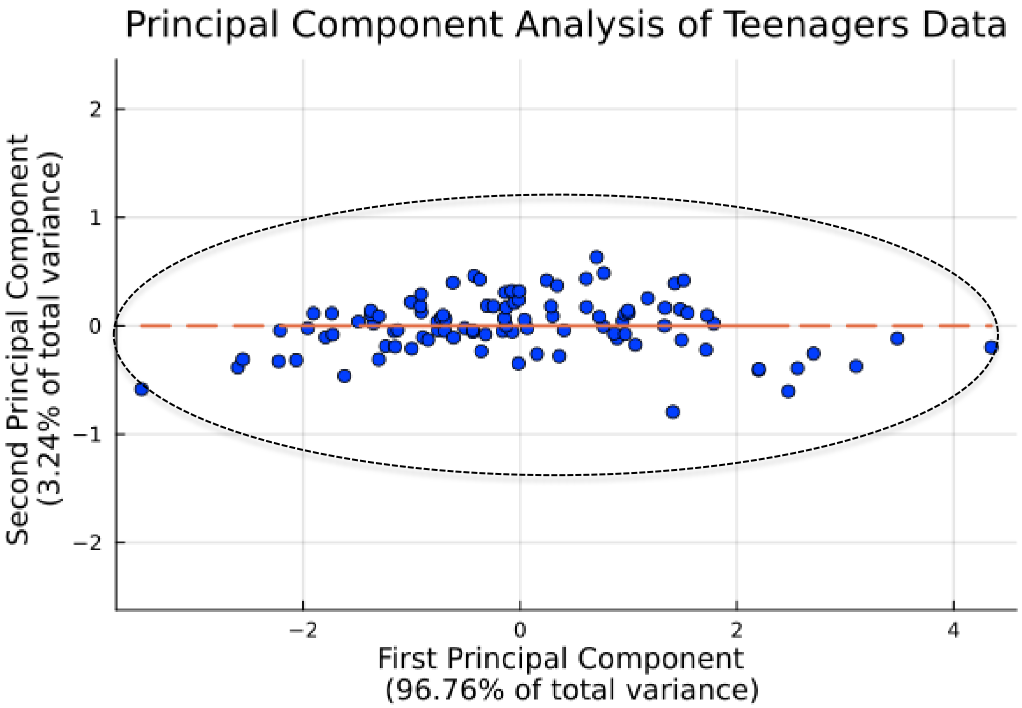
- PCA achieves this by exploiting the variance-covariance structure of the data
  -> Variance: measures the spread of a dimension
  -> Covariance: measures the joint variation of two dimensions

- Dimensionality reduction is achieved by discarding dimensions with little variance
  -> A dimension that varies little is often uninteresting in practice

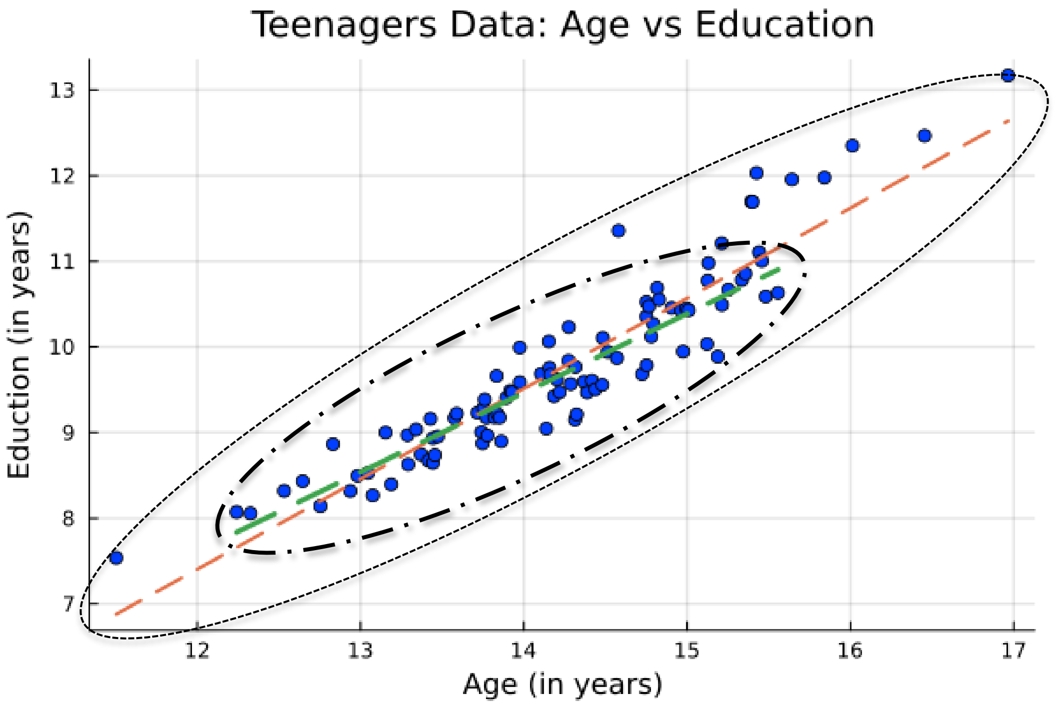
Dimensionality reduction via PCA



Dimensionality reduction via PCA



From covariances to scatters



Invariant Coordinate Selection (ICS)

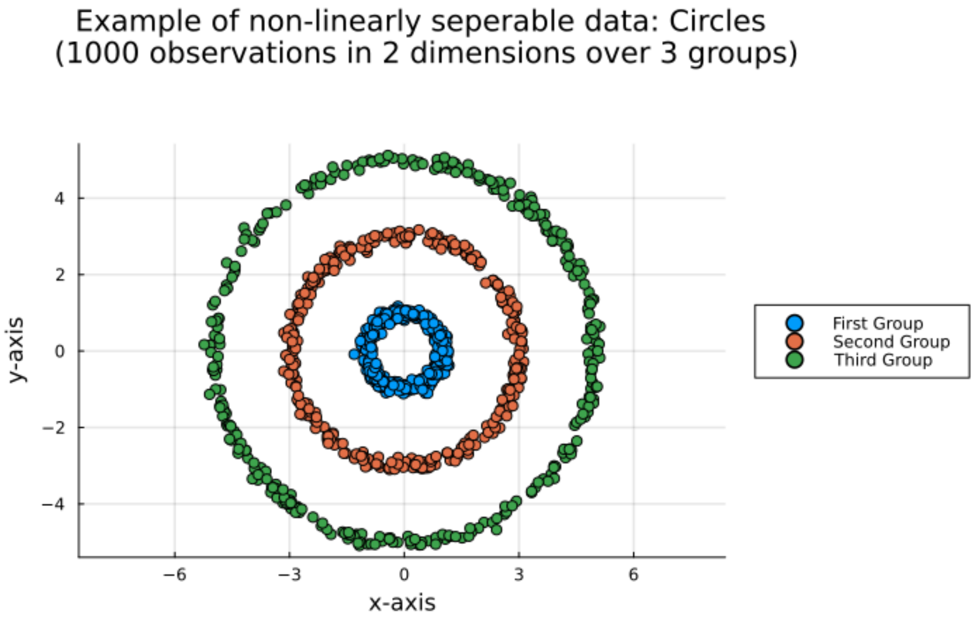
- ICS generalises PCA, as it uses 2 different scatters rather than 1
  -> The two scatters cannot be the same, there must be some form of contrast

- The two scatters jointly determine the type of structure uncovered
  -> For example, using covariance versus robust covariance can reveal outliers

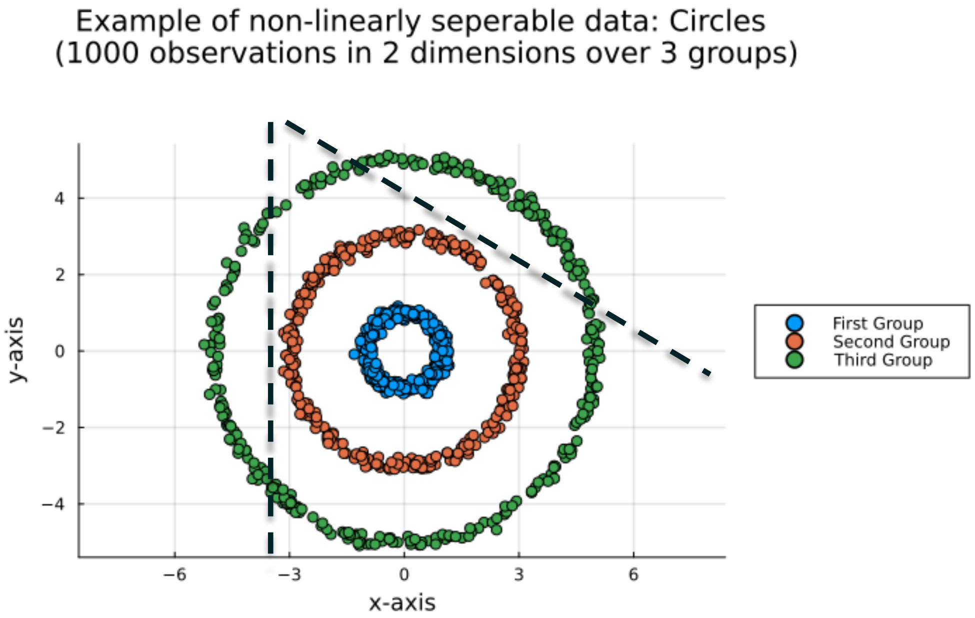
- The structure ICS finds is often on a low-dimensional subspace

- The challenge for using ICS is finding the appropriate scatters

A quick detour: non-linearity



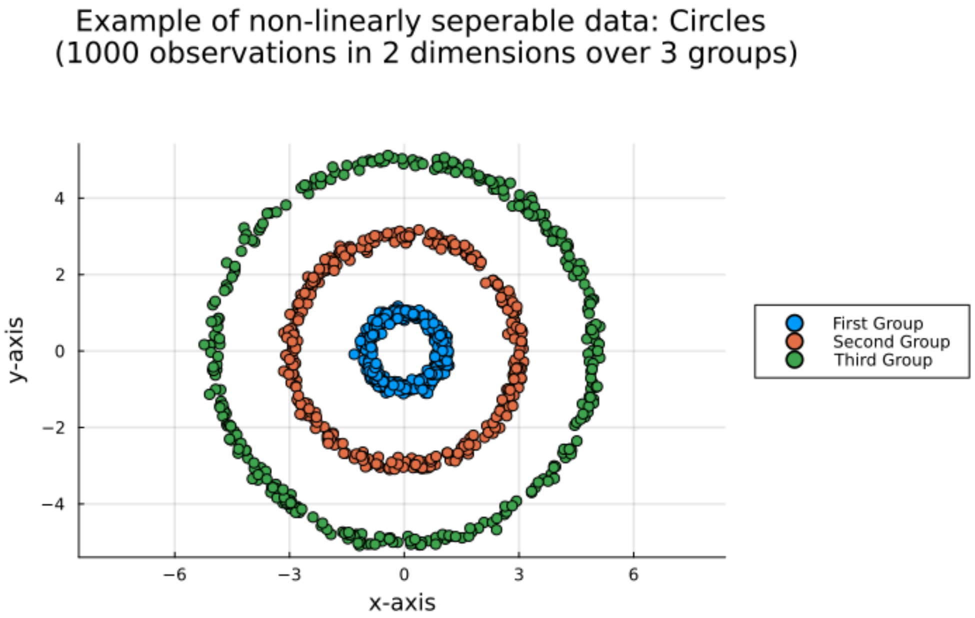
Quick detour: non-linearity



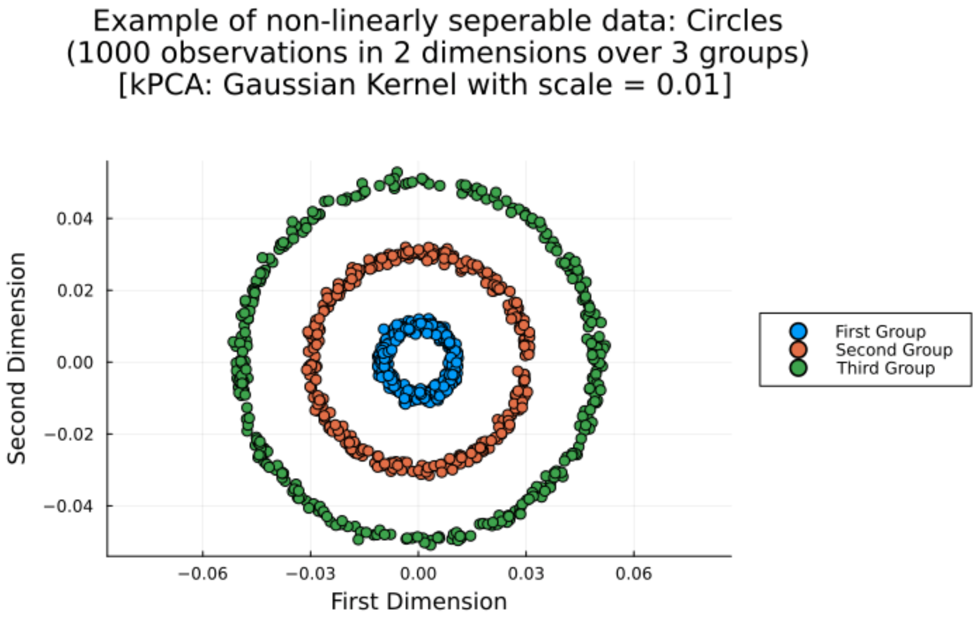
Addressing non-linearity via kernels

- Non-linearity cannot be tackled via rotating and stretching the data
- Kernels capture a certain ‘core’ of non-linearity of the data
  -> Kernels can replace the role of scatters in ICS
- We focus on two types of kernels: reproducing kernels and smoothing kernels
- Reproducing kernels capture non-linear functions of the data
- Smoothing kernels capture the interaction between a point and its close neighbours

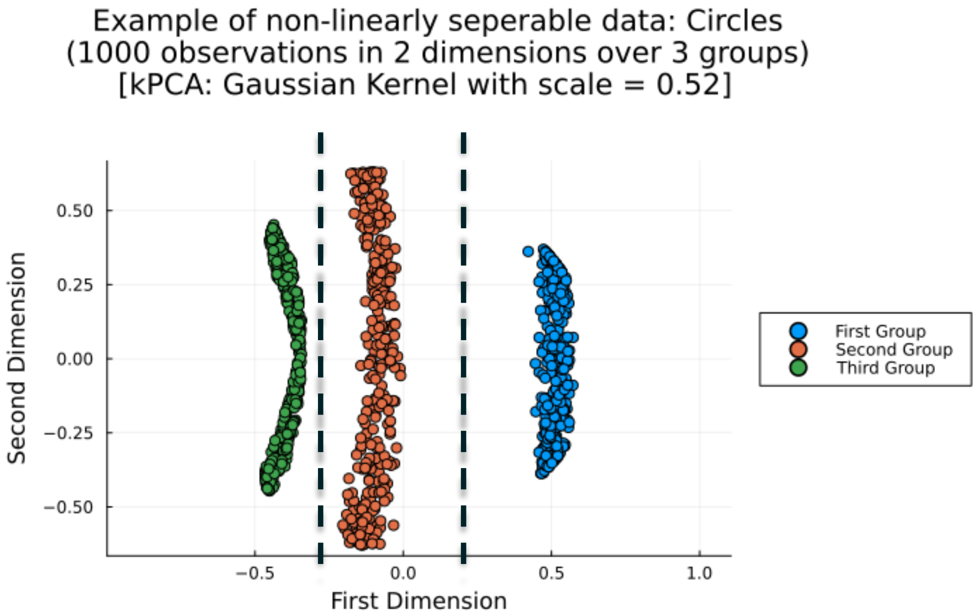
Non-linearity via reproducing kernels: a demonstration



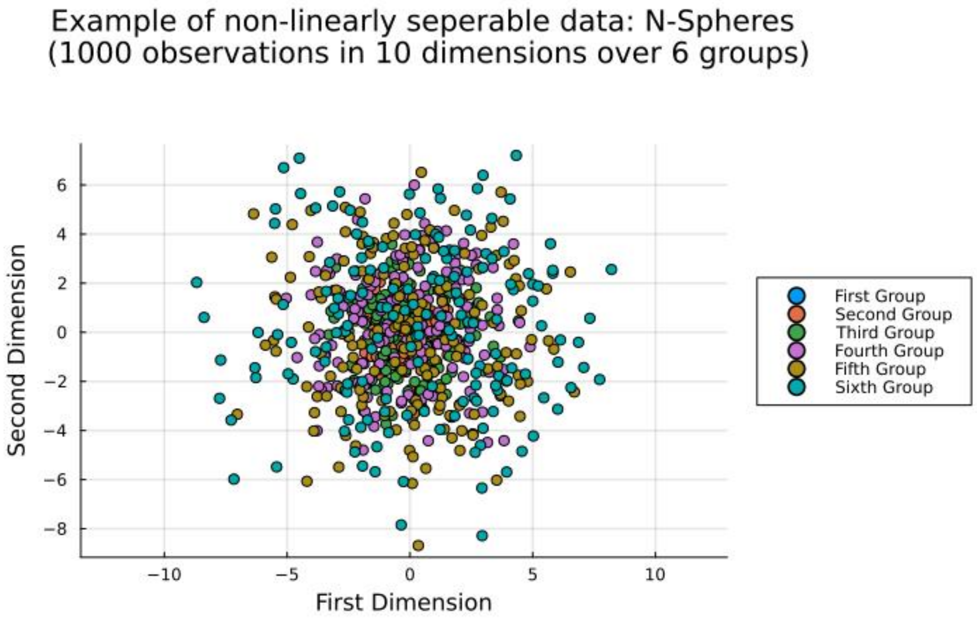
Non-linearity via reproducing kernels: a demonstration



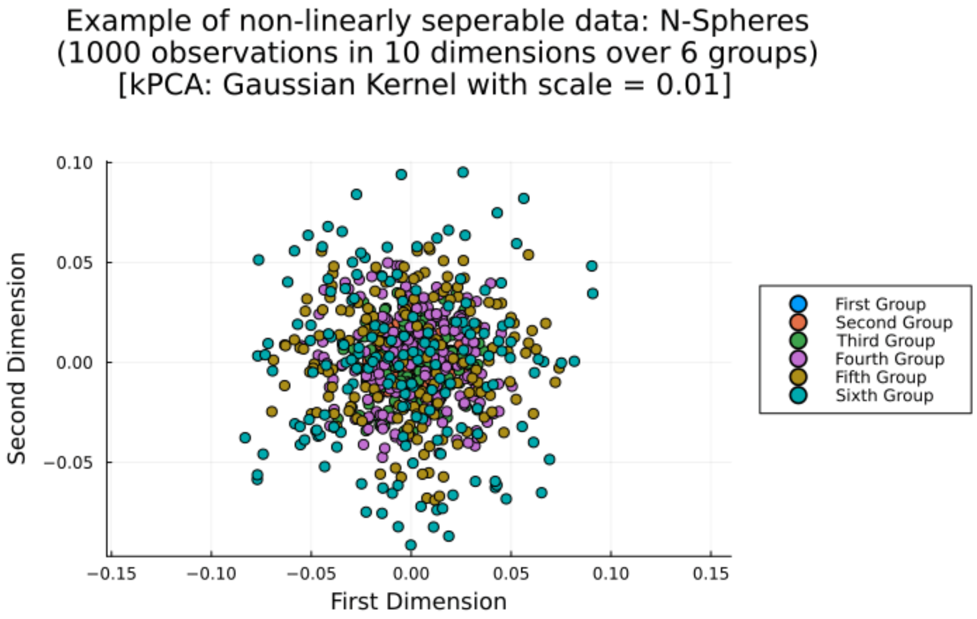
Non-linearity via reproducing kernels: a demonstration



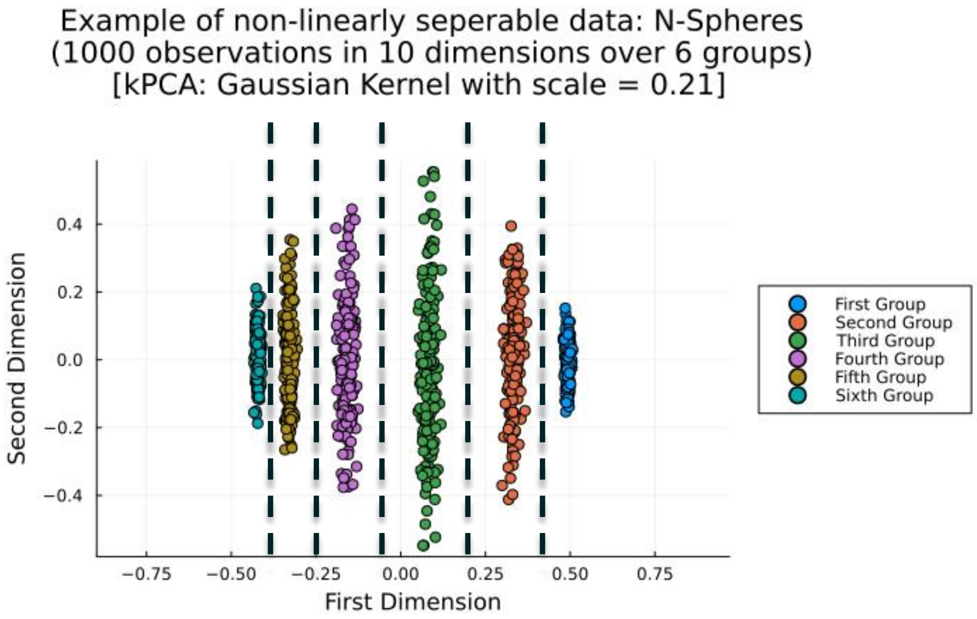
Non-linearity via reproducing kernels: a demonstration



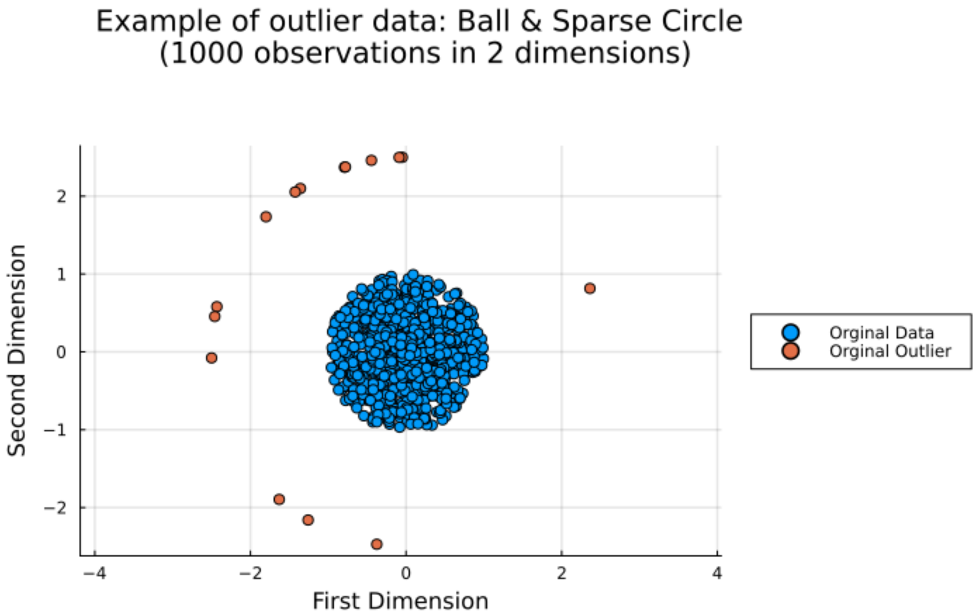
Non-linearity via reproducing kernels: a demonstration



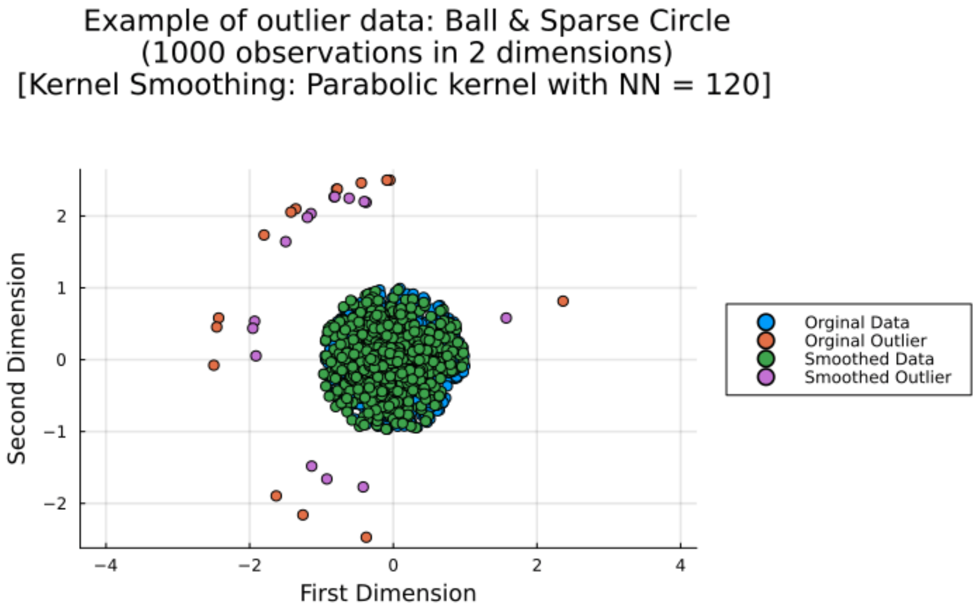
Non-linearity via reproducing kernels: a demonstration



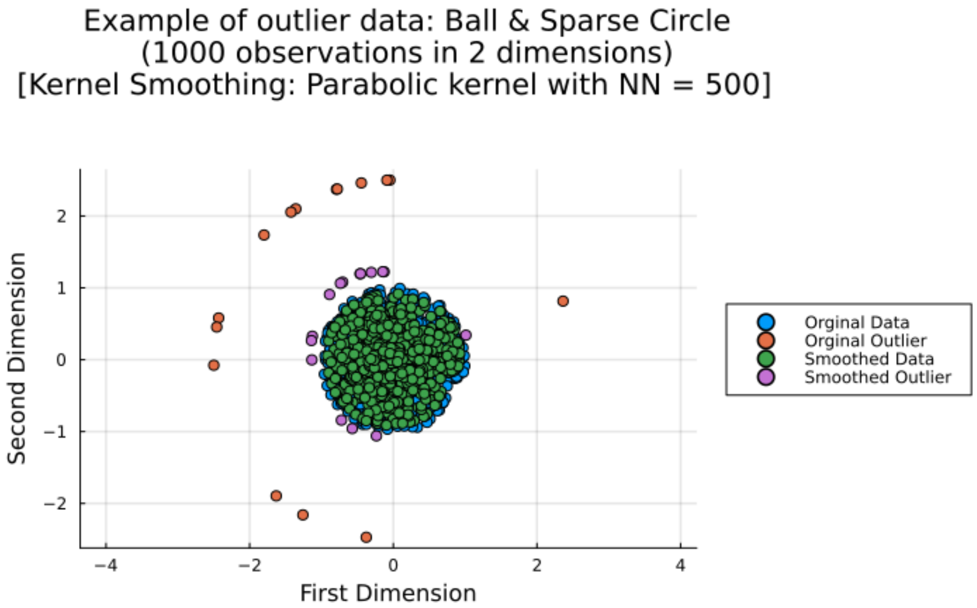
Non-linearity via smoothing kernels: a demonstration



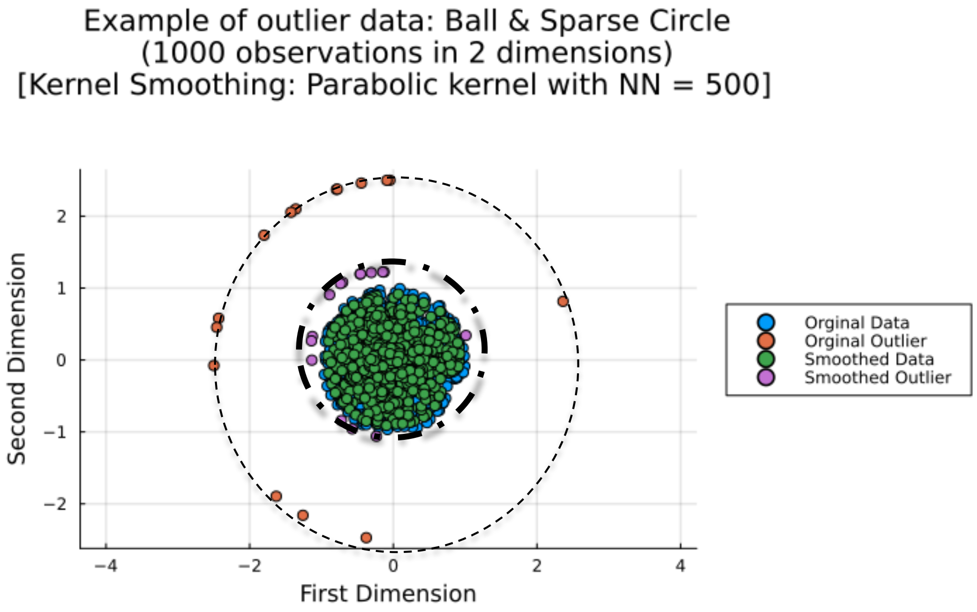
Non-linearity via smoothing kernels: a demonstration



Non-linearity via smoothing kernels: a demonstration



Non-linearity via smoothing kernels: a demonstration



Empirical applications of kernelised ICS

- Dimension reduction via kernelised ICS is able to uncover specific types of structure
  -> In particular, specific types of non-linear structure

- There are many possible subsequent statistical applications, such as:
  -> Data visualisation
  -> Detecting data anomalies
  -> Finding clusters of similar observations in the data

- Using non-linear methods directly for these applications is costlier

Empirical application of ICS: Wine data

- The dataset consists of wine samples originating from different cultivars

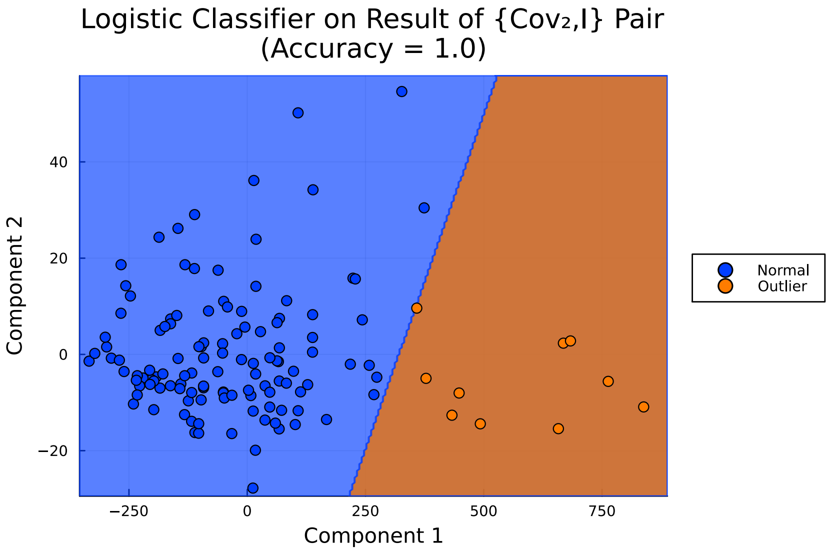
- There are 129 samples with 13 chemical measurements each
  -> Examples of measurements are alcohol contents, sweetness and hue

- The goal is to distinguish between two different cultivars using only few dimensions
  -> For the dimensionality reduction step, we do not use the labels

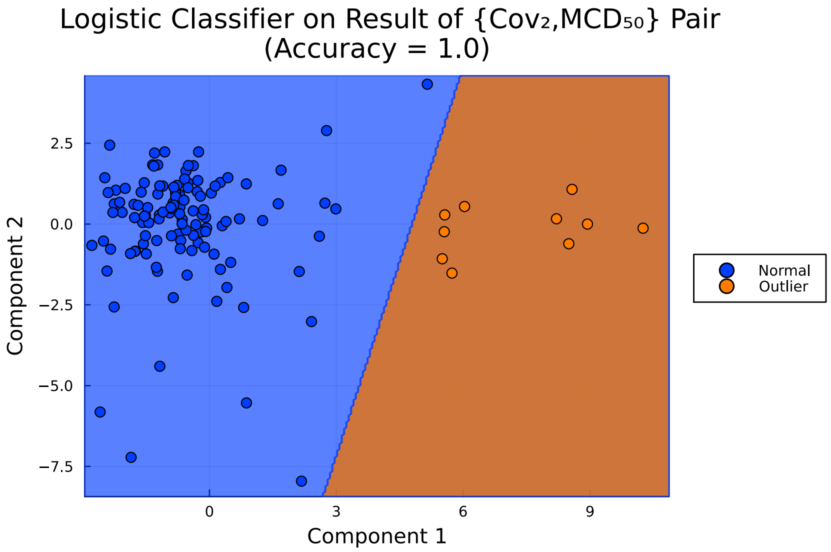
- Other applications that can use this methodology include:
  -> Detecting counterfeit objects
  -> Finding defective machine parts

Empirical application of ICS : Wine data

Example of PCA

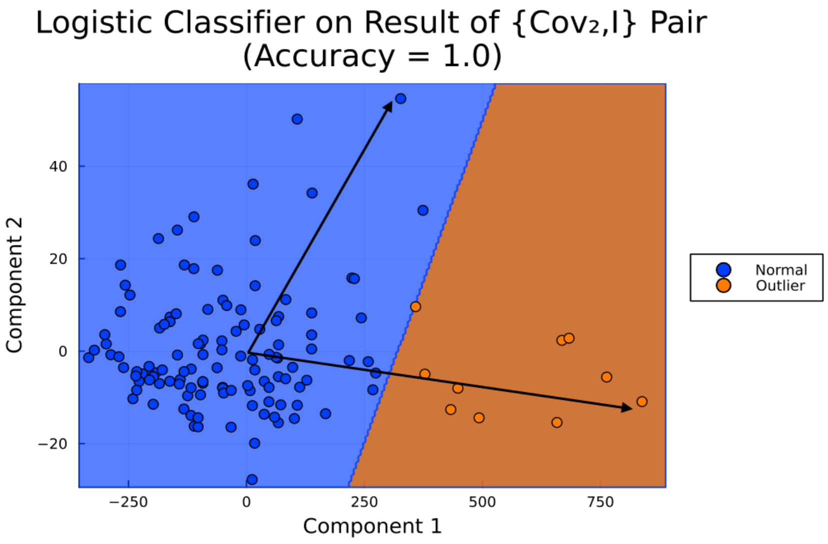


Example of ICS

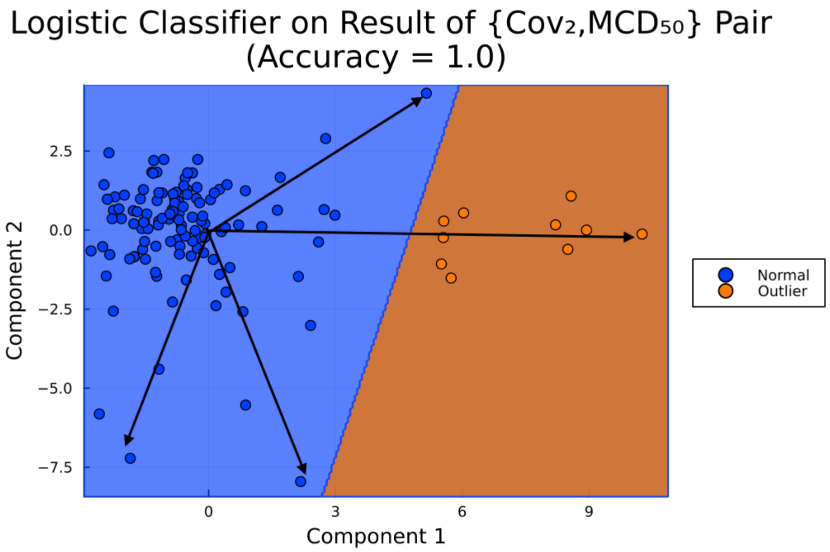


Empirical application of ICS : Wine data

Example of PCA

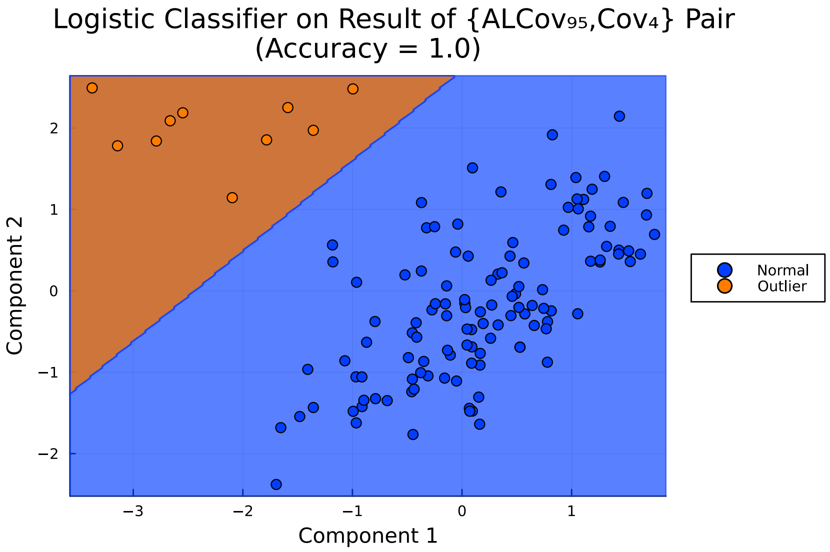


Example of ICS

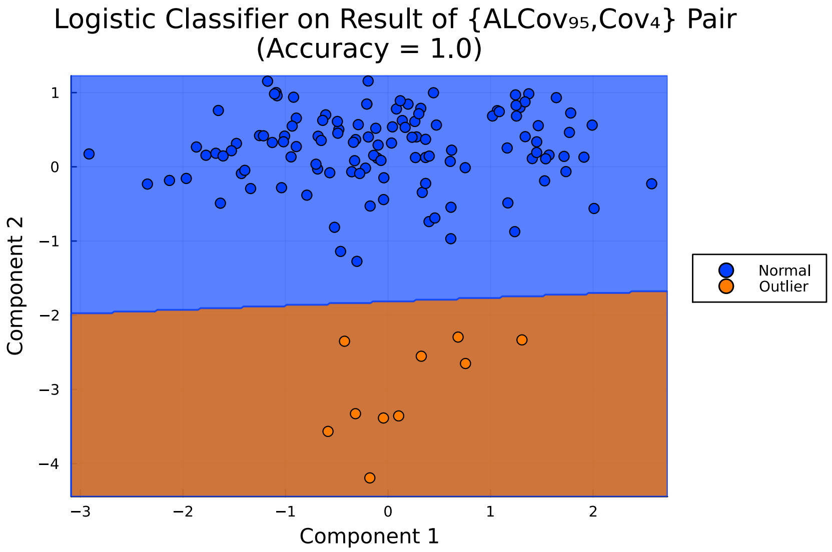


Empirical application of ICS : Wine data

Example of kernel ICS

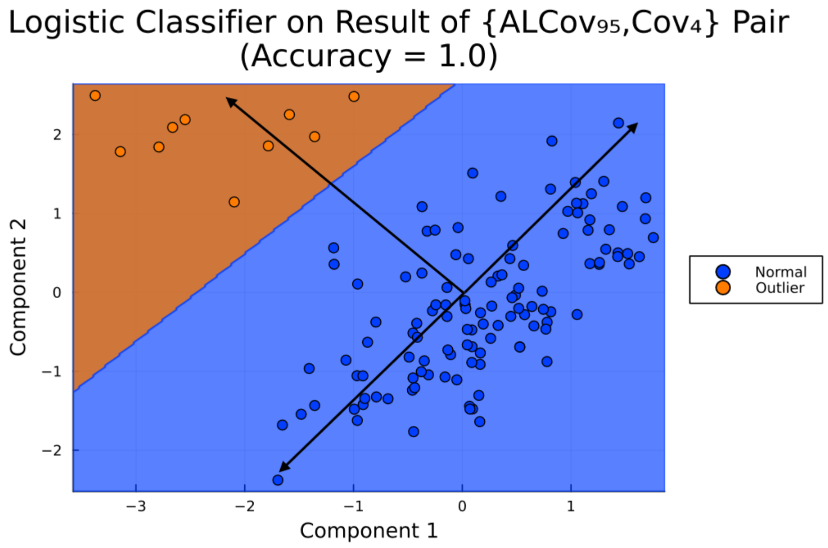


Example of rotated kernel ICS

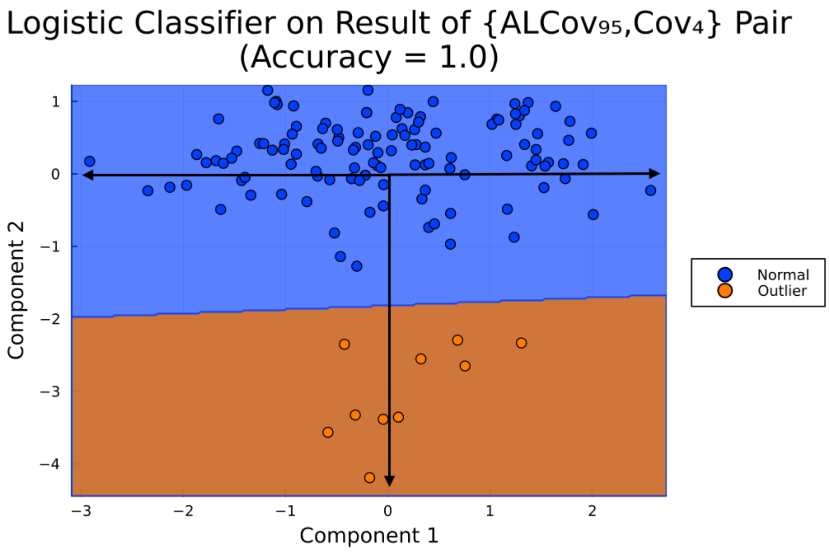


Empirical application of ICS : Wine data

Example of kernel ICS



Example of rotated kernel ICS



Empirical application of ICS : Word embedding data

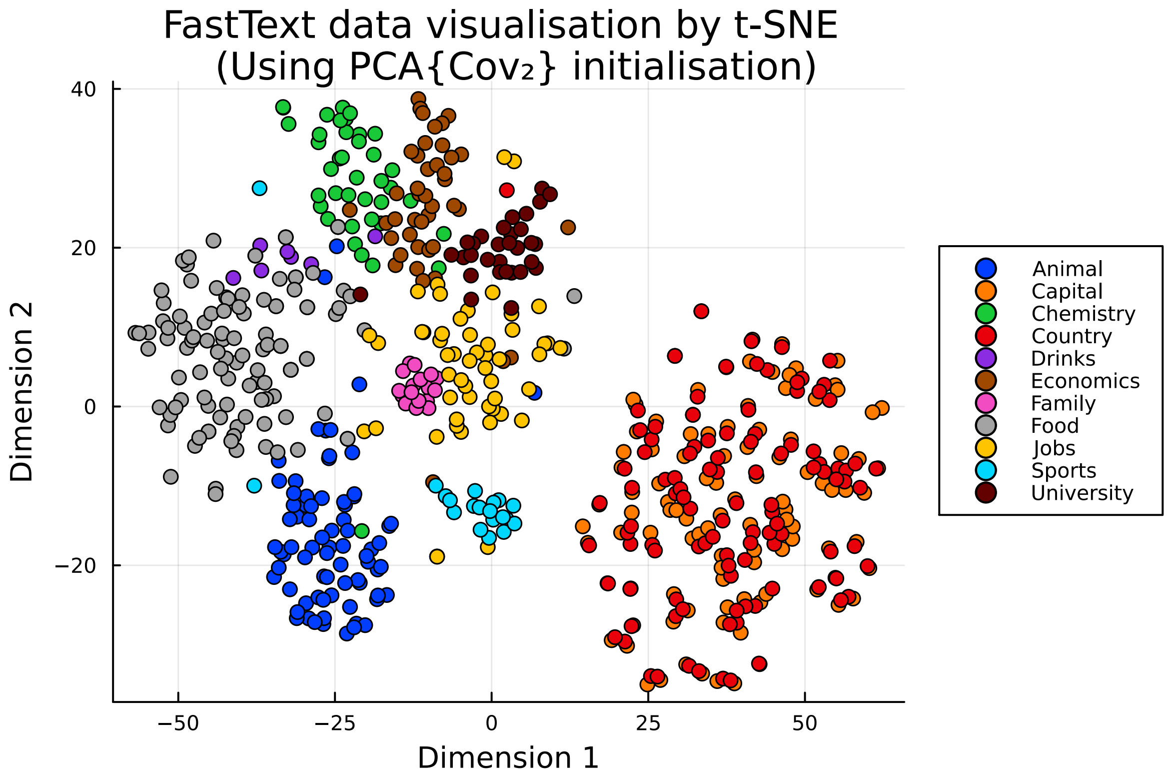
- The dataset consists of word embedding created by machine learning models

- There are 11 classes spread over 601 words with 300 dimensions each
  -> The dimensions do not have a direct interpretation

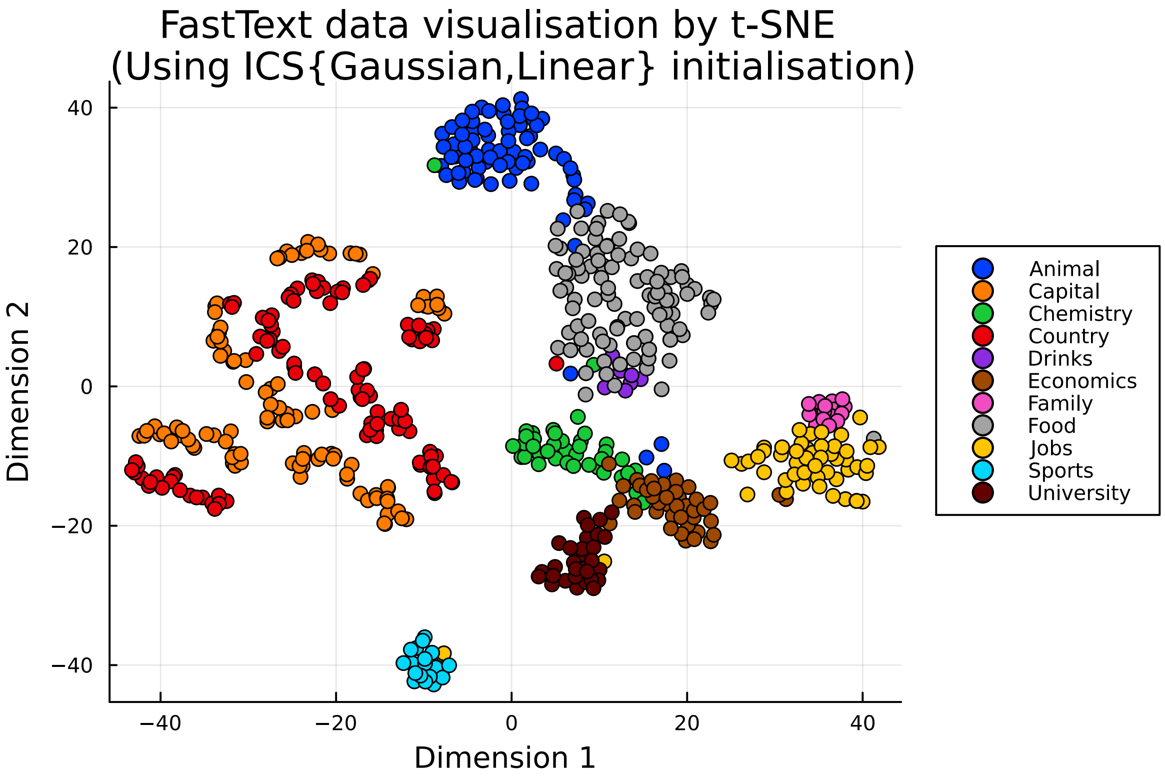
- The goal is to visualise the different word clusters by transfering the information of only a few dimensions to a different method called t-SNE
  -> For the dimensionality reduction step, we do not use the class labels

- Other applications of this type of data include:
  -> Processing written costumer reviews
  -> Creating AI chatbots like ChatGPT

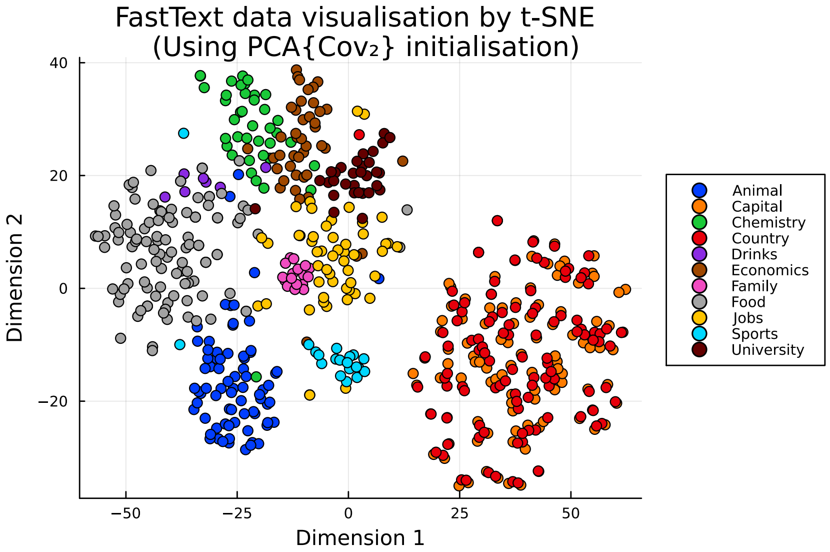
Empirical application of ICS : Word embedding data

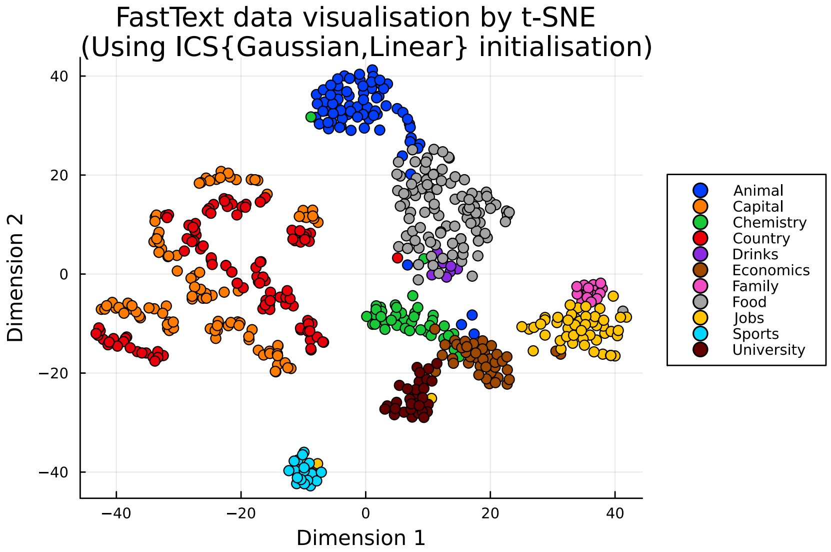


Empirical application of ICS : Word embedding data



Empirical application of ICS : Word embedding data





Challenges and directions for future works

- We have seen that ICS with kernels can be succesfully applied as a dimensionality reduction method for a variety of applications, but this was not without challenges:

  -> Creating a computer package that effectively performs kernelised ICS

  -> Determining which kernel pairs to use

  -> Accounting for numerical instability caused by using kernels

- There is also still potential for improvement of ICS with kernels:

  -> Studying theoretical aspects

  -> Component selection

  -> Kernel hyperparameter optimisation

Thank you for your attention!