Computational Intelligence: Methods and Applications

Lecture 4
CI: simple visualization.

Włodzisław Duch
SCE, NTU, Singapore
Google: Duch

2D projections: scatterplots

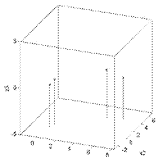
Simplest projections: use scatterplots, select only 2 features.
Example: sugar – teeth decay.

If d=3 than d(d-1)/2=3 subsets in 2D are formed and sometimes displayed in one figure.

Each 2D point is an orthogonal projection from all other d-2 dimensions.

What to look for:

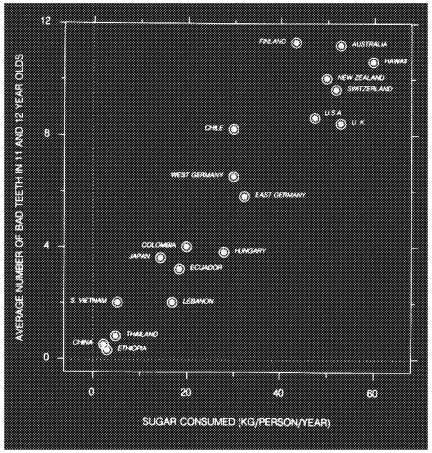
- correlations between variables,
- clustering of different objects.



Problem: for discrete values data points overlap.

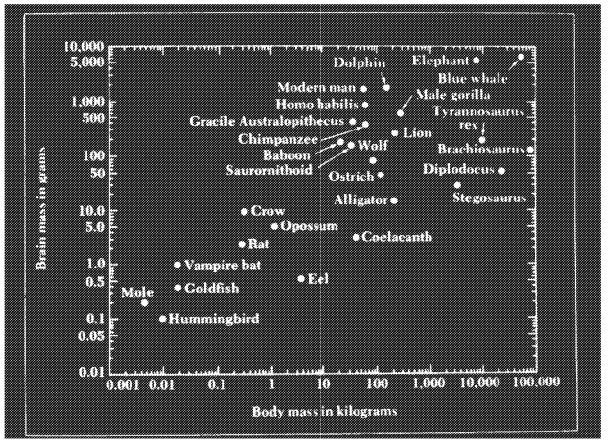
Extreme case: binary data in many dimensions, all structure is hidden, each scatterogram shows 4 points.

Sugar example



What conclusion can we draw?
Can there be alternative explanations?

Brain-body index example



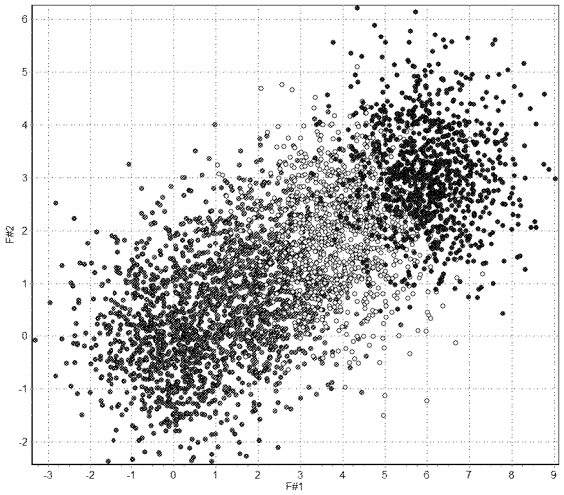
What conclusion can we draw?
Are whales and elephants smarter than man?

4 Gaussians in 8D, X_1 vs. X_2

Scatterograms of 8D data in F1/F2 dimensions.

4 Gaussian distributions, each in 4D, have been generated, the red centered at (0,0,0,0), green at (1,1/2,1/3,1/4), yellow at 2(1,1/2,1/3,1/4) and blue at 3(1,1/2,1/3,1/4)

Demonstration of various projections using Ghostminer software.

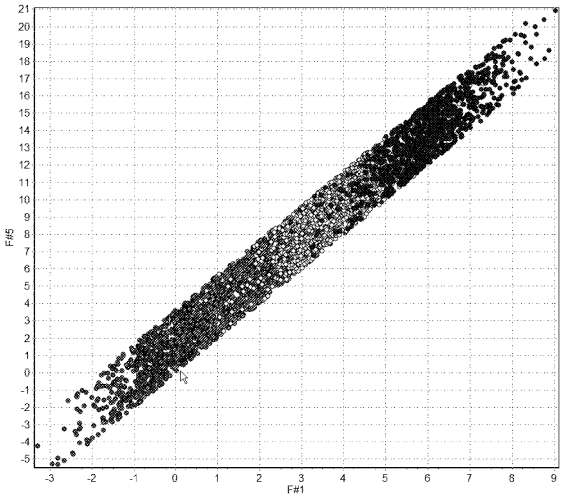


4 Gaussians in 8D, X_1 vs. X_5

What happened here?

All X_i vs. X_{i+4} have this kind of plots.

How were the remaining 4 features generated?



Cars example

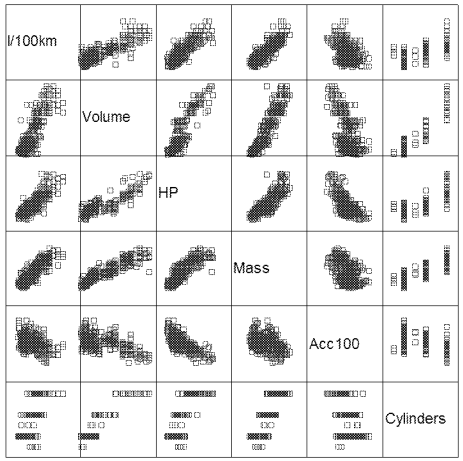
Scatterograms for all feature pairs, data on cars with 3, 4, 5, 6 or 8 cylinders.

To detailed? We are interested in trends that can be seen in probability density functions.

Cluster all points that are close for cars with N cylinders. This may be done by adding Gaussian noise with a growing variance to each point.

See this on the movie:

Movie for cars.



Direct representation: GT

How to deal with more than 3D? We cannot see more dimensions.

Grand Tour: move between different 2D projections; implemented in XGobi, XLispStat, ExplorN software packages.

Ex: 7D data viewed as scatterplot in Grand Tour

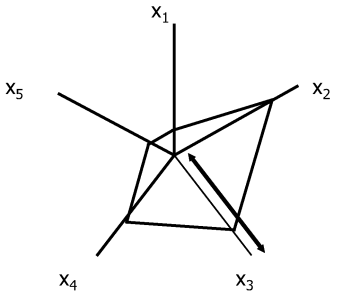
More examples:
http://www.public.iastate.edu/~dicook/JSS/paper/paper.html

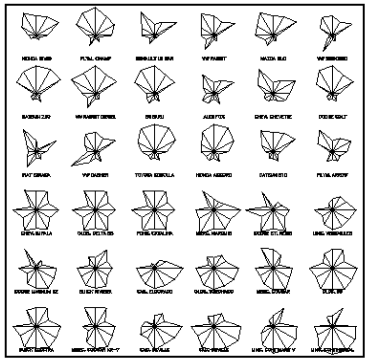
Try to view 9D cube – most of the time looks like Gaussian cloud.

It may take time to “calibrate our eyes” to imagine high-D structure.

Direct representation: star

Star Plots, radar plots:
represent the value of each component in a “spider net”.





Useful to display single or a few vectors per plot, uses many plots.
Too many individual plots? Cluster similar ones, as in the car example.

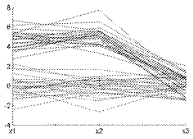
Direct representations: ||

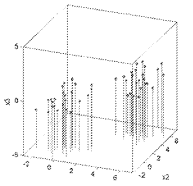
Parallel coordinates: instead of perpendicular axes use parallel!
Many engineering applications, popular in bioinformatics.

Two clusters in 3D

Instead of creating perpendicular axes
put each coordinate on the horizontal x axis
and its value on the vertical y axis.

Point in N dim => line with N segments.



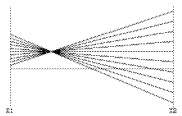


See more examples at:
http://www.nbb.cornell.edu/neurobio/land/PROJECTS/Inselberg/

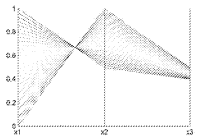
|| lines

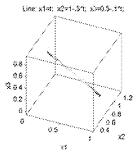
Lines in parallel representation:

2D line

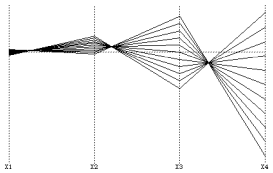


3D line





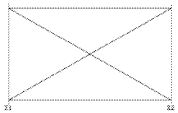
4D line



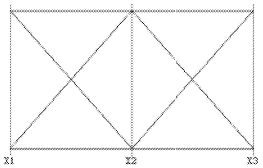
|| cubes

Hypercubes in parallel representation:

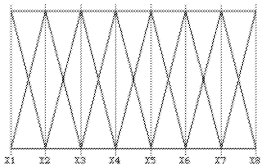
2D (square)



3D cube: 8 vertices



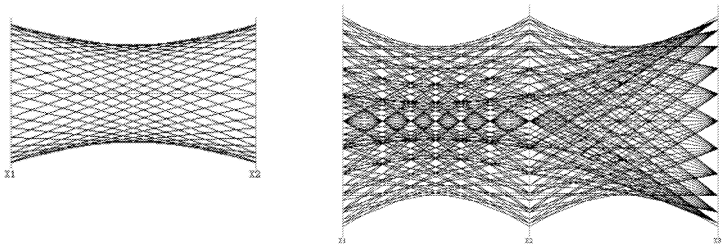
8D: 256 vertices



|| spheres

Hypercubes in parallel representation:

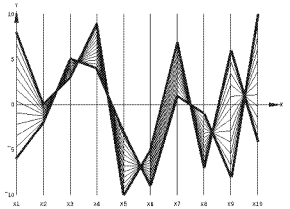
2D (circle) 3D (sphere) ... 8D: ???

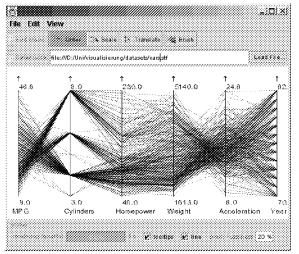


Try some other geometrical figures and see what patterns are created.

|| coordintates

Representation of 10-dim. line ($x_{1,}$,... x_{10}) t, car information data





Parallax software: http://www.kdnuggets.com/software/parallax/
IBM Visualization Data Explorer http://www.research.ibm.com/dx/
has Parallel Coordinates module
http://www.cs.wpi.edu/Research/DataExplorer/contrib/parcoord/
Financial analysis example.

More tools

Statgraphics charting tools

Modeling and Decision Support Tools collected at the University of Cambridge (UK) are at:

http://www.ifm.eng.cam.ac.uk/dstools/