Master BeNeFri in Computer Science

Course: Statistical Learning Methods with R
Spring 2022

Exercise #10. Principal Component Analysis

Download from the ILIAS website the `Boston` dataset (filename: Boston.txt).
This dataset contains information collected by the U.S. Census Service concerning housing in the area of Boston (Massachusetts).
Further information is provided in the file Boston.pdf.

1. Normalize your dataset and consider all the variables except `MEDV`. Create a PCA model and plot it.
2. Which predictor variable contributes the most to component 1? And which contributes the least?
3. Estimate the proportion of variance explained by all the components. If we want to explain only 80% of the original data, how many components should we use?
4. Generate a new dataset using only the components selected in Problem 3. Create a multiple regression model using these components as predictors for the target variable `MEDV`.
5. Compare the model created in Problem 4 with a multiple regression model using all the components.