REVIEW OF ICL CRITERION FOR CLUSTERING IN MIXTURE MODELS

Nusrat Jahan

**Department of Mathematics and Statistics
Mississippi State University
Mississippi State, MS 39762 USA
email:njahan@ra.msstate.edu**

ABSTRACT

Statistical analysis of mixture models are of interest because it is an alternative to nonparametric density estimation and it is a powerful way of modelling in cluster analysis. In case of density estimation, optimization of the Bayesian Information criterion (BIC) generally results in a good approximation of the density to be estimated. But in case of cluster analysis, the BIC tends to overestimate the number of clusters when the data is a poor fit to the mixture model. In this context a modification of BIC, integrated completed likelihood (ICL) criterion has been investigated. In the ICL approach, the integrated completed likelihood is maximized to select both a relevant form of model and relevant number of clusters. In the BIC approach, only the observed likelihood is maximized. Where as the integrated completed likelihood includes the estimated (using maximum a posteriori function) missing data. The ICL criterion penalizes for the complexity of the mixture model, thus ensuring the partitioning of data with the greatest evidence. This paper will focus on the computation of ICL, effectiveness and drawbacks of ICL in the context of cluster analysis. The differences between ICL and BIC will also be investigated.