Assignment 3 Part 2

Please use the Boston Housing data for this assignment.

Principal Components Homework

1. Split sample into two random samples of sizes 70% and 30%.
2. Perform principal components of numeric variables from the Boston Housing 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). Can you interpret each of the components you decide to retain. In case a component is not interpretable, note that.
5. Perform the following:

    a. Show that Component loadings are orthogonal.
    b. Show that Component scores are orthogonal.
    c. Perform Test validation of Principal Components solution.
        i. For Test validation, you will have to
            1. predict the component scores in the Test [using the predict() function in R and transform function in Python
            2. matrix multiply the predicted component scores from (1) 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.
    d. Compute the Variance Account For (R^2) in the Test sample. That yields a measure of Test performance.
    e. [OPTIONAL] Rotate the component loadings using varimax rotation. Look at the Loadings from the varimax rotation. Does it yield any different Interpretation of the Principal Components? **Python Users: Current package is a month old, and buggy.**
    f. [OPTIONAL] 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?

TOTAL POINTS : 5