



Assignment 1

MUTATE: Multiple Training and TEst

- Use GermanCredit data in package caret
- Build a regression model to predict variable "Amount" as a function of other variables (choose variables that you think are necessary and required to build a good model), using following methodology:
 1. Split sample randomly into training-test using a 632:368 ratio.
 2. Build the model using the 63.2% training data and compute R-square in holdout data. (function `lm()` gives R-squares)
 3. Save the coefficients, R-square in training and holdout samples. (To compute R-square in Holdout, take the square of correlation between actual and predicted values)
- Repeat steps 1-3 1000 times. Save all 1000 results.
- Plot the distributions of all coefficients, holdout R^2 , and % fall in R^2 .
- Compute the averages of all 1000 coefficients.
- Compute the standard deviation of all 1000 coefficients (for each beta)
- Compare average across 1000 to single model built using entire sample.
- Sort each coefficient's 1000 values. Compute 2.5%-97.5% Confidence Intervals (CI). Scale these CI's down by a factor of $.632^{0.5}$. How do these CIs compare to CIs computed from single model's CIs? Tighter or broader?
- Summarize your results.