Name: Vaishnavi Gannavaram

Instructor: Dr. Srikanth Mudigonda

Course: 5330-01: Predictive Modelling and Machine Learning

ASSIGNMENT WEEK 05

1. **Run the code for the best subsets shared as part of this week's learning materials and identify the best subsets for the following numbers of predictors, tabulating the results in a table with the following schema:**

No. of predictors	Best combination	RMSE
1	lstat	6.165915
2	rm+lstat	5.505672
3	Rm+ptratio+lstat	5.184455
4	Rm+dis+ptratio+lstat	5.120032
5	Nox+rm+dis+ptratio+lstat	4.993611
6	Chas+nox+rm+dis+ptratio+lstat	4.942955
7	Zn+chas+nox+rm+dis+ptratio+lstat	4.909309
8	Zn+chas+nox+rm+dis+ptratio+black+lstat	4.870871
9	crim+zn+nox+rm+dis+rad+tax+ptratio+lstat	4.835974
10	crim+zn+chas+nox+rm+dis+rad+tax+ptratio+lstat	4.801439
11	crim+zn+chas+nox+rm+dis+rad+tax+ptratio+black+lstat	4.772150
12	crim+zn+indus+chas+nox+rm+dis+rad+tax+ptratio+black+lstat	4.777298
13	crim+zn+indus+chas+nox+rm+age+dis+rad+tax+ptratio+black+lstat	4.788342

2. **Next, repeat the same process, using the function regsubsets() shown in section 6.5.1, again tabulating the results.**

No. of predictors	Best combination using Regsubsets	RMSE
1	lstat	6.209619
2	rm+lstat	5.547375
3	Rm+ptratio+lstat	5.245257
4	Rm+dis+ptratio+lstat	5.265382
5	Nox+rm+dis+ptratio+lstat	5.092283
6	Chas+nox+rm+dis+ptratio+lstat	5.070490
7	black+chas+nox+rm+dis+ptratio+lstat	5.013764
8	Zn+chas+nox+rm+dis+ptratio+black+lstat	5.036151
9	crim+rad+nox+rm+dis+rad+tax+ptratio+lstat	5.088709
10	crim+zn+rad+nox+rm+dis+rad+tax+ptratio+lstat	4.978429
11	crim+zn+chas+nox+rm+dis+rad+tax+ptratio+black+lstat	4.819096
12	crim+zn+indus+chas+nox+rm+dis+rad+tax+ptratio+black+lstat	4.844115
13	crim+zn+indus+chas+nox+rm+age+dis+rad+tax+ptratio+black+lstat	4.844100

3. **Comparing the two sets of results, comment on similarities and differences in terms of**

(a) whether the same predictors were identified for each subset

In both approaches, all the subsets were the same except the 7th, 9th, and 10th subsets of the regsubsets approach.

(b) the RMSE values of k-fold CV, as reported by the two approaches

Although the RMSE values of both approaches differ slightly; the common thing is decreasing values of RMSE'S from 1st set to 13th set in both approaches.

(c) the format in which the output is displayed.

The First model has a clear format of output with a table including the RMSE values when compared to the second model which has an asterisk as an indicator of the presence of a predictor in the subset.

Comment on the time taken to fit the models (you do not have to provide the exact times taken, just a rough perception of which approach appeared to take longer).

Personally, the second model took longer time for me to understand than the first model, but the first model took longer time to run for the 13 subsets.