

Machine Learning Exercise - I

April 27, 2022

Kannan Singaravelu

* * *

Exercise 1

For the Boston housing dataset, construct a dataframe with all features and target values. Perform feature selection to choose the most appropriate features using

1. Variance Inflation Factor
2. SelectKBest
3. Recursive Feature Elimination
4. Recursive Feature Elimination Cross validation
5. Shapley Additive Explanations

Fit the regressor and compare the results. How much better does the model perform? Use Scikit-learn package to perform this task.

Exercise 2

Create a custom transformer that replaces outlier values of 1, 5, 20, 60 and 120 days SPX percentage returns. Determine the lower and upper bound of acceptable values based on the q-th percentile. Compare the result with the original feature set. Use the SPX dataset used in the Python Labs.

* * *