Dimensionality reduction

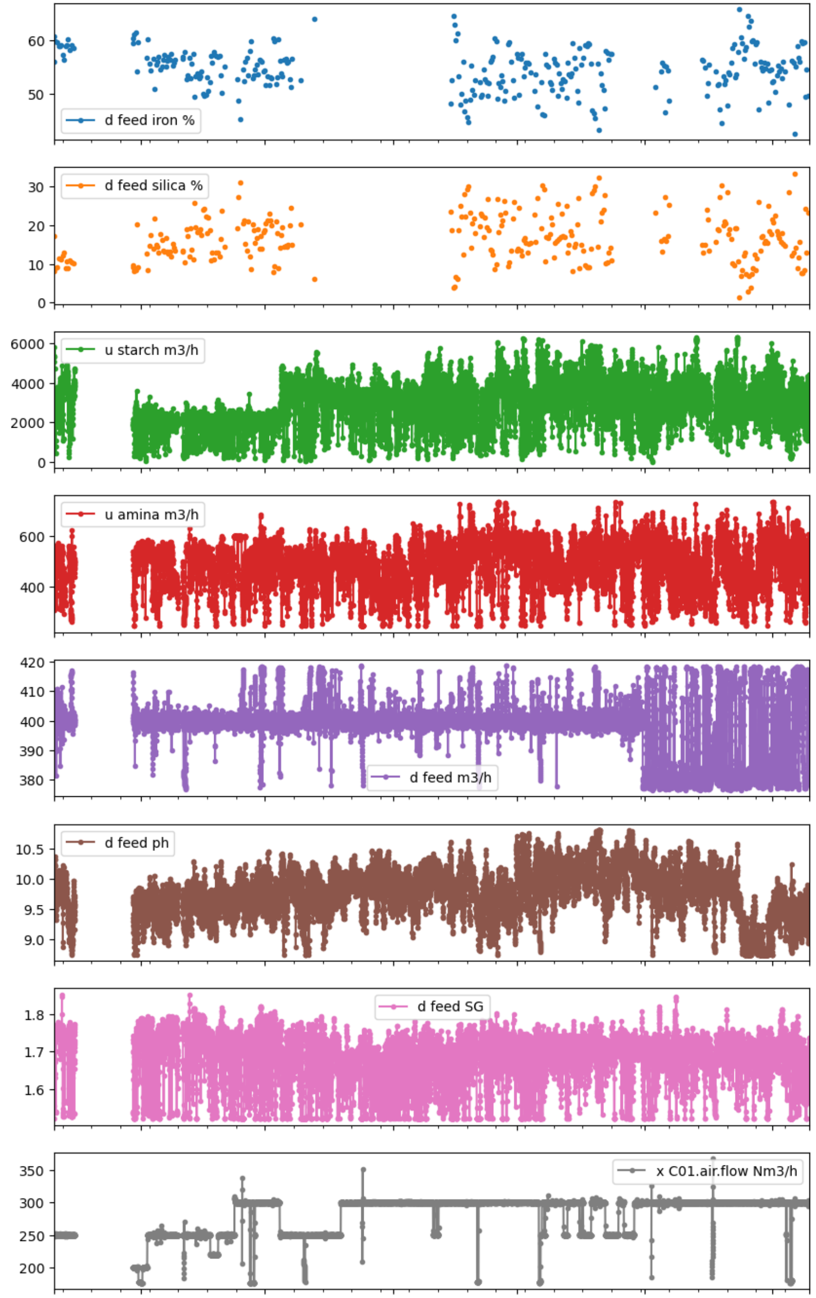
South-African Council for Automation and Control

Exploratory Data Analysis workshop

March 2024

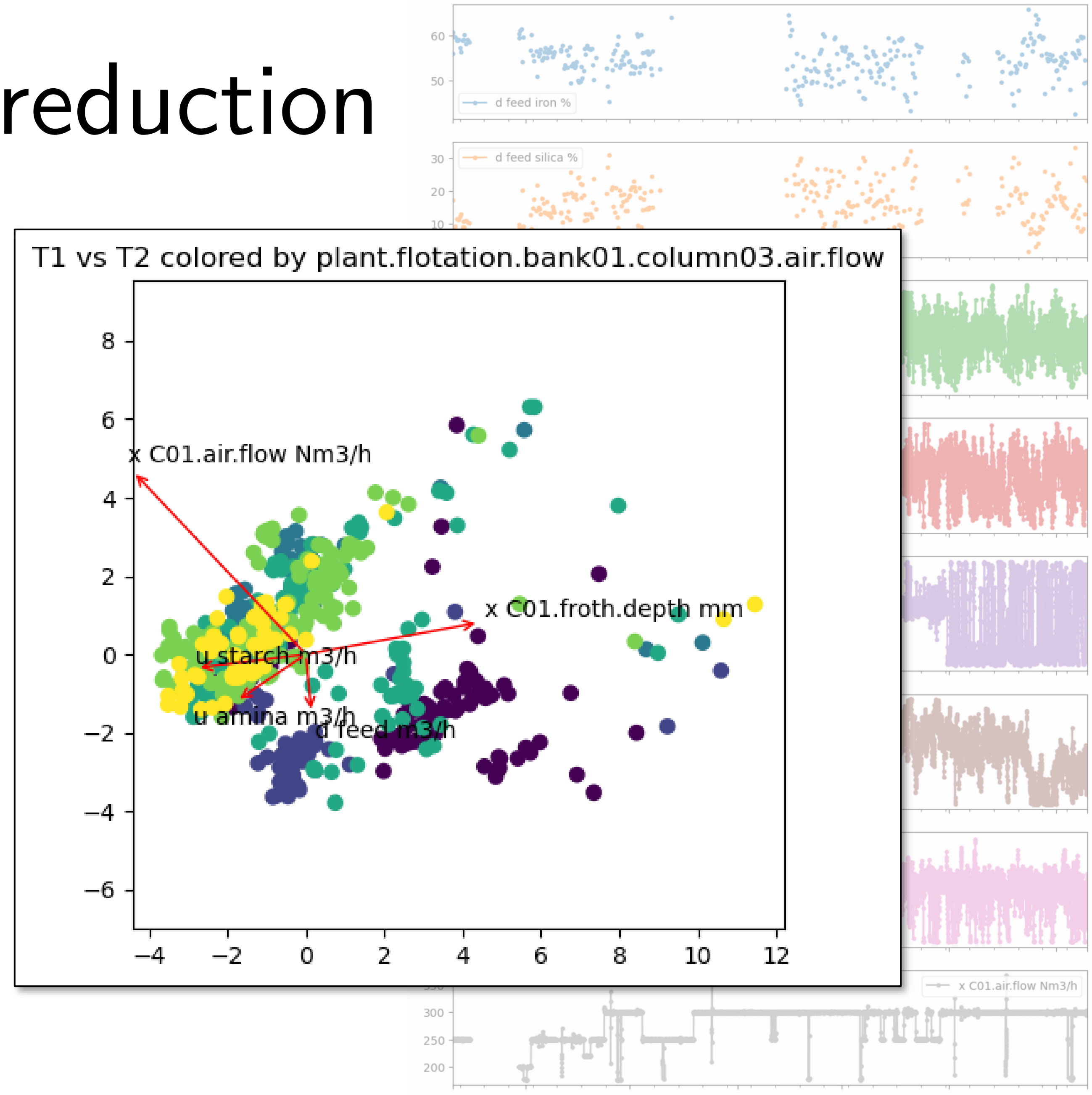
Why dimensionality reduction

- Visualisation
  - Process data consists of many measurements
  - Viewing rows of time series or single feature statistics is not always the most effective method to identify *structure* in the data



Why dimensionality reduction

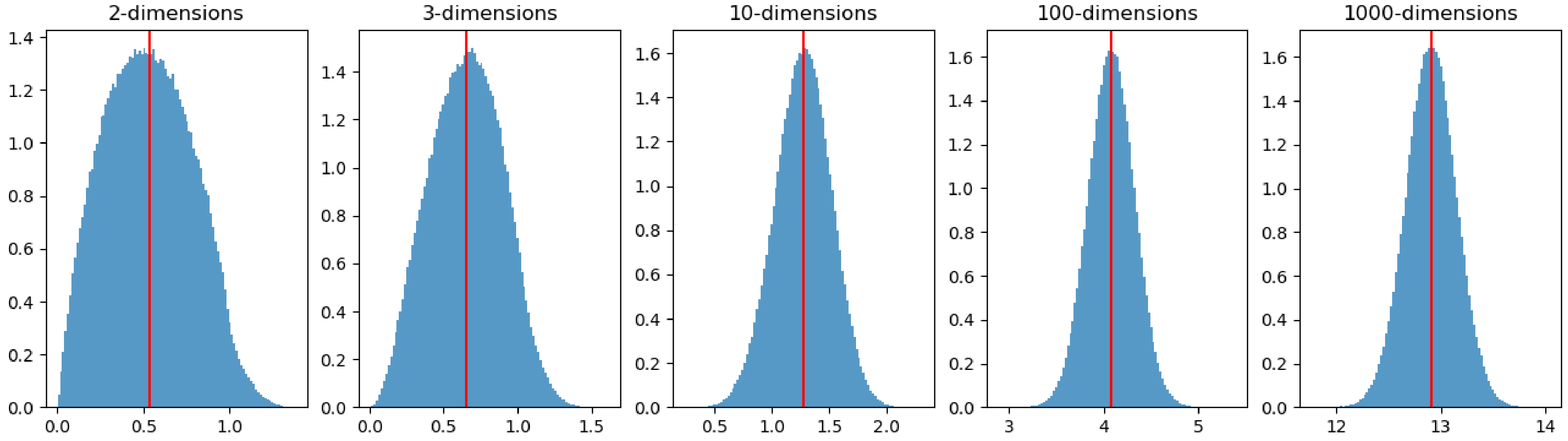
- Visualisation
    - Process data consists of many measurements
    - Viewing rows of time series or single feature statistics is not always the most effective method to identify *structure* in the data
    - Aim to *represent* high dimensional data in two (or maybe three) dimensions



Why dimensionality reduction

- Curse of dimensionality
  - Higher dimensions are unintuitive
  - In higher dimensions, features become more or less equidistant
  - Overfitting due to correlated features

Average distance between points in an n-d unit cube



Goal of dimensionality reduction

- Transform data in a high-dimensional feature space to a lower-dimensional space while retaining the most important *structures*
- Example of *unsupervised learning*: no output data

Supervised learning	**Unsupervised learning**
$f(X) \rightarrow y$	$f(X) \rightarrow ?$

- One easy classification between techniques:
  linear *vs* **non-linear** dimensionality reduction

Linear dimensionality reduction

$$T = XQ$$

- Scores T are a linear transformation of data X subject to constraints
- Usually, we require the transformation to preserve distance, which amounts to Q being a rotation and projection in space
- Most common method: Principal Component Analysis (PCA)

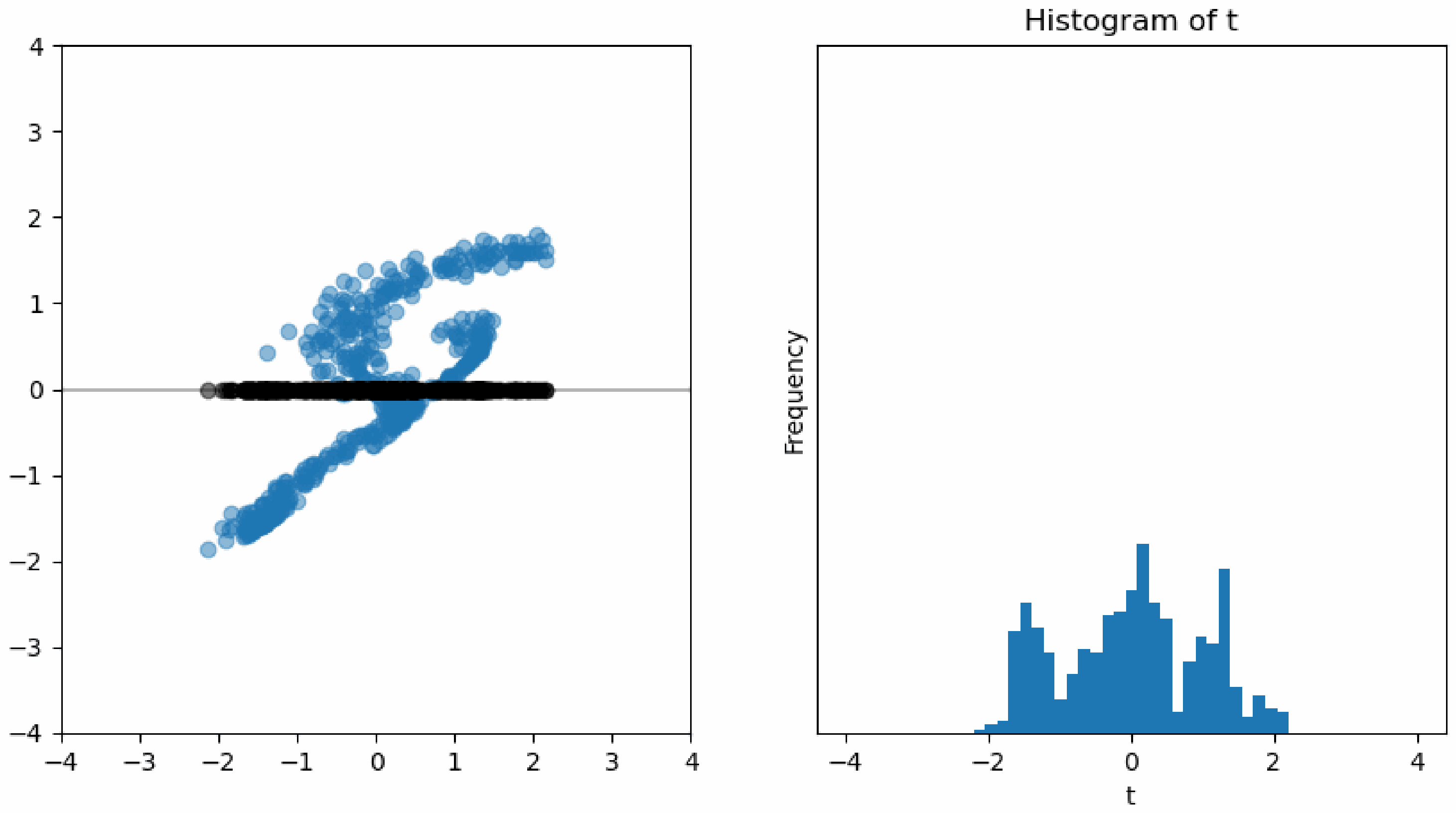
$$q_i = \arg\max_{q_i}[\mathrm{Var}(Xq_i)] \quad \mathrm{s.t.}\ q_i \cdot q_j = 0$$

https://scikit-learn.org/stable/modules/decomposition.html

Linear dimensionality reduction

$$T = XQ$$

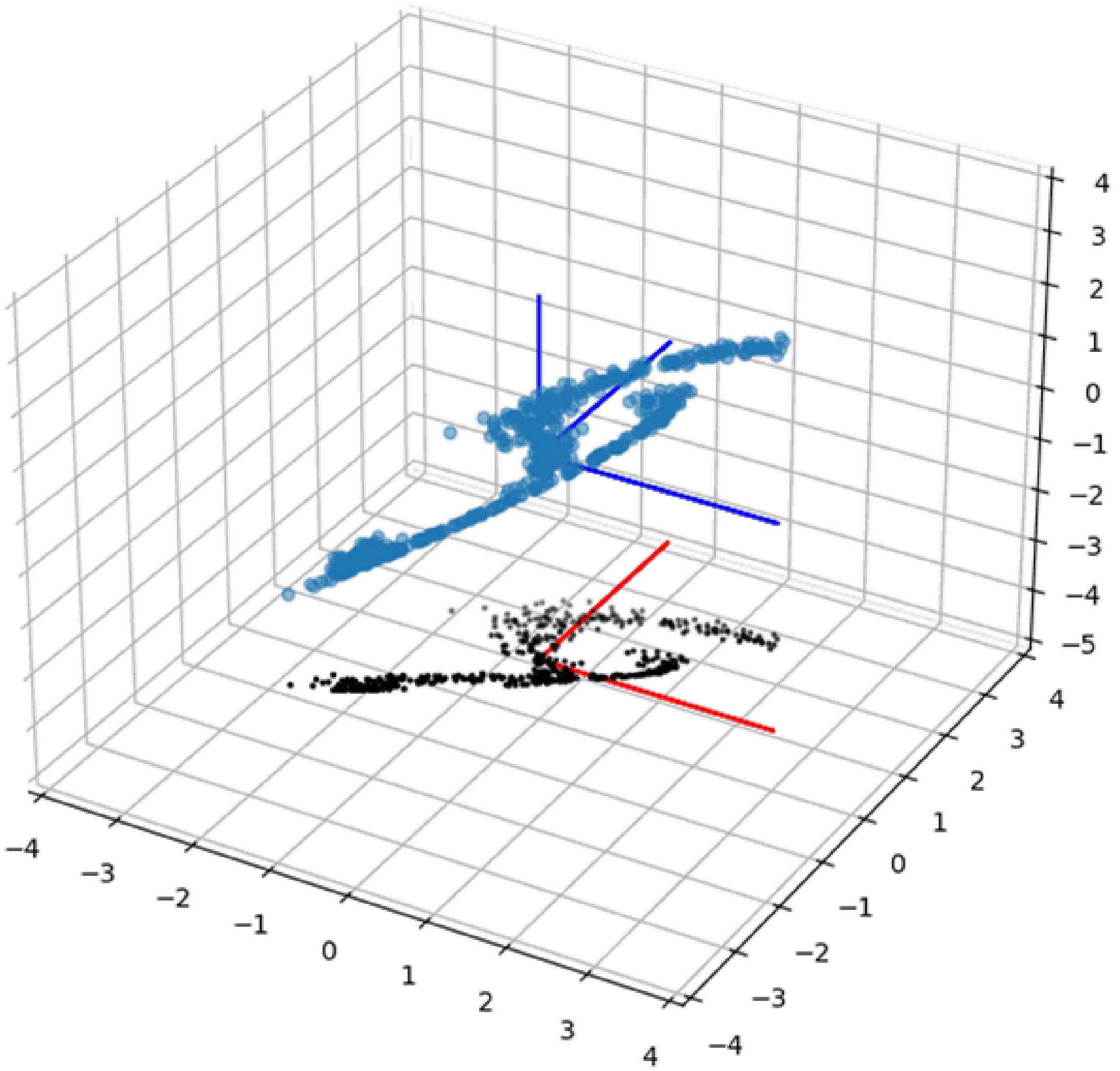
- Scores T are a linear transformation of data X subject to constraints
- Most common method: Principal Component Analysis (PCA)
- Q amounts to a rotation in space that maximizes variance in projected dimension



Linear dimensionality reduction

$$T = XQ$$

- Scores T are a linear transformation of data X subject to constraints
- Most common method: Principal Component Analysis (PCA)
- Q amounts to a rotation in space that maximizes variance in projected dimension



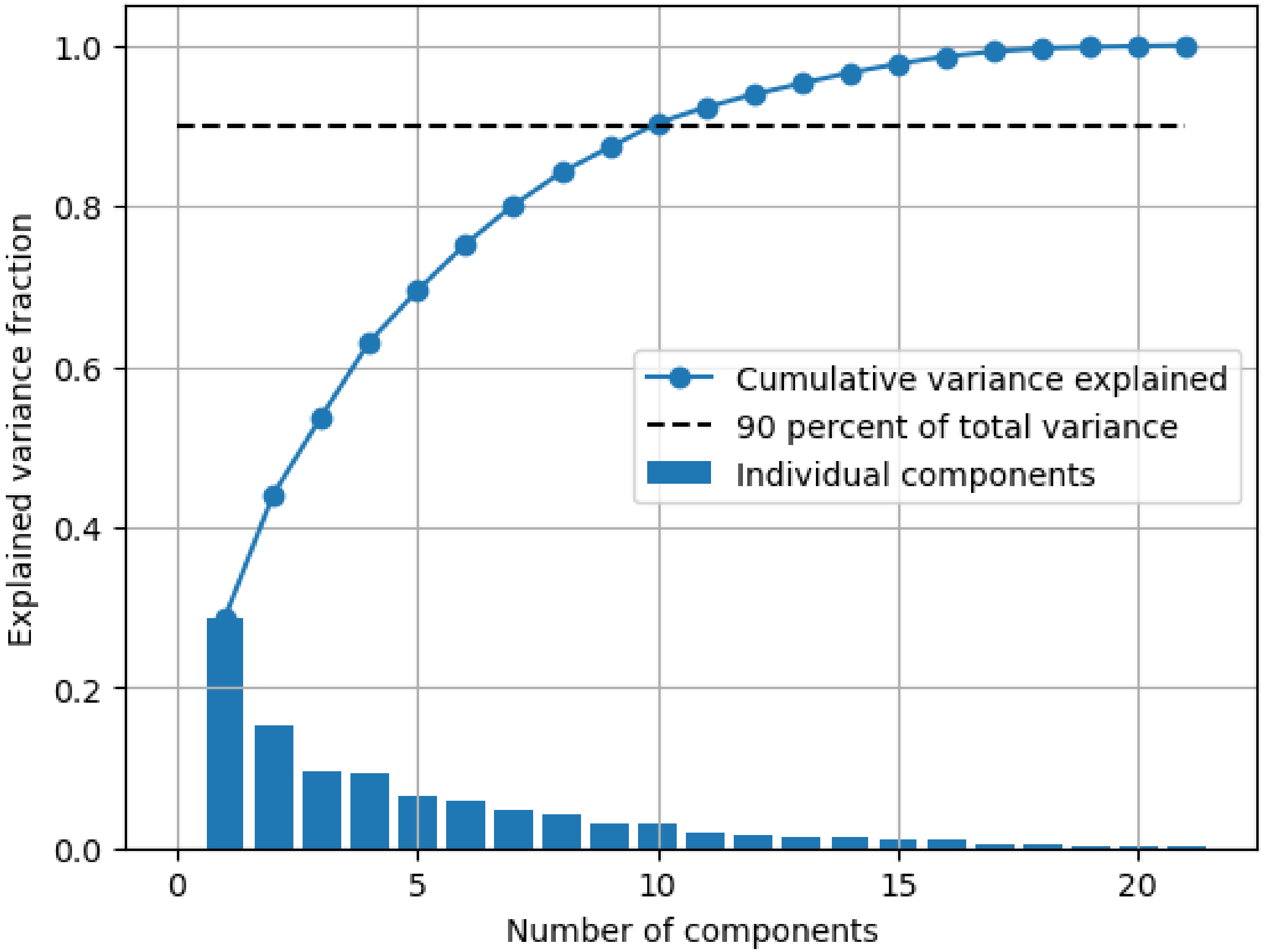
Linear dimensionality reduction

- Often consider the *fraction variance explained* by the principal components

$$\frac{\sum(\mathbf{t}_i - \boldsymbol{\mu}_t)^2}{\sum(\mathbf{x}_i - \boldsymbol{\mu}_x)^2}$$

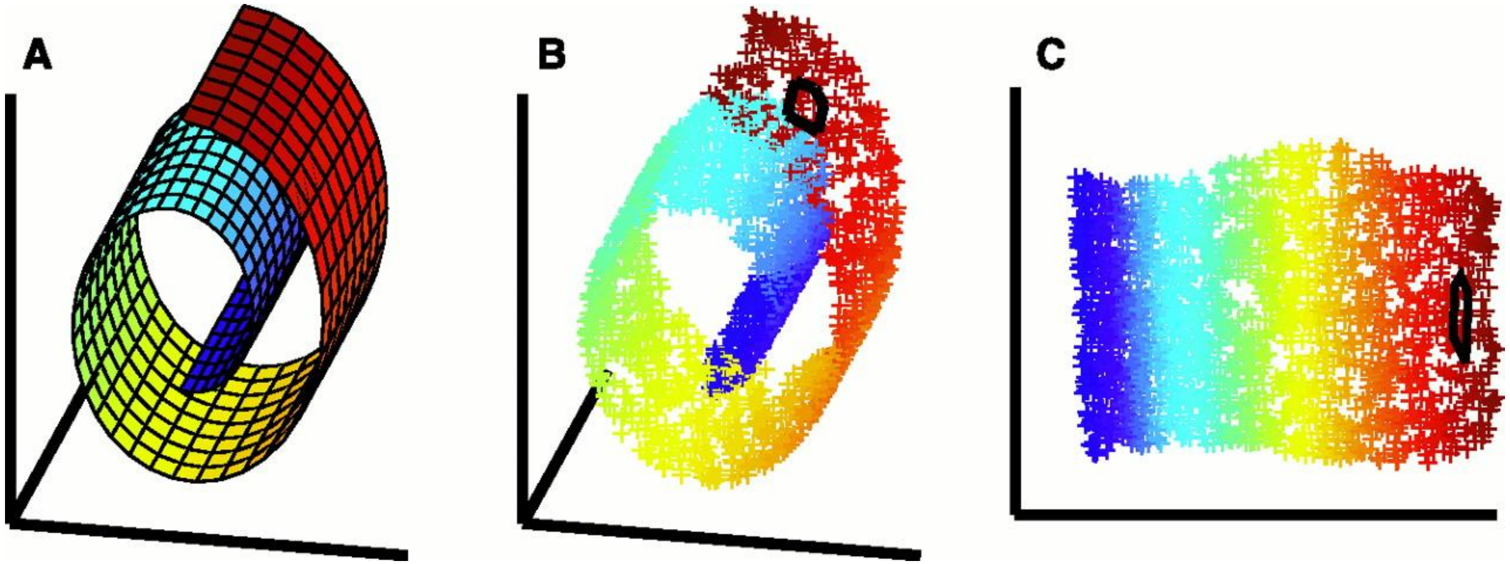
- Retain PCs which explain >90% (>95%?) of variance for further analysis

Scree plot



Non-linear dimensionality reduction

- Many methods for non-linear dimensionality reduction exist
  - Kernel PCA
  - t-SNE
  - Isomap
  - Spectral embedding
  - Autoencoders
  - ...



Locally Linear Embedding, Roweiss and Saul, 2000, https://doi.org/10.1126/science.290.5500.2323

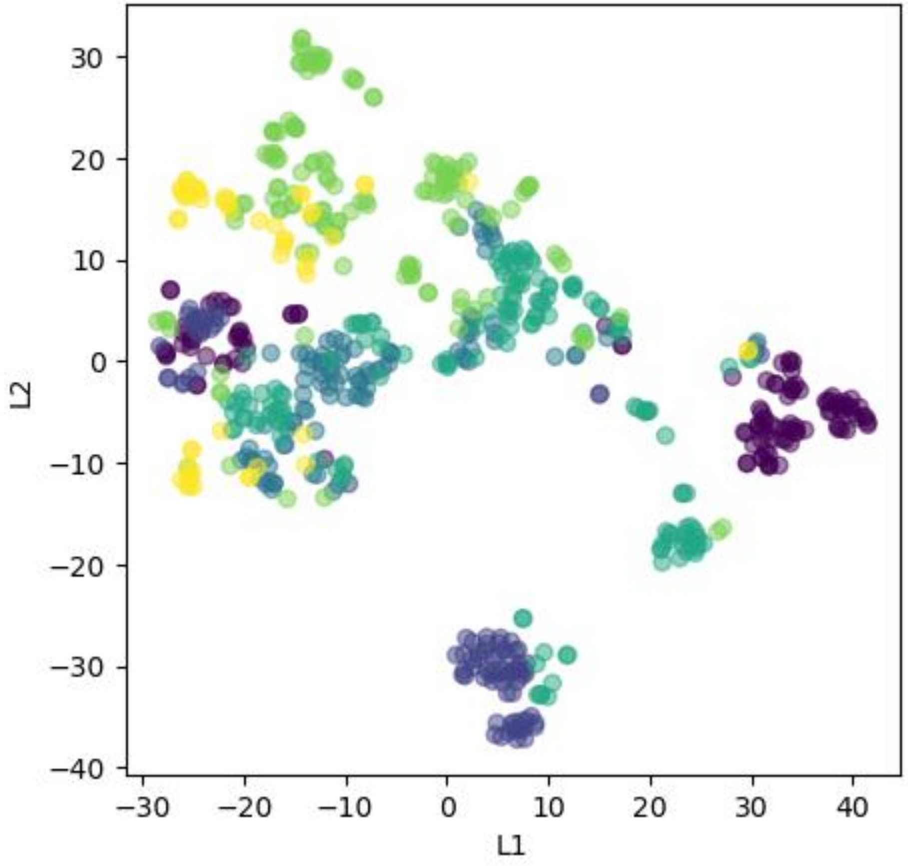
https://scikit-learn.org/stable/modules/manifold.html

Non-linear dimensionality reduction

- t-distributed Stochastic Neighbour Embedding (t-SNE), Van der Maaten and Hinton, 2008, http://jmlr.org/papers/v9/vandermaaten08a.html
- If x in original space and y in latent (embedded) space, minimize the "difference" between conditional distributions $p_{i|j}$ and $q_{i|j}$

$$p_{i|j} = \frac{\exp\left(-\frac{|x_i - x_j|^2}{2\sigma_i^2}\right)}{\sum_{k\neq i}\exp\left(-\frac{|x_i - x_k|^2}{2\sigma_i^2}\right)}$$

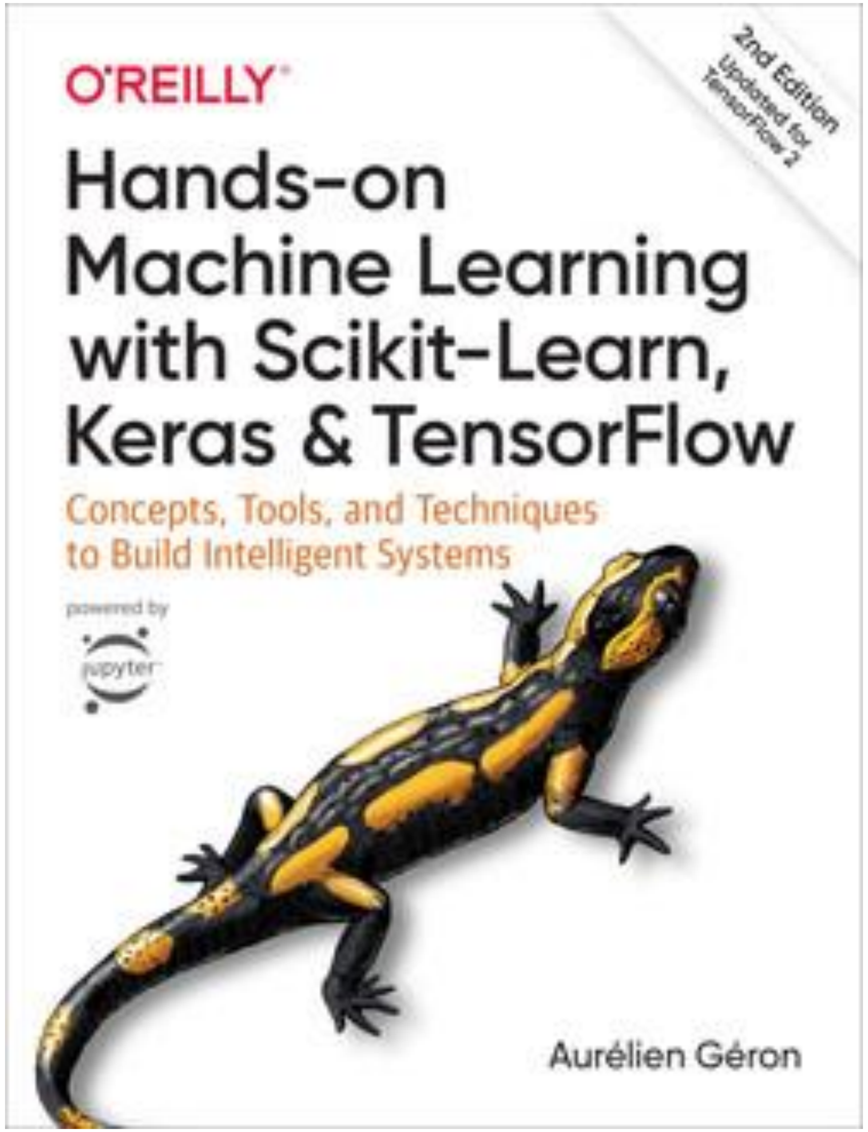
$$q_{i|j} = \frac{\exp\left(-|y_i - y_j|^2\right)}{\sum_{k\neq i}\exp(-|y_i - y_k|^2)}$$



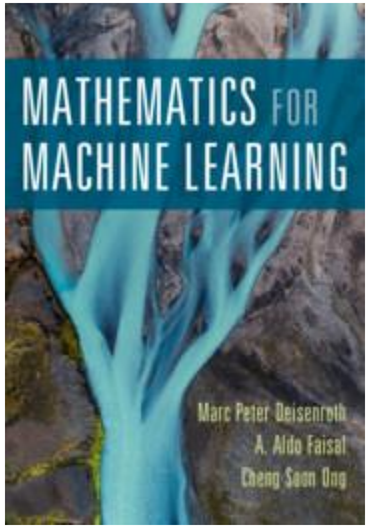
https://scikit-learn.org/stable/modules/manifold.html

Resources

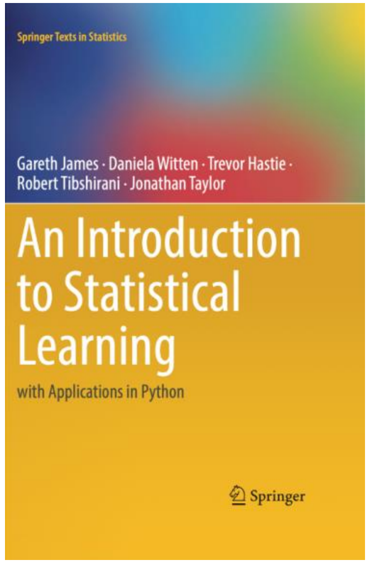
https://scikit-learn.org/stable/unsupervised_learning.html



https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/



https://mml-book.github.io/



https://www.statlearning.com/