THIS DOCUMENT GIVES THE DETAILS ABOUT OUR UNDERSTANDING AND THE IMPLEMENTATION OF THE ALGORITHMS GIVEN IN THE JOURNAL

Stepwise Feature Selection

15th July 2020

INTRODUCTION

Stepwise methods start with some set of selected variables and try to improve it in a greedy fashion, by either including or excluding a single variable at each step.
Two popular members of the stepwise family are as follows,

- Forward Selection
- Backward Selection (also known as Backward Elimination)

Forward selection, starts with a (usually empty) set of variables and adds variables to it, until some stopping criterion is met.
Backward selection, starts with a (usually complete) set of variables and then excludes variables from that set, until some stopping criterion is met.

- Typically, both these methods try to include or exclude the variable that offers the highest performance increase.

Forward-Backward Selection Algorithm

Here we first perform a forward phase and then a backward phase on the selected variables.

Algorithm (FBS)

1. Init empty set S - {set of features}
2. Identify the best feature from the set of n features.
3. At every iteration start including the best feature (only if the performance is increased).
4. Loop step 2 & 3 until the set S does not change.
5. Fetch the final set S from the forward phase, feed it to the next stage.
6. Identify the worst feature from the set S.
7. At every iteration remove one feature (only if performance is not decreased).
8. Loop step 6 & 7 until the set S does not change.
9. Return S

Introduction to Early Dropping

After each forward iteration, remove all variables that do not satisfy the criterion C for the current set of selected variables S from the remaining variables R.

Algorithm (FBSED)

1. Init empty set S - {set of features}, k - Number of runs, C - criterion.
2. Identify the best feature from the set of n features.
3. At every iteration start including the best feature (only if the Criterion C is met).
4. Drop every variable which does not satisfy criterion C.
5. Loop step 2,3,4 until the given set is empty.
6. Return S.
7. Loop step 5,6 until S doesn't change or the number of runs limit is reached.
8. Return S.
9. Fetch the final set S from the forward phase, feed it to the next stage.
10. Identify the worst feature from the set S.
11. At every iteration remove one feature (only if performance is not decreased).
12. Loop step 6 & 7 until the set S does not change.
13. Return S.

In our example,

- We set the criterion by introducing a correlation matrix and setting a limit value to 0.70.
- We experimented with the value of k = 5.
- Finally, we have implemented an exhaustive feature selection algorithm on the same dataset. Taking up a little more time but has also increased the accuracy score. (This is an extra that we have implemented, not in the Journal)

Thus Forward-Backward selection with early dropping is done by finding the independent variables and eliminating them which decreases the run time significantly by giving almost the same accuracy.

Early dropping is done by finding the correlation matrix after performing each iteration of Forward Selection.

Contribution

- All the three of us understood the core algorithms and decided the parameters used. We also preprocessed data and finalized X and Y inputs.

FBS - 17PW07

FBSED - 17PW17

EFS and **Performance Metrics** - 17PW24