Michael Seung
2/17/2015

Statistical Connectomics Homework #2

The statistical decision theoretical problem statement for clustering vertices of a graph can be described by the following 5 components:

[1] Sample space – Collection of all possible observable graphs (directed, simple, attributes)

$$G_n = (V,E,Y)$$

[2] Model – Stochastic Block Model defining a probability distribution over networks

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

[3] Action Space – Space of possible actions with the given data or cluster assignments

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

[4] Loss function – Function quantifying a loss during parameter estimation

$$L: G_n \times A \rightarrow R+$$

[5] Risk function – Expected value of the loss function with respect to the sample space, specifically the Stochastic Block Model

$$R: L \times \emptyset \times P \rightarrow R+$$