HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

A practical guide to dimensionality reduction and feature selection

Elisabeth Georgii [1], Donatella Cea [1]

[1] Helmholtz AI, Helmholtz Munich

Schedule

- Day 1

09:00 - 09:20	Introduction (slides will be made available)
09:20 - 10:00	Breakout rooms: feature transformation notebooks
10:00 - 11:00	Self-study: feature aggregation & autoencoder notebooks
11:00 - 12:00	Breakout rooms: discussion of questions and tasks

- Day 2

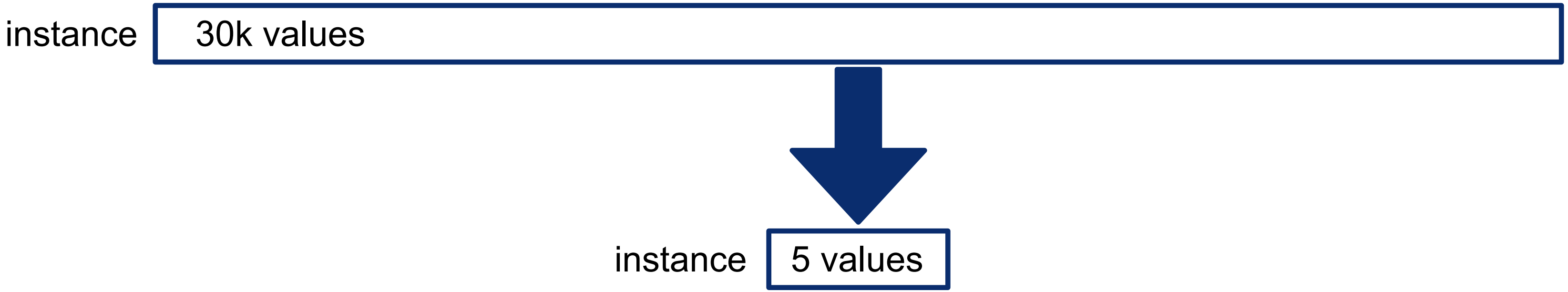
09:00 - 09:15	Questions regarding Day 1 material, introduction Day 2
09:15 - 10:30	Self-study: feature selection & stability optimization
10:30 - 11:45	Breakout rooms: discussion of questions and tasks
11:45 - 12:00	Wrap-up and conclusions

Definition of dimensionality reduction

- Given a dataset of n instances with p-dimensional measurement profiles, use a transformation to represent the n instances by k-dimensional feature profiles, with k<<p

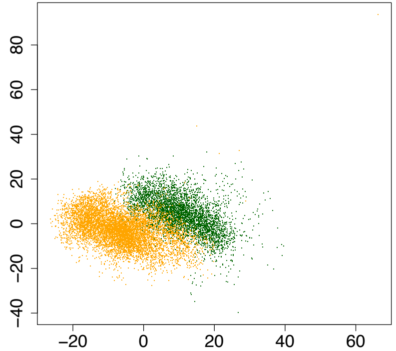
- Example: gene expression measurements
    - n=24 biological samples
    - p=30k genes

- Can we reduce the 30k-dimensional profile of each instance to a 5-dimensional profile?



Purpose of dimensionality reduction

- Visualize the main dissimilarities between instances (e.g., a subgroup structure)
  - Make data easier to visually grasp by humans
  - Not only useful for scatterplots of instances but also for heatmaps (number of columns)



- Facilitate machine learning analysis of instances
  - Clustering for detection of subgroups
  - Robust classification, e.g. healthy vs. cancer
  - Trajectory identification, e.g. cell development

Purpose of dimensionality reduction: further remarks

- Dimensionality reduction facilitates data visualization but not necessarily scientific interpretation: it is different from explainable AI (XAI)!
  - E.g. transformed features are composed of contributions from all original variables, so the influence of single variables may not be easy to trace
- Dimensionality reduction counteracts the curse of dimensionality in machine learning
  - As p increases, the neighborhood at a fixed distance of a specific point becomes less and less populated by instances; thus, instance similarities and groupings get harder and harder to detect
- Dimensionality reduction leads to computational advantages: less storage space, more efficient training of models

Overview of dimensionality reduction approaches

Dimensionality reduction

→ **Feature transformation**

- Linear methods
  - PCA
  - ICA
  - CCA (multimodal)
- Nonlinear methods
  - MDS
  - t-SNE
  - UMAP
  - Autoencoders

→ **Feature aggregation**

- Data-driven methods
  - kMedoid clustering
  - Graph partitioning
- Knowledge-based methods
  - GO slim categories

Mostly unsupervised: no target variable taken into account

→ **Feature selection**

- Filter methods
  - ReliefF
  - Correlation
- Wrapper methods
  - Forward selection
  - Backward elimination
- Embedded methods
  - Elastic net

Mostly supervised prediction

Remarks on supervised vs. unsupervised analysis

- Most often, feature transformation and feature aggregation are performed in an unsupervised manner, i.e. no target variable for the instances (e.g. class label, continuous output) is taken into account during dimensionality reduction
- However, there also exist supervised feature transformation methods, e.g. linear discriminant analysis, and feature aggregation may be targeted toward supervised tasks
- Feature selection methods are typically supervised, but there are also unsupervised methods, e.g. variance-based filters

References

- Altman, N., Krzywinski, M. The curse(s) of dimensionality. Nat Methods 15, 399–400 (2018). https://doi.org/10.1038/s41592-018-0019-x
- Hamer, V., Dupont, P. An importance weighted feature selection stability measure. Journal of Machine Learning Research 22, 1-57 (2021)
- Huang, S.H. (2015). Supervised feature selection: A tutorial. Artif. Intell. Res., 4, 22-37.
- https://www.frontiersin.org/articles/10.3389/fgene.2021.646936