**Semi-Supervised Learning Using Bayesian Hierarchical Methods**

Victor Chen
*Mentors: Andrew M. Stuart and Matthew M. Dunlop*

In semi-supervised machine learning, the task of clustering is to divide the data into groups using a labeled subset. Our Bayesian approach to graph-based clustering views the classifying function as a random variable with a distribution that combines the label model with prior beliefs about the classification. In Bayesian hierarchical methods, hyperparameters governing the prior distribution are introduced and can be sampled as well, with the goals of both deriving a classification and learning the distribution of the hyperparameters. We apply Markov Chain Monte Carlo methods for indirectly sampling the posterior distribution of these random variables, as direct sampling is generally challenging. We focus on priors derived from the graph Laplacian, a matrix whose eigenvectors are known to contain cluster information. We implemented Bayesian hierarchical models that learn different sets of hyperparameters, including ones that govern the scale of the eigenvectors or the number of eigenvectors used. We tested these models on real and synthetic data sets. Our results indicate that there is information to be learned about the distribution of these hyperparameters, which could be used to improve classification accuracy.