

Assignment 5: Part 1

Use the GermanCredit data (package caret) in R

Cluster-wise regression

1. Use the training and Test samples for the GermanCredit data set that you used in Assignment 1. Call them, say, Train and Test.
2. Use the Train data set to build a Cluster-wise Regression Model. Choose "Amount" as the dependent variable. Build 1, 2, and 3 cluster solutions. 4 clusters may be too many for this data set.
 - a. Use the clustreg() function that I uploaded on Canvas. Don't use the "Class" variable as independent variable.
 - b. You don't have to use categorical predictors. You can use only the numeric independent variables 1, and 3 through 7 as predictors. (NOTE: But if decide to use categorical predictors, make sure you select only (K-1) dummies for the categorical variable – not all K dummies. Also, during some of the cluster-wise regression runs, some regressions may give error and the cluster-wise-regressions procedure may error out. This happens because at some stage in the iterations, the categorical dummies may only have all 1s or 0s, for some of the clusters, and regressions will have problems in that case. It might mean (a) you have to use a different starting random number seed, or (b) that you are trying to extract too many clusters for that data – or (c) that you might have to drop some of the categorical variables).
 - c. Plot R^2 as a function of the number of clusters
3. Perform Test validation testing of the cluster-wise regressions using function clustreg.predict() that I have uploaded in R
4. Choose a model with the best regression interpretation on Training Data, R^2 and related significance, and the best Test performance
5. Summarize your results – for both Training and Test

TOTAL POINTS: 5