Multi-class data exploration using space transformed visualization plots

Rida E. Moustafa[1], **Ali S. Hadi**[2], **Jürgen Symanzik**[3]

SUMMARY

Visualization of large data sets is computationally expensive. For this reason, enveloping methods have been used to visualize such data sets. Using enveloping methods, we visualize summary statistics of the data in the space transformed visualization (STV) plots, such as the traditional parallel coordinate plot (TPCP), instead of the actual data records. Existing enveloping methods, however, are limited only to the TPCP and they can also be misleading. This is because the parallel coordinates are parameter transformations and the summary statistics computed for the original data records are not preserved throughout the transformation to the parallel coordinates space. We propose enveloping methods that avoid this drawback and that can be applied not only to the TPCP but also to a family of STV plots such as the smooth parallel coordinate plot (SPCP) and the Andrews plot. We apply the proposed methods to the min-max, the quartiles, and the concentration interval envelopes (CINES). These enveloping methods allow us to visually describe the geometry of given classes without the need of visualizing each single data record. These methods are effective for visualizing large data sets, as illustrated for real data sets, because they mitigate the cluttering effect in visualizing large-sized classes in the STV plots. Supplemental materials, including R-code, are available online to enable readers to reproduce the graphs in this paper and/or apply the proposed methods to their own data [1].

References

[1] Rida E. Moustafa, Ali S. Hadi, Jürgen Symanzik (2011). Multi-Class Data Exploration Using Space Transformed Visualization Plots. *Journal of Computational and Graphical Statistics* **20**(2), 298–315.

[1]dMining Technology, LLC, Ashburn, Virginia, USA
`moustafa@dmining-technology.com`

[2]The American University in Cairo, New Cairo, Egypt
`ahadi@aucegypt.edu`

[3]Department of Mathematics and Statistics,
Utah State University, Utah, USA
`symanzik@math.usu.edu`