Extending Depth Explorer and DEJ Depth Measures

Ashish Datta  Robby Ramdin  Advisor: Diane Souvaine
Tufts University Computer Science

Background
Computational geometry is the branch of computer science that deals primarily with problems framed in geometric contexts. Examples of computational geometry problems include calculating convex hulls and triangulating points. Solutions to these problems have several practical applications, including the efficient geographical distribution of resources.

What is data depth?
Data depth provides a center-outward ordering given a set of points. This allows us to decide how close a point is to the center of the set. Data depth is useful because it provides a fundamentally different analysis than classic statistic tools such as medians, averages, and standard deviations.

About Depth Explorer
Depth explorer (DEJ) is written entirely in Java and based on the tuftsgeometry library. DEJ allows users to specify data and depth measures in depthML, a XML specification, and outputs images of the depth calculations. Since DEJ is written in Java, all of our work has been in Java in order to simplify integration.

Animation in Depth Explorer
The animation extension to DEJ allows the user to interpolate depthML elements between key frames. Because of DEJ's modular design, all of the depthML elements animate themselves which keeps the system extensible. Additionally, the depthML specific code from the animation extension was separated to create a generic XML interpolator.

Generic XML Interpolation
The generic XML interpolator allows a user to interpolate any XML specification file with the appropriate connector. Currently, a complete connector for depthML and a partial connector for SVG exist.

Proximity Depth
Proximity depth is a novel approach to data depth. It works by creating a proximity graph, that is a graph where points that are close enough to each other by some measure have an edge between them; then it determines the shortest path through the unweighted graph from every point to the convex hull.

Why data depth?
Data depth separates itself from classical data analysis because it is non-parametric. Because of this using data depth doesn’t require knowledge about a dataset’s underlying distribution.

Improving Proximity Depth
The shape of the data should not matter, but it does. Some naturally occurring distributions that cause less than ideal behavior are banana shaped clouds, and multimodal distributions.