Statistical Connectomics HW 2

David Lee

February 17, 2015

A statistical decision problem we discussed in class can be divided into 6 parts - the sample space, the model, action space, decision rule class, loss function and risk function.

Sample space The sample space consists all possible arrangements of edges, nodes, and labels.

$$G_n = (V, E, Y) \tag{1}$$

Model A stochastic block model consisting of

$$P = SBM_n^k(\rho, \beta) \text{ where } \rho \in \Delta_2 \text{ and } \beta \in (0,1)^{2\times 2} \tag{2}$$

$$\text{when } k = 2$$

Action space The action space is the assignment given by the clustering algorithm.

$$A = \{y \in \{0,1\}^n\} \tag{3}$$

Decision rule class The decision rule class is given by a method of clustering (such as k-means in our case discussed in class)

Loss function The loss function can be given by the adjusted rand index and gives the cost associated with an action

$$l : G_n \times A \to R_+ \tag{4}$$

$$l = \sum_{i=1}^{n} \Theta(\hat{y}_i = y_i)$$

Risk function The risk function is given by the following equation

$$R = P \times l \tag{5}$$

which can be redefined as the expected value of the loss function:

$$R = E\{l\} \tag{6}$$