Data Mining Principles

Assignment 3 Part 2

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

Principal Components Homework

1. Split sample into two random samples of sizes 70% and 30%.
2. Perform principal components of variables 1:7 from the GermanCredit Data on training sample.
3. Generate Scree Plots and select number of components you would retain.
4. Plot Component 1 loadings (x-axis) versus Component 2 loadings (y-axis). Use this plot to interpret and name the Components. Repeat this by plotting Component (1) separately versus all components you decided to retain from Step 3 (Component 3, Component 4 etc).Make Sure to interpret each of the components you decide to retain. In case a component is not interpretable, note that.
5. Show that Component loadings are orthogonal.
6. Show that Component scores are orthogonal.
7. Perform Holdout validation of Principal Components solution.
    a. For Holdout validation, you will have to
        i. predict the component scores in the Holdout using the predict() function and
        ii. matrix multiply the predicted component scores from (i) above with transpose of component loadings you derived from training data set from Step 2 above. Refer to Page 52 of Class Lecture for Session 4 for details. Understand the R commands on that slide.
8. Compute the Variance Account For (R^2) in the Holdout sample. That yields a measure of Holdout performance.
9. Rotate the component loadings using varimax rotation. Use R function varimax() for it. Look at the Loadings from the varimax rotation. Does it yield any different Interpretation of the Principal Components?
10. Plot rotated loadings(1) versus rotated loadings (2) and (3). Do you think Principal Components reduced this data a lot? Do you like the solution?

Points for each of the 10 parts : 0.5

TOTAL POINTS : 5