

KLIMT

Klassification - Interactive Methods for Trees

Quickstart Guide

Simon Urbanek

Requirements

This is a short guide to show some basic features of KLIMT step by step. For more in-depth discussion of features consult the reference documentation instead. In order to follow the steps exactly as described in this document your system needs to satisfy the the following requirements: Your system must be a PC (Intel compatible) with Windows 95 or higher. Microsoft's Java Virtual Machine must be installed (this is the case if you have some recent Internet Explorer installed). You have also obtained the Windows version of KLIMT, which should consist of a file named `klimt.exe`. If you plan to use the R interface you should also download the `klimt.r` file and put both files in the R-directory¹ which is for example `C:\Programs\R\rw1031` as default for R v1.3.1. You must also have the *tree* package installed (can be obtained from CRAN).

It is possible to follow the steps below with different configurations (such as unix platforms or Mac OS) with only minor changes.

Starting KLIMT from R

In the following example we will use the R's *iris* dataset and generate a tree with default parameters. If you don't have R or don't want to generate own trees, skip this section, download the prepared `iris.data` file, double-click on the KLIMT application and select the supplied `iris.data` file which contains both the *iris* dataset and a tree.

If you want to use R, start it and use following commands to generate a tree and launch KLIMT:

```
> source("klimt.r")
> data(iris)
> t <- tree(Species~., iris)
> Klimt(t, iris)
```

The file `klimt.r` defines the `Klimt(tree, dataset)` function and it also loads the *tree* library automatically. After fetching the preinstalled *iris* dataset the actual tree is built. The model definition `Species~.` causes *Species* to be explained using all other variables. You could preview the generated tree in R with `plot(t); text(t)`. Finally KLIMT is started with the tree *t* and dataset *iris*.

¹The reason is to have **Klimt.exe** in the working directory. Alternatively it is possible to place KLIMT in its own directory and add that one to your PATH environment.

Working with KLIMIT

After KLIMIT loaded the dataset two windows appear: Variables window and tree window. In the tree window you can move nodes by dragging them. If you hold the <Shift> key, movement is constrained to horizontal direction. Holding <Ctrl> allows you to move individual nodes instead of branches. Use <r>² key to re-arrange nodes the way KLIMIT proposes. Try following different keys and watch their influence on the tree display: <Shift><r>, <c>, <ℓ>, <f> and <s>. The later switches between proportional view, where the size of a node is proportional to the number of contained cases, and the fixed view, where each node has the same size.

You can query an individual node by right-clicking on a node³. Extended query containing additional information can be obtained by additionally holding the <Shift> key.

KLIMIT provides a way to visualize the deviance gain of each split and the final deviance of a node. Press <d> to include this information in the plot. Red circles appear beside each inner node representing the deviance gain in that node. Red rectangles beside leaves correspond to the deviance in the leaf, thus pure leaves have no symbol. For alternative visualization press <Shift><d> to obtain a deviance plot. Select one of the leaves to see how the deviance changes on the way down to the split. In this plot use <ℓ> to switch between bar and line representation or <c> to switch between cumulative view (i.e. absolute deviance plotted) and deviance gain view.

Tools

In the tree window there are four different tool modes, which can be seen in the *Tools* menu. In the default mode *select cases* clicking on a node selects all cases in the node. Selected cases are denoted by green color in all plots. The proportion of the green area relative to total area of the node corresponds to the proportion of the selected nodes. Holding <Shift> key during the selection causes the selection of the cases to be inverted in the corresponding node and does not affect cases in all other nodes. If no other keys are pressed a new selection replaces the previous one. If you hold the <Ctrl> key, newly selected cases are added to the previous selection. Both selection methods and graphical methods of highlighting apply to all other plots, such as histograms, barcharts or treemaps.

Besides the *select cases* tool three other modes can be used. Pressing <n> switches to *node select* mode. Case selection is not changed in this mode, instead clicking on nodes only causes the node itself to be selected. Select various nodes and note the different coloring used for the nodes. Dark blue color indicates the currently selected node, inner leaves of the same classification as the current node are painted yellow and leaves of the same class are light blue. In addition names of the same class are drawn in red. The node selection is important especially when tree enhanced plots are used such as scatterplots or histograms.

²All commands described here as key strokes can be also issued from the menu. You can see the list of all available key commands by selecting *Help* → *Shortcuts*.

³On some other platforms with single mouse button this corresponds to holding the <Alt> key and clicking on the node.

Last two available tools are *zoom* (<z> key) and *move* (<v> key). In the zoom mode clicking inside the tree performs zoom in by factor two. To zoom out hold the <Shift> key pressed while clicking. KLIMT supports logical zoom, try to zoom out multiple times to observe the effect. In the *move* mode, dragging moves the entire canvas in the direction of the mouse movement. The *move* mode can also be switched on temporarily at any time by holding the <Space> key as known from Adobe products.

Entire branches of the tree can be pruned by selecting the last node to be retained and pressing the <p> key. Pruned branches can be expanded back to full size by clicking on the plus sign denoting a pruned branch. Once the tree is pruned to your satisfaction you can generate a new, pruned tree by pressing <Shift><n>. Both, the previous and the new tree, are linked via highlighting.

Plots

Beside hierarchical tree views KLIMT supports many different plots. Alternative view called *treemap* can be obtained by pressing the <m> key. The space is partitioned according to splits in each node. Each partition corresponds to one leaf as you can easily verify by clicking inside the partitions and hence selecting the corresponding cases in all plots, including the tree. You can toggle the orientation of the treemap by pressing <Shift><r>.

Pressing <a> in the treemap window toggles between a treemap and spineplot of leaves. The difference is that treemap partitions space horizontally and vertically in alternate fashion, whereas spineplot is made by partitioning in one direction only. Both plots are helpful in comparing proportions.

To highlight cases of a specific class we can use a barchart. Go to the variable window, click on the *Species* variable and further on the *Hist/Barchart* button. You will obtain a barchart of the *Species* variable. Click on each bar to select cases of the category and watch the corresponding distribution in the tree and the treemap. You can change the order of the categories by dragging the bars.

Click on the *Petal.Length* variable and then on *Hist/Barchart* to obtain a histogram. You can change the bin width by dragging the right tick mark on the *x*-Axis. Anchor point can be changed by dragging the left tick mark. If you select a node which was obtained by splitting on the *Petal.Length* variable, the corresponding split will be shown in the histogram. Go to the tree window and select one of the children of the root. You will see a red line denoting the split appear in the histogram.

Even more information about the partitioning can be seen in scatterplots. Leave *Petal.Length* selected in the variables window, hold the <Shift> key and click on *Petal.Width* to have both variables selected. Now click on *Scatterplot*. If you still have a node selected, you will see a partitioned scatterplot. It is a two-dimensional projection of the splits. You can select cases in a scatterplot by either clicking on a case or by dragging a selection rectangle. Axes can be swapped by pressing <Shift><r> (rotate), labels can be toggled by the <l> key and alternative visualization of the selected points can be activated with the <e> key.

Finally let us look at boxplots. Select just the *Petal.Length* variable in the variable window and click on the *Boxplot* button to get a single boxplot. Select additionally the *Species* variable and click on the *Boxplot* button. Now you see multiple boxplots of *Petal.Length* side-by-side grouped by the categorical

variable *Species*. Boxplots are the only plots where you can't select cases. Click instead on *Scatterplot* to obtain a parallel dotplot, which is technically a scatterplot with equidistant axis by category for the categorical variable.

Common functionality

All plots support exporting functions allowing you to use all the graphics in other programs. Press <Shift><p> to produce an EPS file or <Shift><x> for PoGraSS meta file (see documentation for details about PoGraSS). The resulting graphics is vector-based and can be loaded directly by many programs such as TeX or Adobe Illustrator. For Windows users there is also a PoGraSS meta to WMF/EMF convertor available so you can use the graphics in Microsoft products such as Word, Excel etc.

Most plots also support export of selected cases in an ASCII, tab separated text file. Use <Shift><c> to export the current selection.

All plots have a menu item called *Window* which can be used to quickly switch between windows or to close the active window. Use *File* → *Quit* to close KLIMT.

KLIMT can be used to open any dataset even without a tree, as long as it is in space separated ASCII format. KLIMT can handle missing values, it currently supports R's NA notation, question mark or Apple's dot symbol. You can load multiple trees in KLIMT by using the *Open tree..* button. Each tree must be in a separate file generated by „sink“ing the `print(tree(...))` command.

You cannot work with the same instance of R while KLIMT is running. Use the bi-directional R interface as described in the appendix of the documentation if you need cooperation between KLIMT and R both ways.

KLIMT is still work in progress, I am thankful for any suggestions or bug reports. Some fetatures may have changed in the meantime, this guide is based on KLIMT version 0.95g.

Simon.Urbanek@math.uni-augsburg.de