Russell Ricker: Mathematical Sciences
Mentor: Albert Parker, Tomas Gedeon -- Mathematical Sciences

Creating Soft Clusterings of Data Via the Information Bottleneck Method

Information-based distortion methods have been used successfully to analyze the relationship between stimulus and reaction spaces. Distortion methods make few assumptions concerning the correspondence between the two spaces, providing maximally informative relationships between them. I used the Information Bottleneck technique to create soft clustering of a synthetic data set with 50 stimuli and 50 neural responses with a multivariate Gaussian (either with 4-blobs or 10-blobs) describing their hypothetical relationship. The algorithm utilized an annealing method to solve the high-dimensional non-linear problem, and was implemented using Matlab. As the annealing parameter increased, the solution to the problem underwent a series of phase transitions, or bifurcations, that eventually stabilized to a nearly deterministic clustering. By calculating the matrix of second derivatives (Hessian), we are able to determine when the bifurcations occur. By calculating the third and fourth derivatives we are able to determine whether the bifurcations are subcritical or supercritical. The existence of subcritical branching implies that several solutions not found by the method of annealing exist. Because the method of annealing is guaranteed to converge, the subcritical branch must turn at a later bifurcation and become optimal.