

R Shiny's Self-Organizing Map

Zury Betzab Marroquin¹, Joshua Walsh², Trenton Wesley³

1 - Scripps College, Claremont, CA

2 - University of Alaska Southeast, Juneau, AK

3 – Harvey Mudd College, Claremont, CA

Abstract

There exist oceans of data, and into these oceans flow wide rivers of more data from the various disciplines in the applied sciences. The majority of these data are high-dimensional, and relationships among the underlying variables are unknown. The initial task of researchers who explore such data typically involves using technology or advanced software to unveil, or at least better understand the nature of those unknown relationships. Preparing suitable data analysis code can be time consuming and advanced data analysis software can be expensive, and it is to circumvent these two hurdles faced by data analysts that this group prepared an R Shiny app. This app takes in high-dimensional data and identifies clusters that may exist in the data. This is done using an algorithm that produces a two-dimensional discrete map of the dataset, by running a self-organizing map (SOM) on the dataset, whereby the results are then displayed graphically. In particular, the SOM function used is a two-layered unsupervised neural network with competitive and cooperative learning.