InVis is a tool that enables a user to interactively explore a dataset. At the moment there is only one algorithm implemented for the interactive analysis, “Least Squared Error Projections” (LSP). Unfortunately the constrained Kernel PCA version is not fully build in yet. However, there is already an option visible in the GUI to select c-KPCA (it just doesn’t do anything).

Loading and storing datasets

In general, you can load csv, arff and libsvm data-files. Note, that for loading csv files, the parser is quite strict. The file needs a header-line and every row is interpreted as a data-record; with its name as first and its labels as last entry. The values have to be separated by commas and only the name entry is allowed to be non-numeric. Also the entries may not be quoted (e.g. ”12.54”). You can store and load a running analysis session and export the current projection matrix, including the parameters like e.g. the indices and locations of the selected control-points.

<table>
<thead>
<tr>
<th>Good csv</th>
<th>Bad csv</th>
</tr>
</thead>
<tbody>
<tr>
<td>name,a1,a2,a3,label</td>
<td>#header</td>
</tr>
<tr>
<td>Dino,1.0,2.3,2.1,100</td>
<td>”Dino”;”1.0”;”2.3”;”2.1”;”A”</td>
</tr>
<tr>
<td>Dani,3.0,1.3,2.3,80</td>
<td>”Dani”;”3.0”;”1.3”;”2.3”;”B”</td>
</tr>
</tbody>
</table>

Figure 1: Load datasets/sessions, save sessions and export the projection matrix.
Classic embeddings

The option ”Classic embeddings” lets you have a look at your data, using commonly used embedding techniques:

- **PCA** (principal component analysis)
- **LLE** (locally linear embedding)
- **Isomap** (Isomap: basically, multidimensional scaling applied to the knn graph)
- **ICA** (independent component analysis)
- **Back to LSP** (lets you get back to the interactive surface)

While you are in a static embedding, you can see all your control-points, but you can not toggle their selection, or relocate them!

![Classic embeddings screenshot](image.png)

Figure 2: *Check out some classic embedding methods.*

Mouse interaction:

- **Left-click & drag** lets you drag the control-point nearest to your click around.
- **Right-click** displays infos of the clicked data-record (e.g: attribute_name:values;0)
- **Middle-click** lets you select or de-select a data-record as control-point.
- **Mouse-wheel** lets you zoom in and out.
- **x+left-click** lets you construct a cocktail (this is a bit experimental). At the moment the cocktail is constructed by the ”most probable” method, using only ingredient from the 5 nearest neighbours in the embedding.
Search

You can search for a substring that is contained in the data-records name (bottom right in the GUI). The matching results will be highlighted in red. To clear the canvas from the highlights (and also info-boxes) you can use the "Clear" button or alternatively press "ctrl+c".

Editing attributes

Important: All manipulations on attributes, including the label, have to be applied by using the "Update" button (or pressing ctrl+u). You can ignore an attribute in the calculation of the embedding by un-checking it (in the right side of the GUI). You can also fade-out data-records that do not meet a constraint on an attribute. To formulate a constraint, double-click the attribute; you can now edit the name. Note that this is ONLY to add constraints in a specific syntax.
changing the attribute name to a more suitable name will break the program!
The constraint syntax is quite easy, attribute-name, constraint and value are
specified in a colon separated manner, e.g.: vodka:>:0 or orange juice:=:0.5
The only constraints available at the moment are <, > and =. In the example
below all cocktails that do not contain vodka and orange juice are faded out.

Figure 5: *Fade-out data-records that do not satisfy a constraint on a set of attributes.*

You can also change the label that is used to color the embedding, simply
change the attribute in the "Label" field to the desired attribute and click on
"Update". In the example below you can see how the data-records that are
closer to the manchurian candidate (the most vodka-containing cocktail in the
database) also have a more intense blue color.

Figure 6: *Change the labels that is used for the coloring.*