STAT 771 Project Summary

Stewart Kerr (Professors Menggang Yu & Guanhua Chen, BMI Department)

Mixture proportion estimation (MPE) is the problem of estimating the weight of a component distribution in a mixture given samples from the mixture and component. Solving this problem is an important step in “weakly supervised” learning tasks in which one has access to only positively labelled data or there is noise in the labels in the training set. Specifically, we are interested in MPE to assess the similarity between two different populations of hospital patients. For example, we wish to use data observed from patients in Iowa to build upon models learned using data from Wisconsin patients. That is, Iowa is defined as our **source** population while Wisconsin is our **target** population. Among several methods proposed to solve MPE problems, we adapt methods proposed by Ramaswamy et. al via kernel embedding of distributions. This approach utilizes an efficient algorithm for MPE along with guaranteed convergence to estimate the true proportion under certain less restrictive conditions. Using this method, we can obtain an estimate, of the proportion of the source population