Using Data Images for Outlier Detection

David Marchette
Naval Surface Warfare Center

The data image maps high dimensional data into an image. This can be done directly, resulting in an image of size nXd, where n is the number of observations and d is the dimension. Alternatively, one can display the interpoint distance matrix as an nXn image. Using clustering techniques, we show that outliers, including clusters of outliers, have a tendency to stand out quite well in the data image, making this a potentially useful tool for exploratory data analysis.