# Exercise Sheet 9

1. Describe the general idea of dimension reduction, why it is particularly useful for high-dimensional data and what the drawbacks are.
   - Remove irrelevant dimensions (*feature selection*) or
   - Transform data in another space where less dimensions represent the data (*feature extraction/transform*)
     - Data are projected from HD to LD space with techniques that aim to preserve features.
   - DR involves a *loss of information*. Goal: preserve important structures, such as clusters, outliers, correlations and manifolds (a 1D manifold is a line or circle; a 2D manifold a plane, torus or sphere)
   - represents similarity (in HD) with proximity,
   - enables an overview of the data, its structure, relations and distribution,
   - supports navigation and browsing when the user supplies a query (with values and ranges for attributes) leading to a 2D point region
   - **Drawbacks:**
     - Dimensions need to be centralized (zero means) and normalized (divide by the range or σ) (*auto scaling*)
     - Drawback of auto scaling: Noisy measurements are scaled up whereas large peaks in meaningful data get reduced
     - Strongly correlating dimensions hamper the result (should be removed upfront, remember feature selection)
     - Interpretability of the new dimensions challenging; often domain scientists are not satisfied

2. Explain the relation between dimension reduction and subspace clustering
   - A dimension that does not contribute to a clusterable subspace probably is not important.
   - A dimension that contributes to clusters where users are confident in representing true phenomena likely should be preserved.
   - Data has a dimensionality given by the observations and it has a lower *intrinsic dimensionality* that captures most of the variance.
   - Dimension reduction is about finding the intrinsic dimensionality.

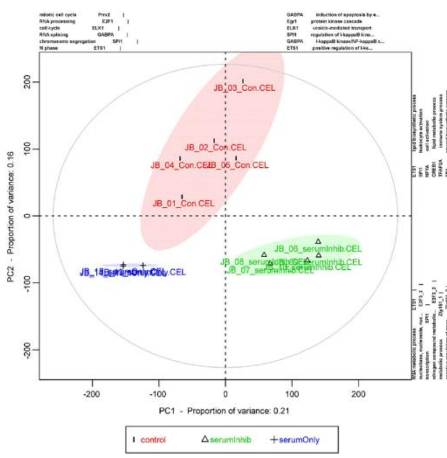
3. Explain the general concept of the linear dimension reduction method PCA.
   - generate a new set of dimensions that is a linear combination of the original dimensions
   - Based on the original dimensions $x_1, \ldots, x_p$ PCA generates a new coordinate system with orthogonal dimensions
   - New dimensions (Principle Components, PC) are *linear combinations* of the original dimensions and are *sorted according to variance*
   - Each PC carries a *loading* that characterizes how much variability of the data is explained.

- Starting from the dimension with the highest loading, take the first *n* dimensions until their cumulative loadings exceed a threshold, e.g. 95%
- The projection error (in a least square sense) is minimized for any selection of *n*

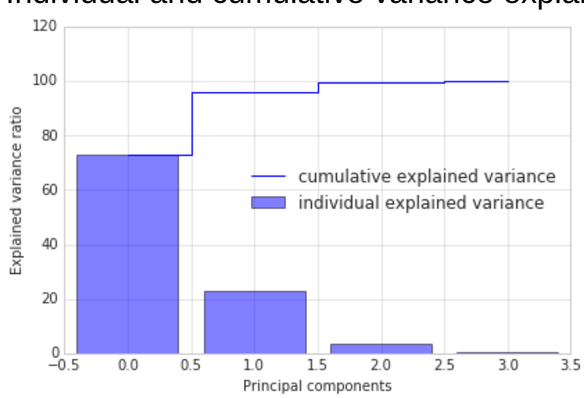
- **Approach:**
    - Normalize the data
    - Determine the covariance matrix $Cov = 1/(n-1)\, XX^T$
    - Apply an Eigenvalue analysis
    - $Cov = U \lambda U^T$
    - $\lambda$ is a diagonal matrix (all elements are zero except the diagonal) with Eigenvalues $\lambda_1 \leq \ldots \leq \lambda_p$
    - $U$ is a orthogonal matrix containing the Eigenvectors (sorted according to the Eigenvalues).

4. What are score plots and scree plots?

- **Score plots:** indicate distribution in the direction of the 2 or 3 largest components




- **Scree plots:** Individual and cumulative variance explained by the first *n* PCs



5. What is the motivation for non-linear dimension reduction methods?
   - suitable for skewed or multimodal (e.g. bimodal) distributions
   - Large distances are often not interesting or reliable.
   - Instead small distances often on a manifold should be preserved.
   - Non-linear techniques have more degrees of freedom

6. Explain the general concept of the non-linear dimension reduction method multidimensional scaling.

- Non-linear iterative optimization method where the distance of points in $R^n$ is preserved optimally when transforming to $R^k$
- An optimization problem based on a *stress function* is solved with a non-linear optimization method (e.g. gradient descent, simulated annealing)
- Gradient descent: simpler, but more sensitive to local minima
- MDS is a similarity-based projection in LD-space
- At the core of MDS is a **distance metric**.
  - Typical choices are
  - Euclidean distance,
  - city block distance or
  - angle between feature vectors, e.g. in text analytics, where large set of documents are shown
- Users may select a metric, fix certain points as constraints, employ cluster labels (in the distance metric)

$$\sum_{j=1 \ldots N-1} \sum_{i=i+1..j} \left( \|(x_i, y_i) - (x_j, y_j)\|) - d_{ij} \right)^2$$

The true distance of the points in HD space

The distance in LD space

7. Describe forms of interaction between an analyst and the dimension reduction algorithm.

   - Interactive PCA
   - Users may:
     - adjust weights for each dimension
     - manipulate points (e.g. by selecting a rectangular region and dragging it)
     - see different aspects in 4 coordinated views
   - ***Visual hierarchical DR***
     - A hierarchy is created by means of dissimilarity and used to select dimensions
     - Dimensions are analyzed w.r.t. similarity and are clustered.

    - Clustering is performed with different similarity thresholds, leading to a hierarchy of clusters
    - For each low level cluster, a representative dimension is computed or selected by the user.
    - Dimension hierarchy is navigated to select/deselect clusters and dimensions.
    - Results are shown with PC plots, SP matrices
  - ***Interactive DR through user-defined quality metrics***
    - QM relates to the preservation of correlation, clusters, …
  - ***Assisted search for multiple subspaces***
    - A combination of visual representations (views) is used to interpret subspaces.
    - Filters enable reduction in item and dimension space.
    - Reduction may be started by automatically detected patterns (A) or by user-defined subspaces.
    - Subspaces are iteratively refined
  - ***Guidance* for dimensional reduction**
    - More comprehensive support, including filtering, feature transformation

8. Name three application examples for linear and non-linear dimension reduction
   - Show the results of subspace clustering
   - Text analytics (visualization of textual documents with various tags and properties such as creation date, word count, authors