

STATISTICAL CONNECTOMICS: A POSED STATISTICAL DECISION THEORETIC FOR BRAIN GRAPH CLUSTERING IN THE SUPERIOR COLLICULUS

ERIKA DUNN-WEISS

1. OVERVIEW

The superior colliculus (SC) both receives sensory inputs and sends outputs to motor command centers, making it a key center for sensorimotor integration. In the big brown bat, it's been shown (Moss 1997) that the SC is involved in head aiming and call production for goal-oriented echolocation. Given a directed graph of activity in the SC, where the nodes are different neurons in different layers of the SC and edges are synapses from one neuron to another, we propose a statistical decision theoretic for clustering the graph into sensory-input neurons (0) and motor-output neurons (1).

2. POSED STATISTICAL DECISION THEORETIC

2.1. Sample Space. $\Omega\mathcal{G}_n = (V, E, Y)$, the space of graphs with fixed vertices, edges between them, and a label indicating one of two categories (sensory or motor).

2.2. Model. $P = SBM_n^k(\rho, \beta)$ for $k = 3$ for the superficial, intermediate, and deep layers of the superior colliculus. Thus the model used to simulate this data within this space is the stochastic block model with three distributions. $\rho \in \Delta_8$ and $\beta \in (0,1)^{3 \times 3}$.

2.3. Action Space. $A = \{y \in \{0,1\}^n\}$.

The action space consists of the cluster assignments to either sensory (0) or motor (1) clusters.

2.4. Decision Rule Class. $\delta : \Omega \rightarrow A$

Since it's important to specify the number of clusters beforehand, and since we expect anatomical organization of these neurons based on functionality (i.e. clusters shouldn't largely overlap or have large variance, which is an issue for centroid-based clustering), the choice for the clustering algorithm will be k -means.*

2.5. Loss Function. $l : \Theta \times A \rightarrow \mathbb{R}_{>0}$ where $l(\theta, \delta(\Omega))$ is the cost of action δ under parameter θ . More explicitly, $l(\theta, \delta(\Omega)) = \sum_{i=1}^n \Theta\{\hat{y}_i = y_i\}$, where y_i is the cluster assignment of vertex v_i and \hat{y}_i is the correct assignment of vertex v_i . Here we use the adjusted rand index (ARI).**

2.6. Risk Function. $R(\theta, \delta) = E_p(L(\theta, \delta(\Omega)))$. The risk function is taken as the expected loss.

Date: February 16, 2015.

3. DISCUSSION

1. *Minimally overlapping and small variance relative to the mean are assumptions I would rather not have to make a priori. I would also prefer to be able to leave some unclustered that may be implicated in sensory and motor activities. But I don't know of any other clustering algorithm that allows you to specify the number of clusters up front that does not have these issues. Does anyone have another recommendation?

2.**I'm not certain if stating the loss function in this way is still appropriate given that the \hat{y} 's are unknown.

3. Would a latent position model be preferable to the SBM in this case? I thought perhaps that would help with point 1. But, latent position models are not as clear to me, so I wasn't sure how to state one.

Thanks!