Statistical Connectomics: A posed statistical decision theoretic for brain graph clustering

Greg Kiar

February 15, 2015

Overview

A relevant task in brain graph classification is the clustering and recognition of similar cliques within the graph. Here, we pose a statistical decision theoretic to aid in the development and evaluation of such clustering techniques. As we know, a statistical decision theoretic must define the following: a sample space, model, action space, decision rule class, loss function, and risk function. Shown below are these elements of the theoretic designed.

Posed Statistical Decision Theoretic

Sample Space $\mathcal{G}_n = (V, E, Y)$
The sample space for clustering is the space of graphs with fixed vertices, edges between them, and a label indicating one of two categories.

Model $P = SBM_n^k(\rho, \beta)$, where, for $k = 2$: $\rho \in \Delta_2$, $\beta \in (0, 1)^{2x2}$
The model used to simulate data within this space is the stochastic block model with two distributions.

Action Space $A = \{y \in \{0, 1\}^n\}$ The action space consists of cluster assignments made by the decision rule class.

Decision Rule Class The clustering algorithm of choice can be stated and implemented here.

Loss Function $l : \mathcal{G}_n \times A \to R_+$, $l = \sum_{i=1}^{n} \Theta(\hat{y}_i = y_i)$
Loss in this statistic can be measured by observing incorrect assignment of class. The adjusted rand index (ARI) is a good measure to do this, as it allows you to compute a normallized to $\{0, 1\}$ loss dependent on the number of correcty and incorectly assigned labels.

Risk Function $R = P \times l$, e.g. $R = E_P(l)$
The risk function can, in the simple case, be the expected loss of the function.