Data Mining Principles

Assignment 3 Part 1

Use the GermanCredit data (package caret) in R or from UCI Machine Learning's website.

Latent Class Analysis Homework

1. Perform latent class analysis of only the categorical variables for market segmentation using (function poLCA in package poLCA). Remember: the local optima problem is big for all the clustering and latent class methodsThe data for analysis should only include the variables that you think have business relevance for market segmentation.
2. Determine 2, 3,..,K class/cluster solutions. Remember to run from multiple random starts. Use AIC criterion and interpretation based on graphs to interpret LCA solutions.
3. Perform Holdout validation of LCA.
   a. For Holdout, use the centers class-conditional probabilities - probs - from training set as input to probs.start for holdout (generated from the training set LCA solution, as the starting point for holdout. Use similarity of relative class sizes and hold-out class conditional probabilities as measures of stability.
4. Provide implications / commentary on the goodness, interpretability, stability, and adequacy of solutions.
5. Comment on the similarity/differences between the clustering solutions you generated in Assignment 1 with the solution you generated using LCA.

Each of the (5) parts is worth 1 point each. TOTAL: 5 points.