Multidimensional Data Visualization

Overview

Strategies for Multidimensional Data Visualization

- Direct visualization methods, where each feature, characterizing a multidimensional object, is represented in a visual form;
- Projection, so-called dimensionality reduction, methods, allowing us to represent the multidimensional data on a low-dimensional space. Artificial neural networks may also be used for visualizing multidimensional data. They realize various nonlinear projections.

Direct Visualization

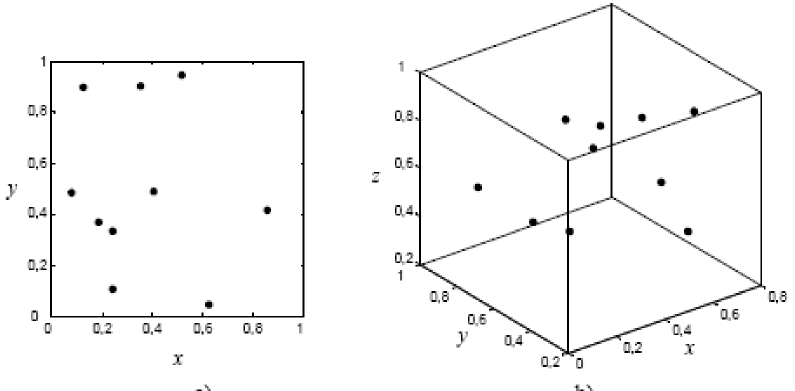
- The direct data visualization is a graphical presentation of the data set that provides a qualitative understanding of the information contents in a natural and direct way.
- The commonly used methods are scatter plot matrices, parallel coordinates, Andrews curves, Chernoff faces, stars, dimensional stacking, etc.
- The direct visualization methods do not have any defined formal mathematical criterion for estimating the visualization quality.
- Each of the features $x_1, x_2, \dots, x_n$ characterizing the object $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i \in \{1, \dots, m\}$, is represented in a visual form acceptable for a human being.

Direct Visualization Methods

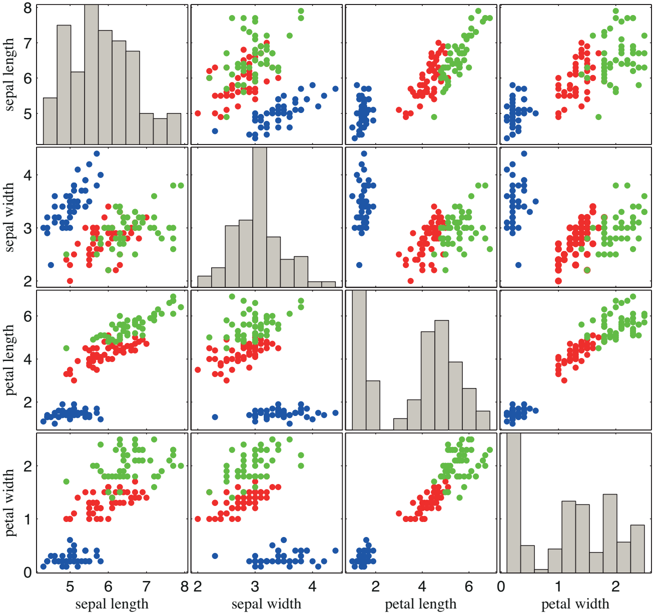
1. Geometric methods:
   a) scatter plots,
   b) matrix of scatter plots,
   c) multiline graphs,
   d) permutation matrix,
   e) survey plots,
   f) Andrews curves,
   g) parallel coordinates,
   h) radial visualization (RadViz) and its modifications GridViz and PolyViz.
2. Iconographic displays:
   a) Chernoff faces,
   b) star glyphs,
   c) stick figure,
   d) color icon.
3. Hierarchical displays:
   a) dimensional stacking,
   b) trellis display,
   c) hierarchical parallel coordinates.

Geometric Methods

- Geometric visualization methods are the methods where multidimensional points are displayed using the axes of the selected geometric shape.
- *Scatter plots* are one of the most commonly used techniques for data representation on a plane $\mathbb{R}^2$ or space $\mathbb{R}^3$. Points are displayed in the classic (x, y) or (x, y, z) format.



Scatter Plot Matrix of the Iris Data



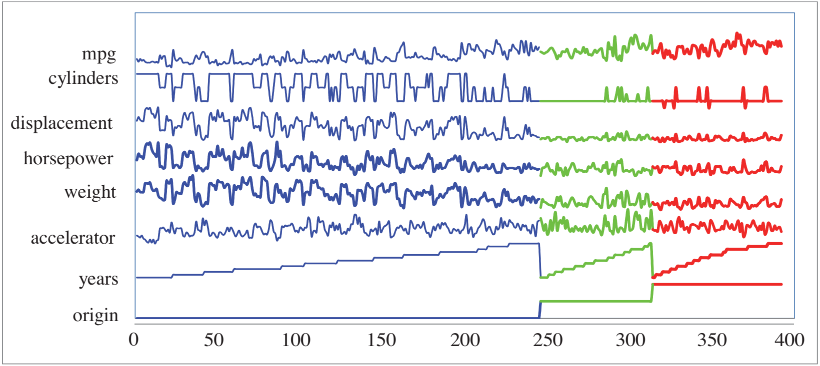
Setosa

Versicolor

Virginica

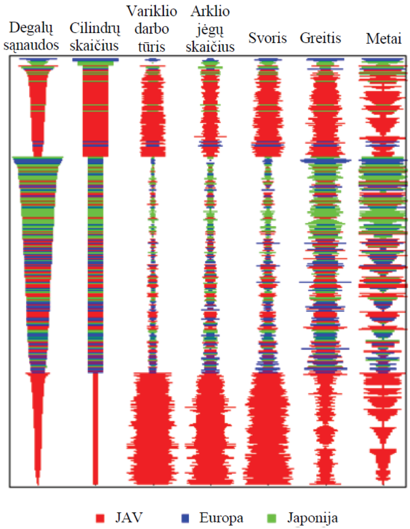
Multiline Graphs of the Auto MPG Data

- The data are aligned according to origin (USA, Japan, Europe).

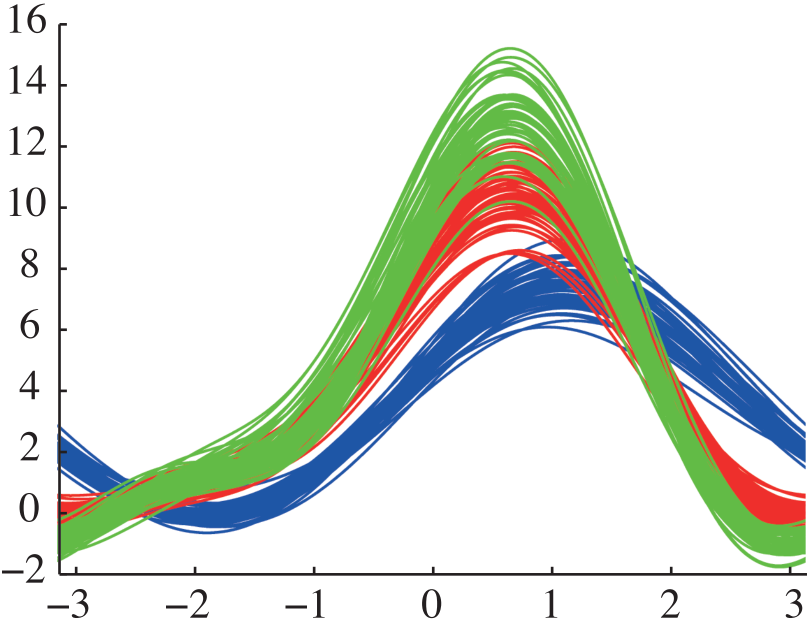


Survey Plot of Auto MPG Data

- Sorted according to the number of cylinders and MPG.

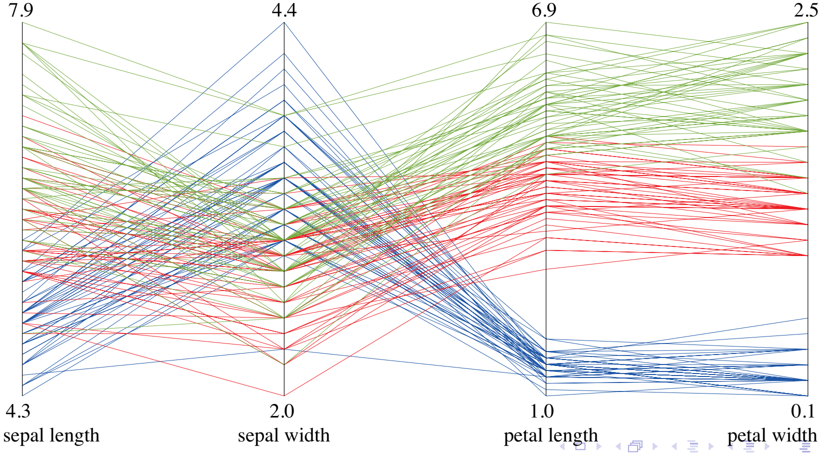


Andrews Curves of Iris Data



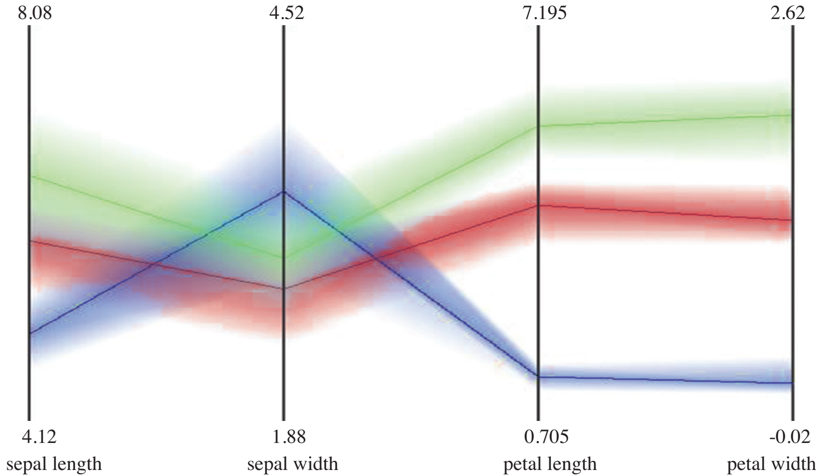
Iris Data Represented on the Parallel Coordinates

- The image is obtained using the system *Orange*.
- Different colors correspond to the different species.
- We see that the species are distinguished best by the petal length and width. It is difficult to separate the species by the sepal length and width.



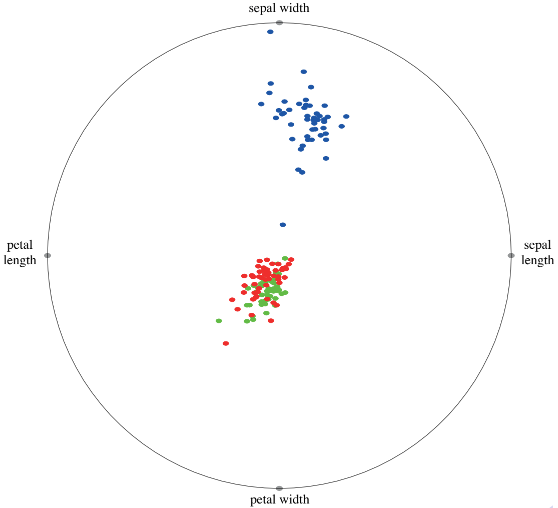
Iris Data Represented on the Hierarchical Parallel Coordinates

- The image is obtained using the system *Xmdv*.

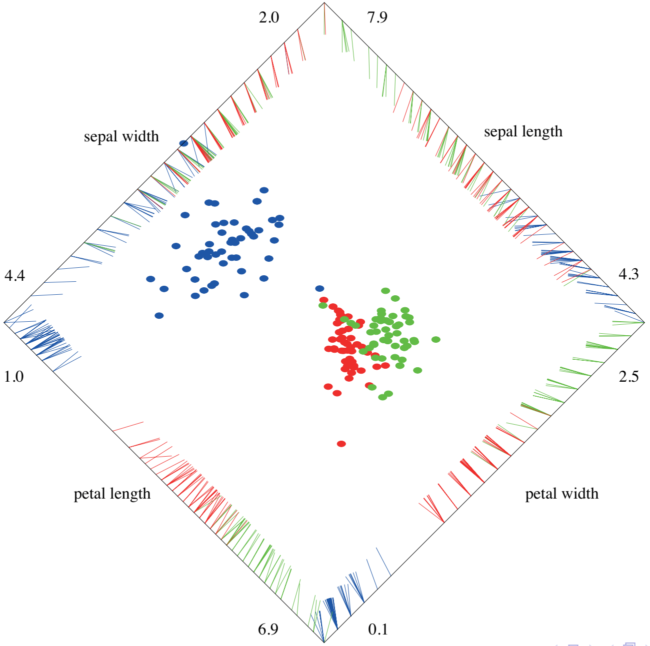


Iris data visualized by the RadViz method

- The petal length, petal width, sepal length, and sepal width are dimensional anchors. The image is obtained using the system *Orange*.



Iris Data Visualized by the PolyViz Method

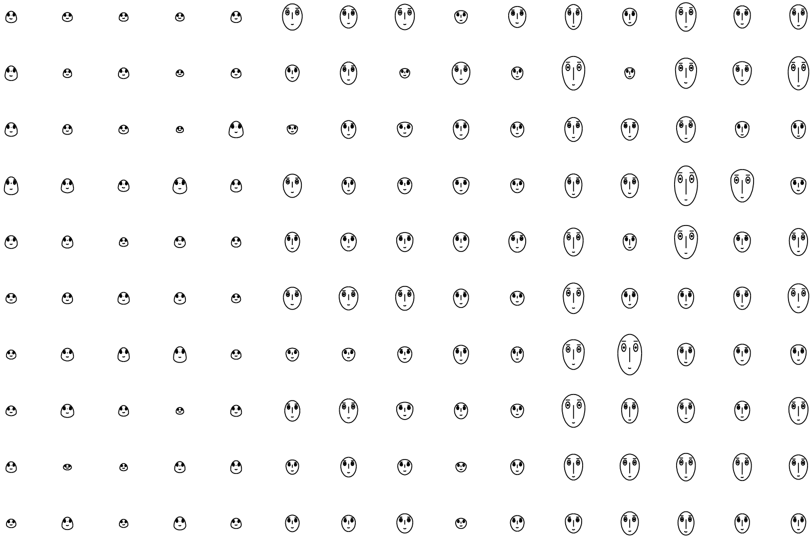


Iconographic Displays

- The aim of visualization of multidimensional data is not only to map the data onto a two- or three-dimensional space, but also to help perceiving them.
- The second aim may be achieved visualizing multidimensional data by *iconographic display* methods. They are also called *glyph* methods.
- Each object that is defined by the n features is displayed by a glyph. Color, shape, and location of the glyph depend on the values of features.
- The most famous methods are *Chernoff faces* and the *star* method, however some methods of more complicated other glyphs may be used as well.

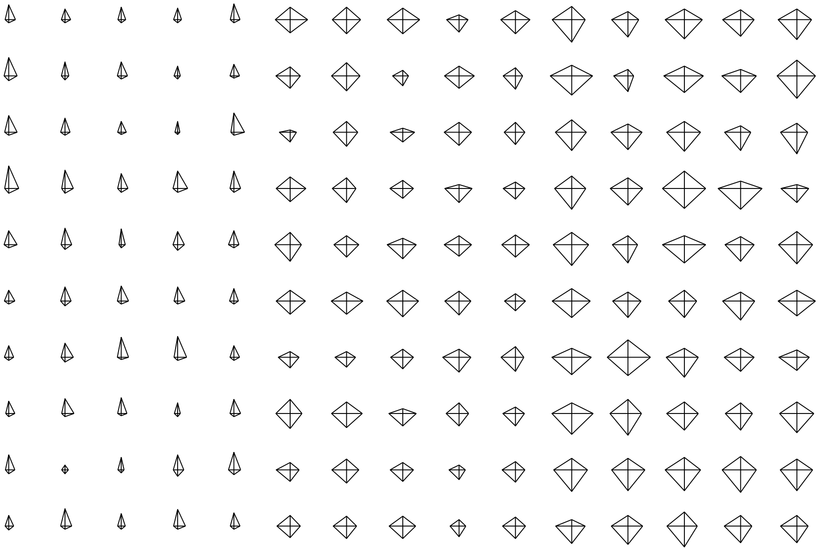
Iris Data Visualized by Chernoff Faces

- Sepal length corresponds to the size of face, sepal width – shape of forehead, petal length – shape of jaw, and petal width – length of nose. *Matlab* is used to obtain the image.



Iris Data Set Visualized by Star Glyphs

- The stars, corresponding to Setosa irises, are smaller than the other. The larger stars correspond to Virginica irises.

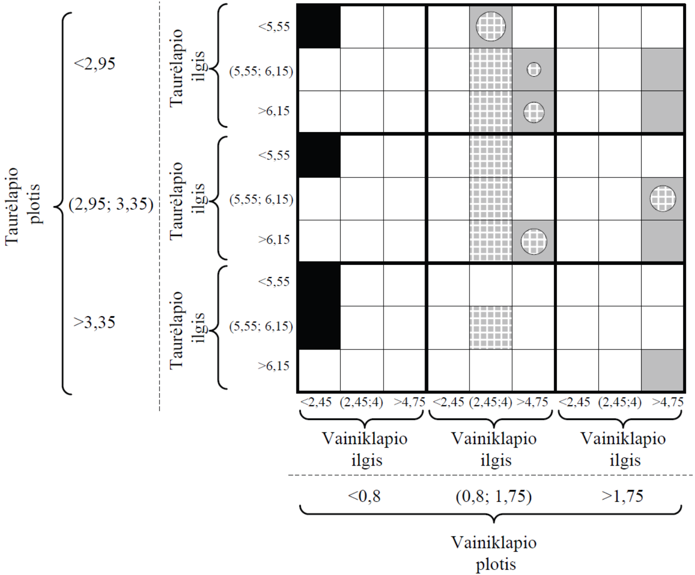


Hierarchical Displays

- *Hierarchical displays* create a structure of an image such that some features are embedded in displays of other features.
- Visualization of some features is displayed in the structure depending on the values of other features.
- Here we discuss two such techniques: *dimensional stacking* and *trellis display*.

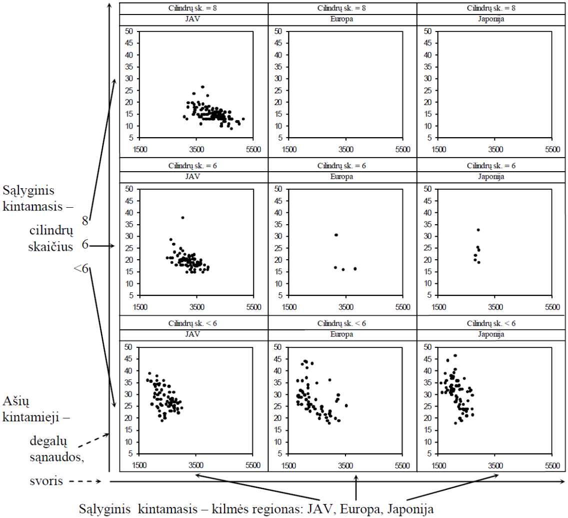
Iris data visualized by dimensional stacking

- Setosa irises (black cells) are displayed separately from the other two species. The other two species overlap.



Auto MPG Data Visualized by Trellis Display

- The axis variables are weight and miles per gallon (MPG).
- The origin and number of cylinders are conditioning variables.



Projection Methods

- Methods that allow us to represent multidimensional data from $\mathbb{R}^n$ in a low-dimensional space $\mathbb{R}^d$, $d < n$, are called projection (dimensionality reduction) methods.
- If the dimensionality of the *projection space* is small enough ($d = 2$ or $d = 3$), these methods may be used to visualize the multidimensional data.
- In such a case, the projection space can be called a *display*, *embedding* or *image space*.
- These methods usually invoke formal mathematical criteria by which the projection distortion is minimized.
  1. Linear projection methods:
     a) principal component analysis,
     b) linear discriminant analysis,
     c) projection pursuit.
  2. Nonlinear projection methods:
     a) multidimensional scaling,
     b) locally linear embedding,
     c) isometric feature mapping,
     d) principal curves.

Criteria of the Projection Quality

- There are some formal mathematical criteria of the projection quality. These criteria are optimized in order to get the optimal projection of multidimensional data onto a low-dimensional space.
- The main goal is to preserve the proportions of distances or estimations of other proximities between the multidimensional points in the image space as well as to preserve, or even to highlight other characteristics of the multidimensional data (for example, clusters).

Linear Transformation

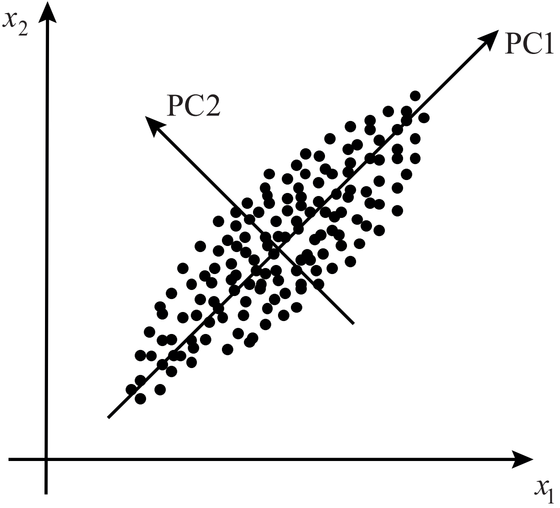
- There are linear and nonlinear projection methods.
- Linear projection methods pursue a linear transformation of data. There are various linear transformations: rotation, shearing, reflection, scaling, etc.
- A *linear transformation* may be described by linear equations
$$Y_i = X_i A.$$
- If $d = n$, i.e. $Y_i = (y_{i1}, y_{i2}, \ldots, y_{in})$ and $X_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, then A is a square matrix, consisting of n rows and n columns. The matrix A is called a *transformation matrix*.
- If a linear transformation is used for dimensionality reduction, then $d < n$, $Y_i = (y_{i1}, y_{i2}, \ldots, y_{id})$, $i = 1, \ldots, m$, and A is a matrix, consisting of n rows and d columns.

Principal Component Analysis

- The principal component analysis (PCA) is a well-known data analysis technique invented in 1901 by Pearson.
- It is a way of linear transforming a set X of n-dimensional points $X_1, X_2, \ldots, X_m$ into another set Y of n-dimensional points $Y_1, Y_2, \ldots, Y_m$.
- The property of the set is that the largest part of its information content is stored in the first few coordinates (components) of points Y_i, $i = 1, \ldots, m$.
- The principal component analysis is often used to reduce the dimensionality of multidimensional points X_i, $i = 1, \ldots, m$, by discarding some of the components of the points Y_i and by leaving only the first (principal) d ones.
- The principal component analysis projects the data linearly into a low-dimensional space preserving the variance of the data best.

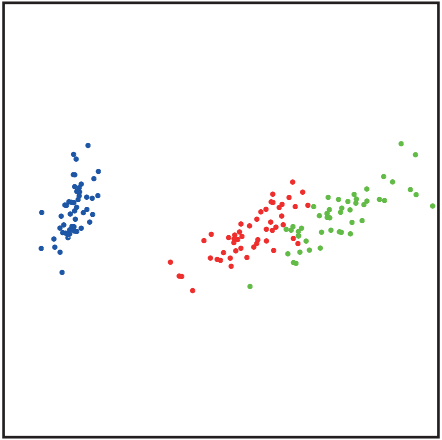
Illustration of Principal Component Analysis

- Figure illustrates a two-dimensional case with two principal components PC1 and PC2.



Iris Data Set Visualized Using PCA

- Setosa irises (marked in blue) are faraway from Versicolor (red) and Virginica (green) irises. There is no exactly expressed boundary between these two species.

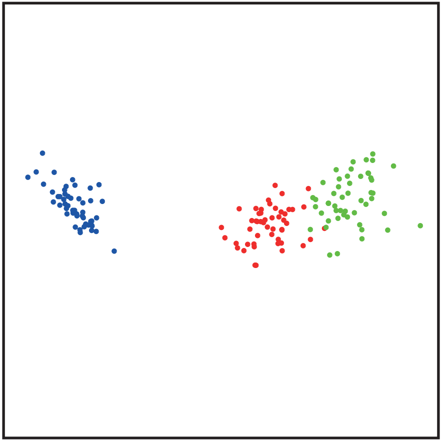


Linear Discriminant Analysis

- In contrast to most other dimensionality reduction methods, a *linear discriminant analysis* (LDA) is a supervised method.
- The method is often called Fisher's discriminant analysis.
- In a supervised strategy, some known properties of data (for example, belonging of the objects to one of classes) are applied.
- LDA transforms multidimensional data to a low-dimensional space, maximizing the linear separability between objects belonging to different classes.

Iris Data Set Visualized Using LDA

- The difference between LDA and PCA is that, in addition, the known classes of objects are applied.



Nonlinear Transformation

- A nonlinear transformation may be described as follows:

$$Y = f(X),$$

where f is a nonlinear function and

$$Y = \{Y_1, Y_2, \dots, Y_m\} = \{y_{ij},\ i = 1, \dots, m,\ j = 1, \dots, n\}.$$

- The nonlinear transformation is more complicated than the linear one and requires more time-consuming computations. However, such a transformation allows us to preserve the characteristics of multidimensional data better as compared with the linear transformation if $d < n$, i.e. the data are projected to a lower-dimensional space.

Multidimensional scaling (MDS) – a technique for exploratory analysis of multidimensional data

- Pairwise dissimilarities between n objects are given by a matrix (δ_{ij}), $i, j = 1, \dots, n$, it is supposed that $\delta_{ij} = \delta_{ji}$.
- The points representing objects in an m-dimensional embedding space $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \dots, n$ should be found whose inter-point distances fit the given dissimilarities.
- The problem is reduced to minimization of a fitness criterion, e.g. so called *Stress* function

$$S(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(d\left(\mathbf{x}_i, \mathbf{x}_j\right) - \delta_{ij} \right)^2 ,$$

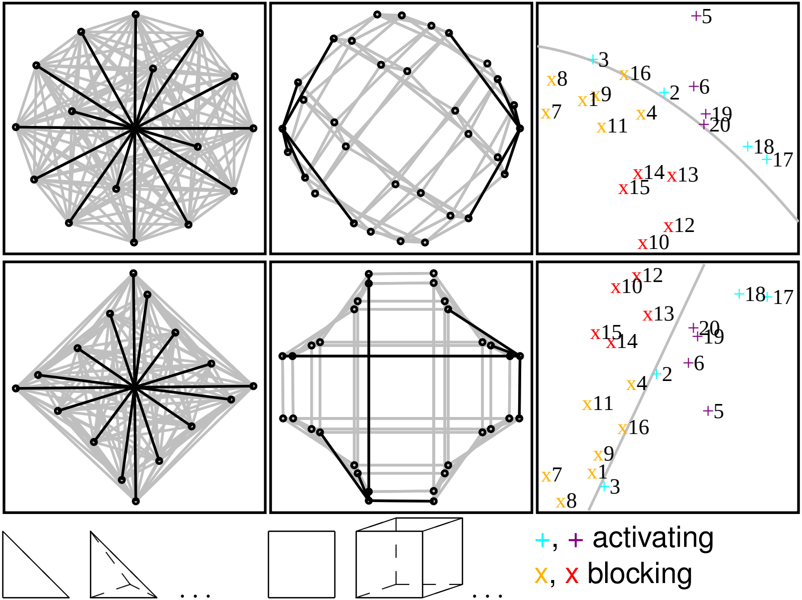
where $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$; $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between the points $\mathbf{x}_i$ and $\mathbf{x}_j$; weights $w_{ij} > 0, i, j = 1, \dots, n$.

MDS is a difficult global optimization problem

- Although *Stress* function is defined by an analytical formula which seems rather simple, it normally has many local minima.
- The problem is high dimensional: $\mathbf{x} \in \mathbb{R}^N$ and the number of variables is equal to $N = n \times m$.
- *Stress* function is invariant with respect to translation, rotation and mirroring.
- Smoothness of *Stress* function depends on distances $d(\mathbf{x}_i, \mathbf{x}_j)$, however, non-differentiability normally cannot be ignored. Minkowski distances

$$d_r(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|^r \right)^{1/r}.$$

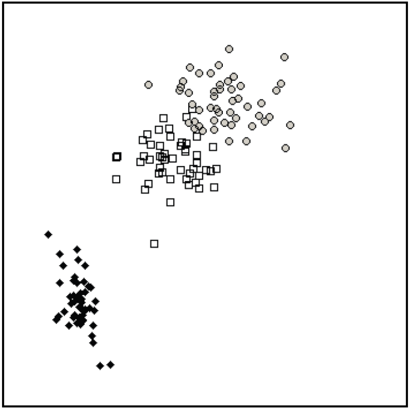
MDS with Euclidean and city-block distances



+, + activating

x, x blocking

Iris Data Set Visualized Using Multidimensional Scaling

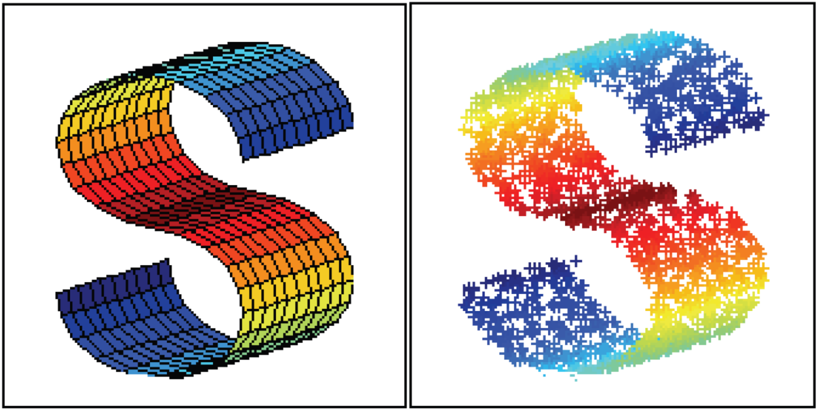


Isometric feature mapping (ISOMAP)

- *Isometric feature mapping* (ISOMAP) can be assigned to the group of multidimensional scaling. ISOMAP is designed for dimensionality reduction as well as for visualization of multidimensional data. An assumption is made that the multidimensional points are located on a lower-dimensional manifold. Therefore geodesic distances are used as a measure of proximity between the multidimensional points.
- Usually Euclidean distances between the points (as a proximity measure) are used in multidimensional scaling. In this case, the existence of a manifold is not taken into consideration.
- In ISOMAP, the geodesic distance is a proximity measure between the multidimensional points. A *geodesic distance* is the length of the shortest path between two points along the surface of a manifold.

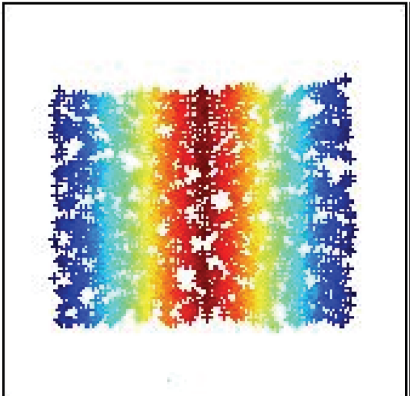
S-manifold

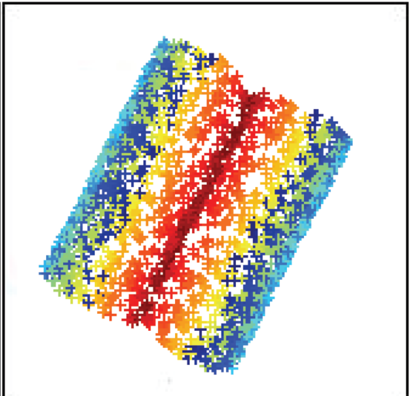
- The S-manifold is presented, $n = 3$. The points on the manifold are also shown, $m = 1000$.



ISOMAP and MDS Projections of S-manifold

- The structure of the manifold is well preserved by ISOMAP, because the S-manifold is unfolded: the farthest points on the manifold remain the farthest ones on the projection.
- The farthest points obtained by MDS are pale blue. These points are the farthest in multidimensional space in the sense of Euclidean distances, but they are not farthest in the sense of geodesic distances on the S-manifold.

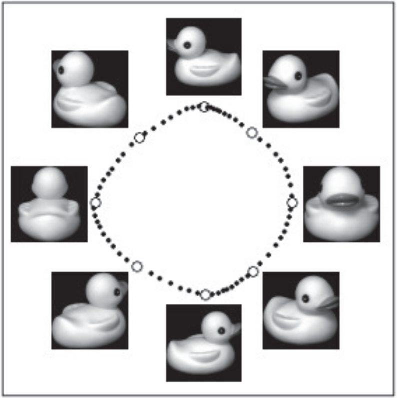




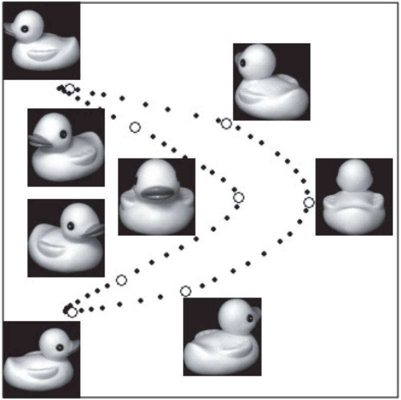
Locally Linear Embedding

- *Locally linear embedding* (LLE) is a nonlinear method for dimensionality reduction and manifold learning.
- Given a set of data points distributed on a manifold in a multidimensional space, LLE is able to project the data to a low-dimensional space by unfolding the manifold.
- LLE works by assuming that the manifold is well sampled, i.e. there are enough data, each data point and its neighbors lie on or close to a locally linear patch.
- Therefore, a data point can be approximated as a weighted linear combination of its neighbors. The basic idea of LLE is that such a linear combination is invariant under linear transformations (translation, rotation, and scaling) and, therefore, it should remain unchanged after the manifold has been unfolded to a low-dimensional space.
- The low-dimensional configuration of data points is obtained by solving two least squares optimization problems.

Visualization of Pictures of a Rotating Duckling by LLE



$k = 4$



$k = 9$