Visualizing high dimensional data: Applying graph theory to data visualization

R.Wayne Oldford
University of Waterloo
Canada

Abstract

In statistical data analysis, we are often looking for structure in high dimensional data. In three or fewer dimensions, our visual system is an important asset, as much (even unanticipated) structure can be recognized effortlessly when points can be plotted so few dimensions. Unfortunately, modern digital data sources (e.g. images, text documents, sound files, ...) often have 100s or 1000s of dimensions. Even after formal dimension reduction methods have been applied, we are often faced with many more dimensions than three.

In this talk, I will explore some visualization methods for high dimensional data. I will review and illustrate methods based on radial, parallel, and orthogonal coordinates. These three axis systems have different strengths and weaknesses. In all cases however, improvements may be had by casting the axis arrangement in a graph theoretic framework. I will explore the relevant graph theoretic representations and illustrate their use on real data sets (e.g. Frey image data). I will pay particular attention to the orthogonal axis system and show how graph traversal can be used to meaningfully navigate through high dimensional space.

All software used is available from `R` packages `PairViz` and `RnavGraph`. This talk is based on joint work with Catherine Hurley, NUI Maynooth (Ireland) and Adrian Waddell, Waterloo (Canada).