Regularized logistic clustering for binary data[a]

Hosik Choi, Kyonggi University
Seokho Lee, Hankuk University of Foreign Studies

Summary: Recently, many binary data have been produced. To cluster such data, we require some adequate metric on binary vector unlike continuous data. But, it is difficult to find good 'metric' to clustering. As clustering's objective, we consider binomial likelihood as a natural distribution for binary values and equal constraints on latent values of those probabilities. The proposed clustering method is defined on the logit transform of the success probabilities of the binary observations and pairwise fused penalty between principal scores of binary vectors. Further we propose the optimization algorithm to solve it where the algorithm consists of two routines. Outer routine is a Majorization-Minimization(MM) algorithm and inner routine consists of two steps. First is to find sparse principal loading vector for enhanced interpretability and more stable extraction of the principal components. Second step is to find clusters using splitting method like alternative direction multipliers method(ADMM) to deal with equal constraints of principal scores. Numerical studies show that the proposed method works quite well comparative to the competing methods.

Keywords: shrinkage, pairwise fused penalty, dimension reduction

[a]This work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2013R1A1A2007611)