Stat635 Project Proposal

Arthur Lui

11 March 2015

In the plot below, we see that there are three groups or clusters of observations. One could fit one linear model

$$y = \beta_0 + \beta_1 x + \epsilon$$

through all three clusters together (the black line); but it would be more appropriate to fit three seperate models to each cluster (the colored lines). Using linear mixed models (LMM), we could fit the model

$$y_i = \beta_0 + \beta_1 x_i + Z_i \gamma + \epsilon_i$$

to account for the effect of each cluster. The problem with an LMM is that if the clusters are latent, and not observed, we cannot specify out design matrix.

Using an Indian Buffet Prior, a distribution for sparse and infinite binary matrices, I propose a model and sampling algorithm to estimate Z. Time permitting, I will fit the model.