CSCI 6360 DATA SCIENCE II

MINGYU SUN, SHUBHANGI RAI, YE TIAN

MARCH 31, 2021

This report analyses and presents comparison between the performance of various feature selection to traditional Regression (project 1) and feature selection of Transformed regression Perceptron, Neuralnet-3L and Neuralnet-XL on the following 5/6 datasets from UCI Machine Learning Repository.
These datasets include a wide range of number of instances and features.

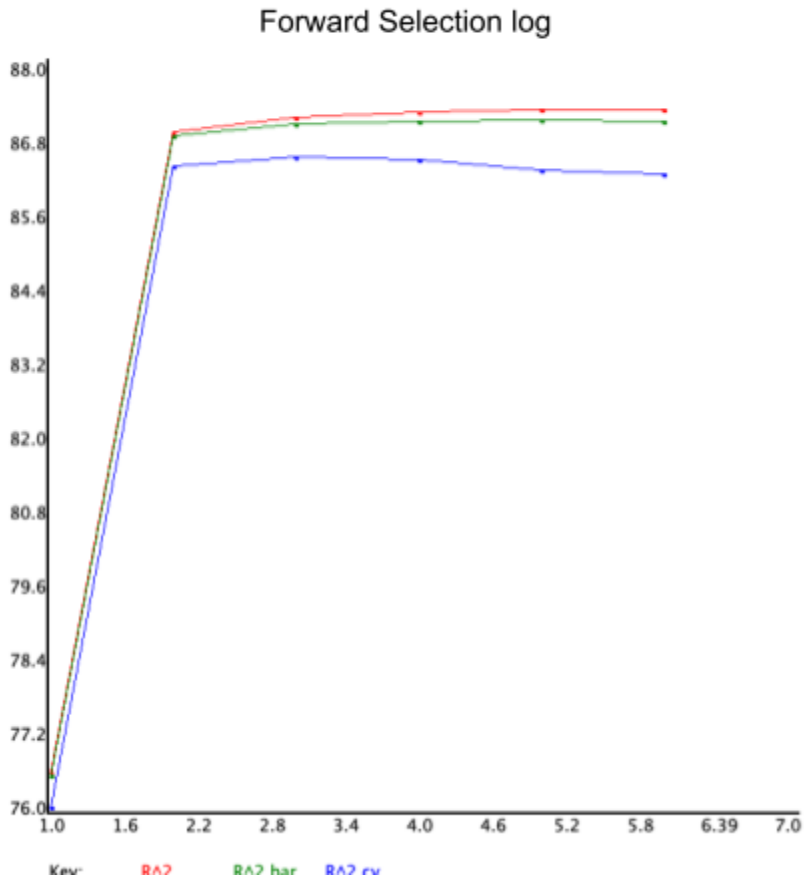
1) Auto MPG
2) Concrete Compressive
3) SkillCraft_1
4) Bias_correction
5) Forest + Fires
6) GPs Trajectories

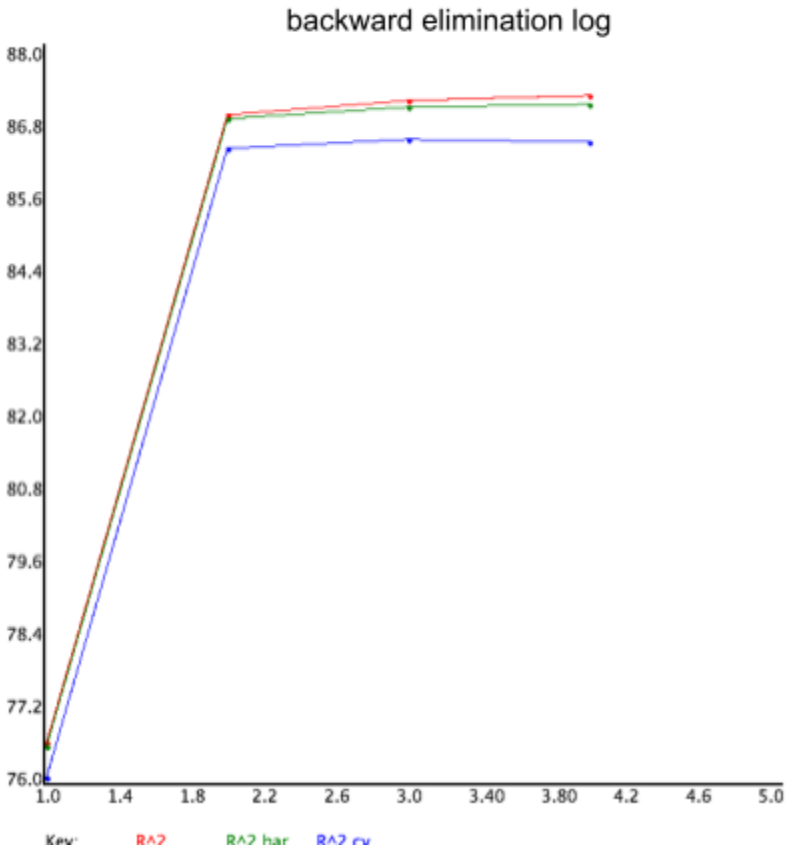
Plots of all the datasets through different techniques are saved in the `plots` folder in the ZIP file. A README has been attached in the folder to explain the naming convention.

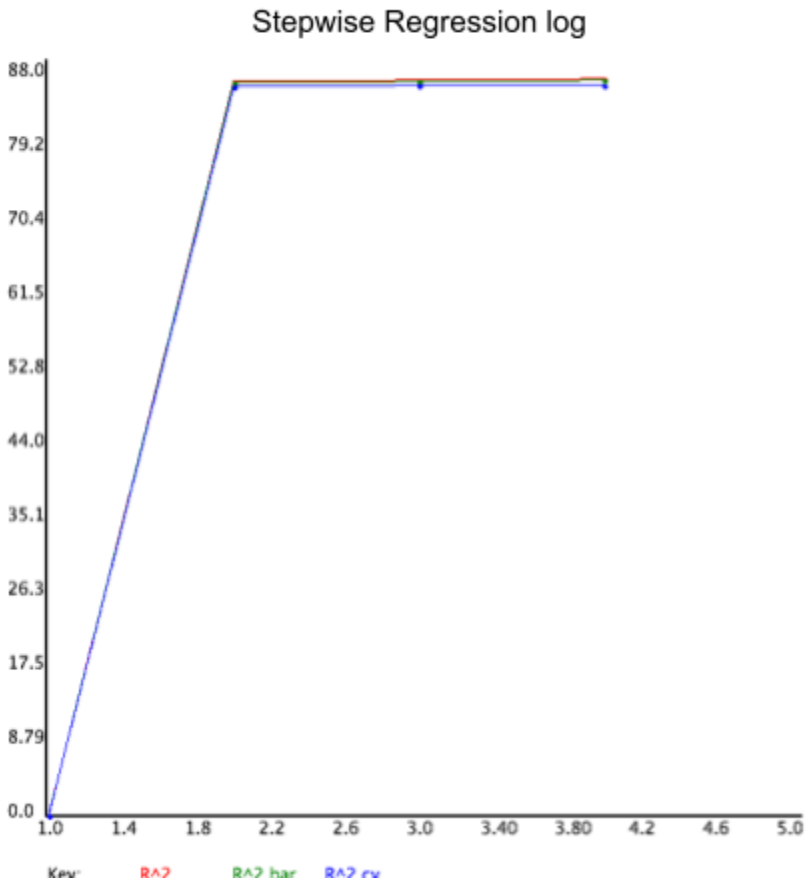
Below are the plots obtained for the **‘AutoMPG’ dataset** through different techniques with **SGDM** optimizer and **Sigmoid** Activation Function-

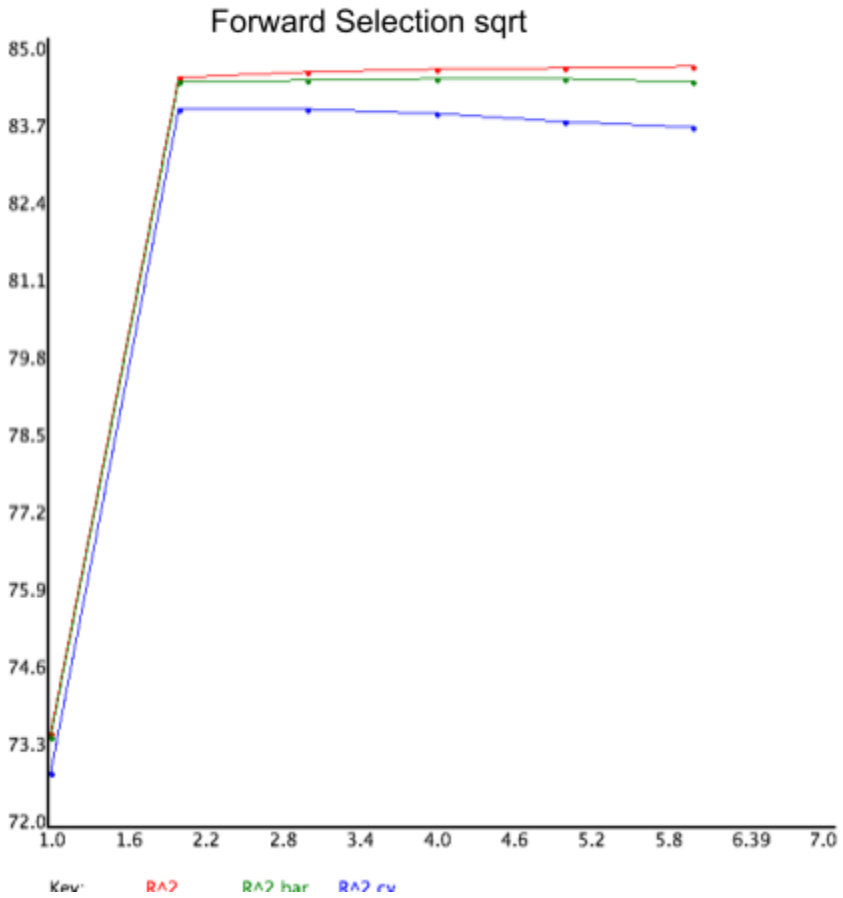
TranRegression:

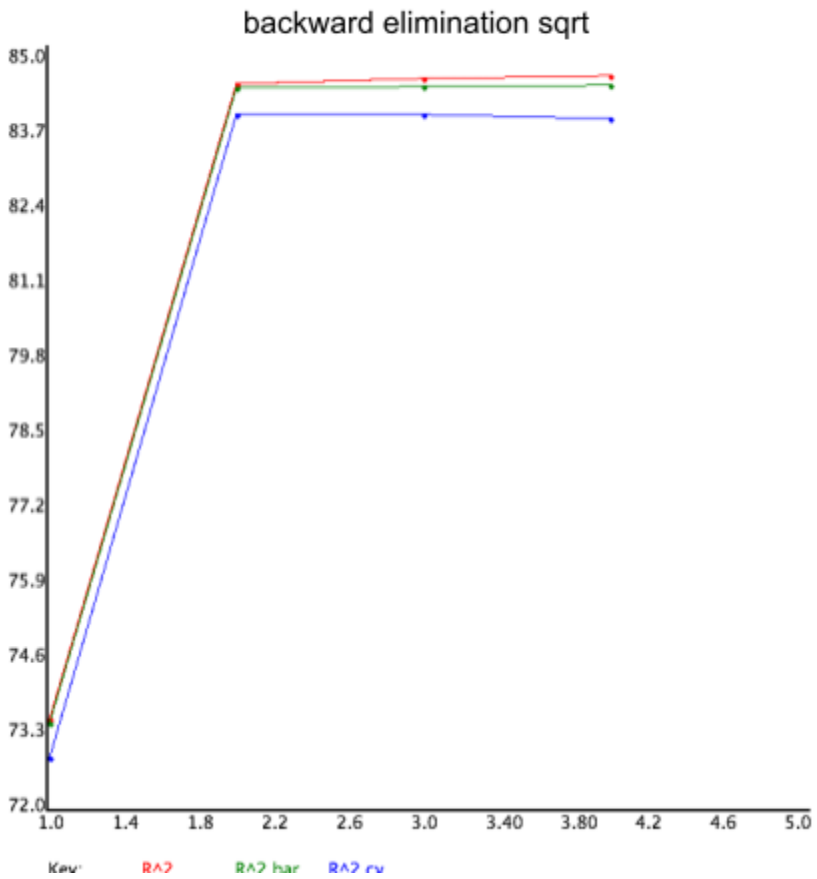
For TranRegression, the three feature selection techniques behave in a similar fashion to those in project 1, different transformations (log, sqrt, reciprocal) yield similar results. To summarize, as the number of features increases, the coefficient of determination R^2increases, the adjusted coefficient of determination $\bar{R}^2$follows closely to R^2, and the cross validationR^2_{cv}also increases at first, though it may dive down with more features. As for the comparison among the three feature selection techniques, stepwise regression should be the most robust one, while forward selection is the least. However, similar to project 1, as auto-mpg dataset is not a big one, all three feature selection techniques perform similarly, it could behave differently for datasets with more features. One difference that is showed with auto-mpg dataset is that, $\bar{R}^2$ R^2_{cv} are close to R^2 for stepwise Regression, which also proves our former statement.

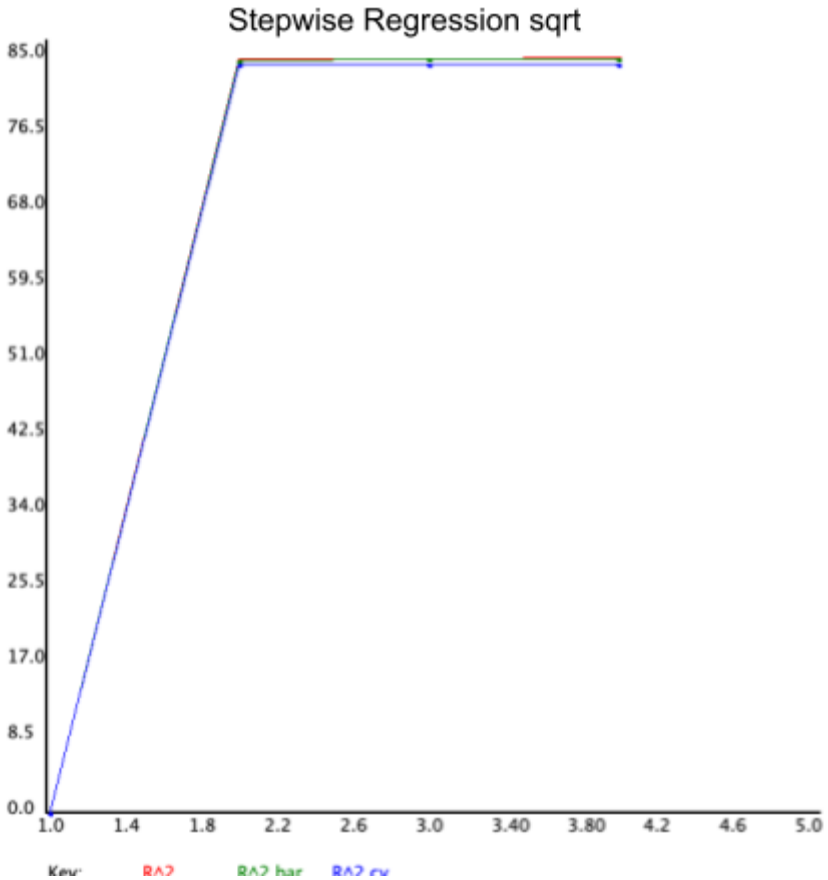


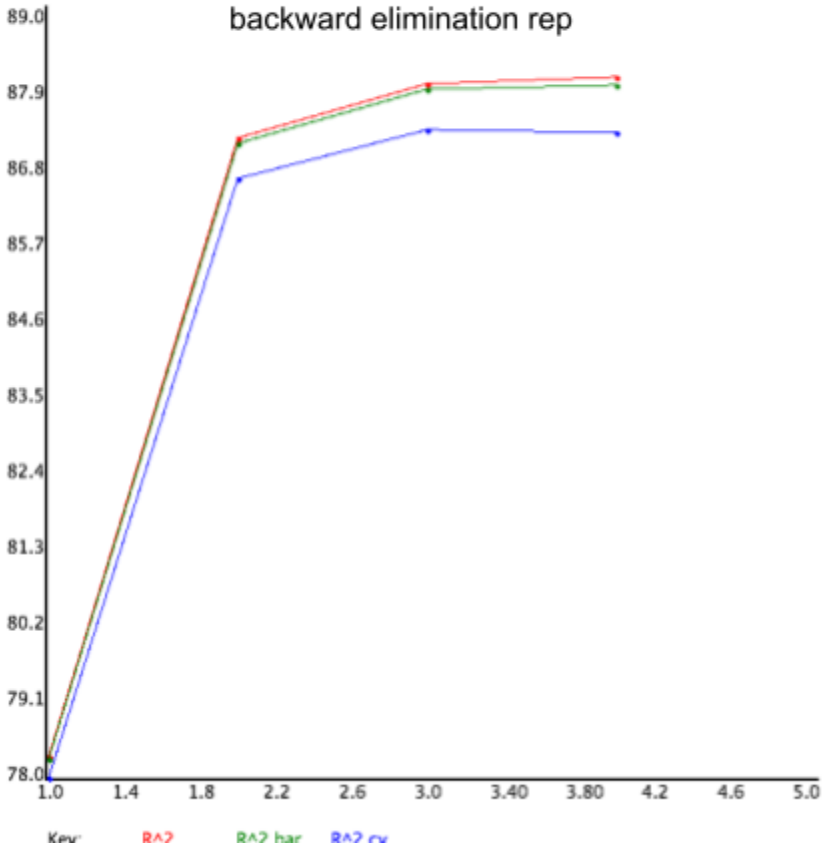

backward elimination log
88.0
86.8
85.6
84.4
83.2
82.0
80.8
79.6
78.4
77.2
76.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
RA2
RA2 bar
RA2 cv

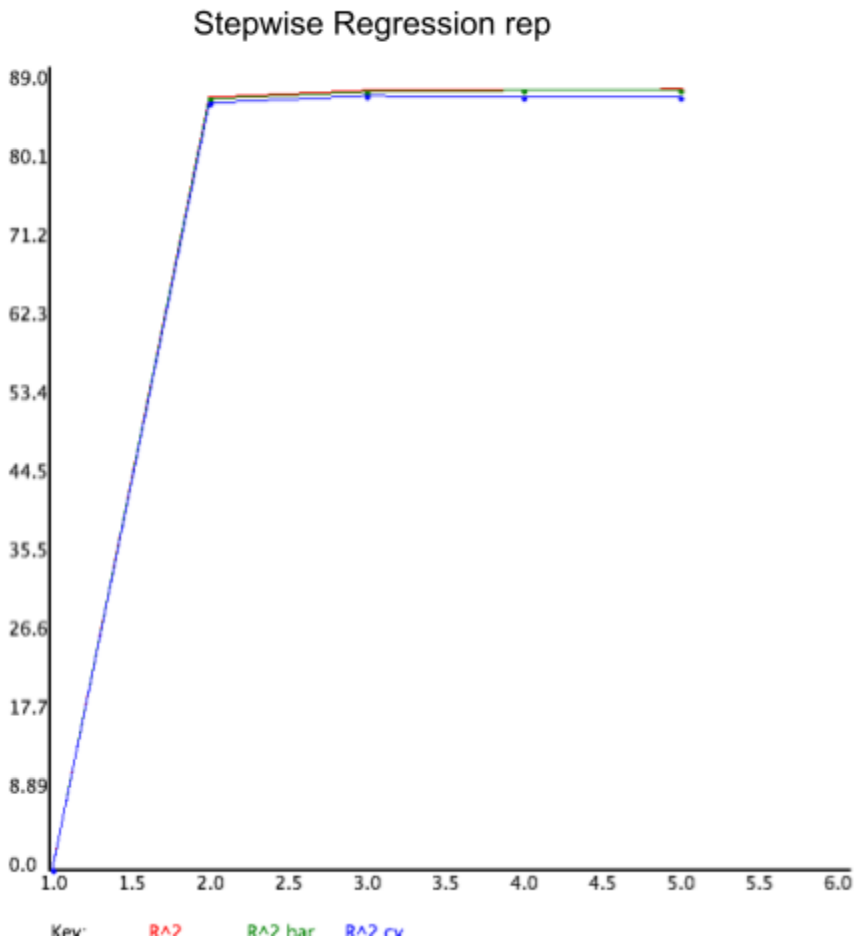

Stepwise Regression log
88.0
79.2
70.4
61.5
52.8
44.0
35.1
26.3
17.5
8.79
0.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
R^2
R^2 bar
R^2 cv


Forward Selection sqrt
85.0
83.7
82.4
81.1
79.8
78.5
77.2
75.9
74.6
73.3
72.0
1.0
1.6
2.2
2.8
3.4
4.0
4.6
5.2
5.8
6.39
7.0
Key:
RA2
RA2 bar
RA2 cv


backward elimination sqrt
85.0
83.7
82.4
81.1
79.8
78.5
77.2
75.9
74.6
73.3
72.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
RA2
RA2 bar
RA2 cv

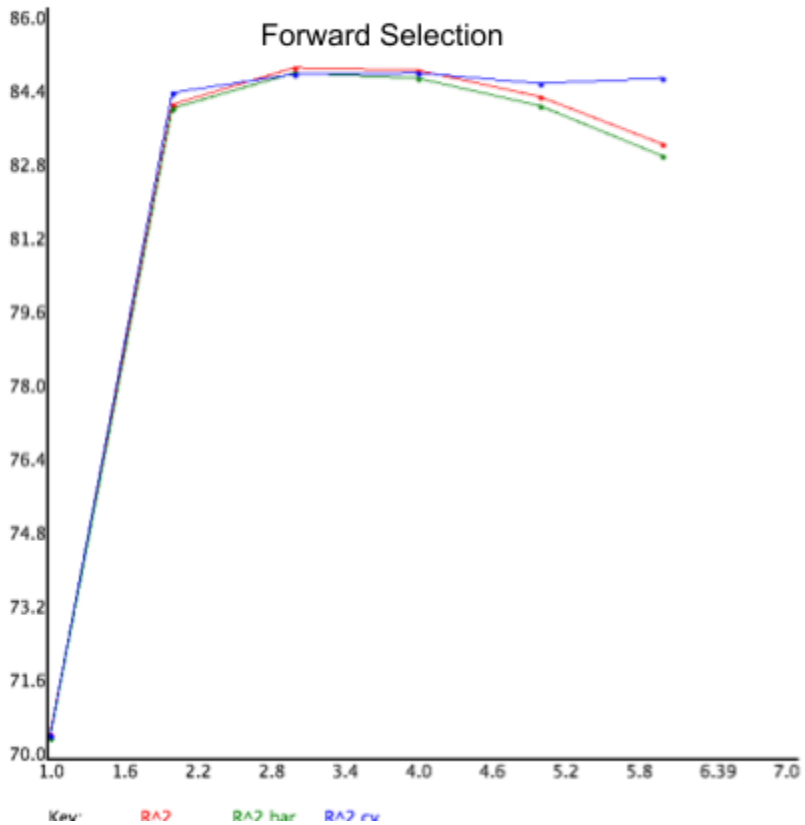

Stepwise Regression sqrt
85.0
76.5
68.0
59.5
51.0
42.5
34.0
25.5
17.0
8.5
0.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
R^2
R^2 bar
R^2 cv

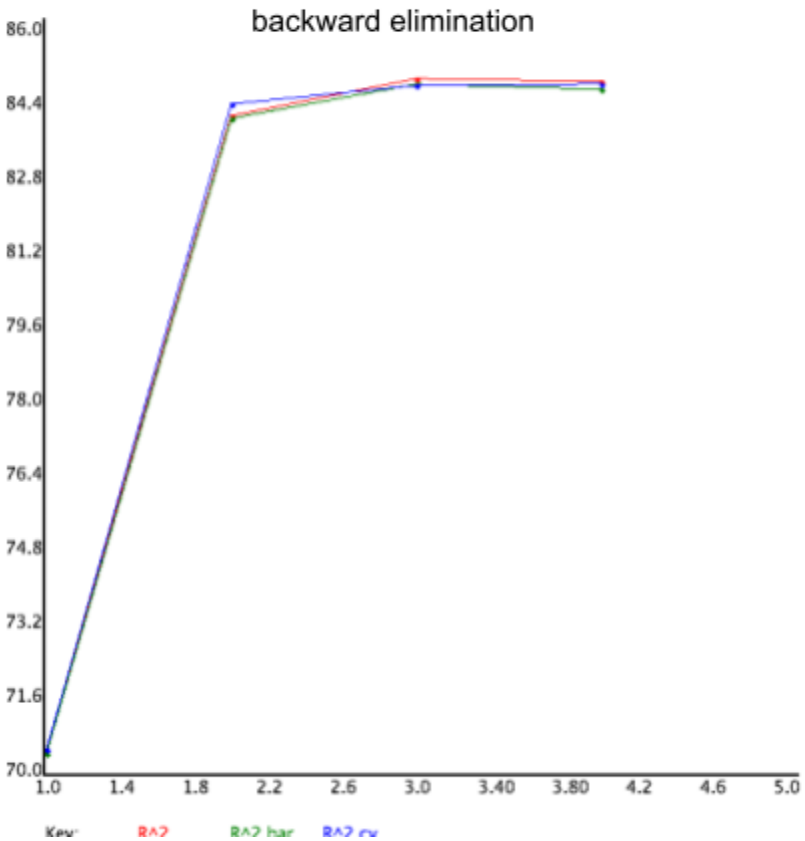

backward elimination rep
89.0
87.9
86.8
85.7
84.6
83.5
82.4
81.3
80.2
79.1
78.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0

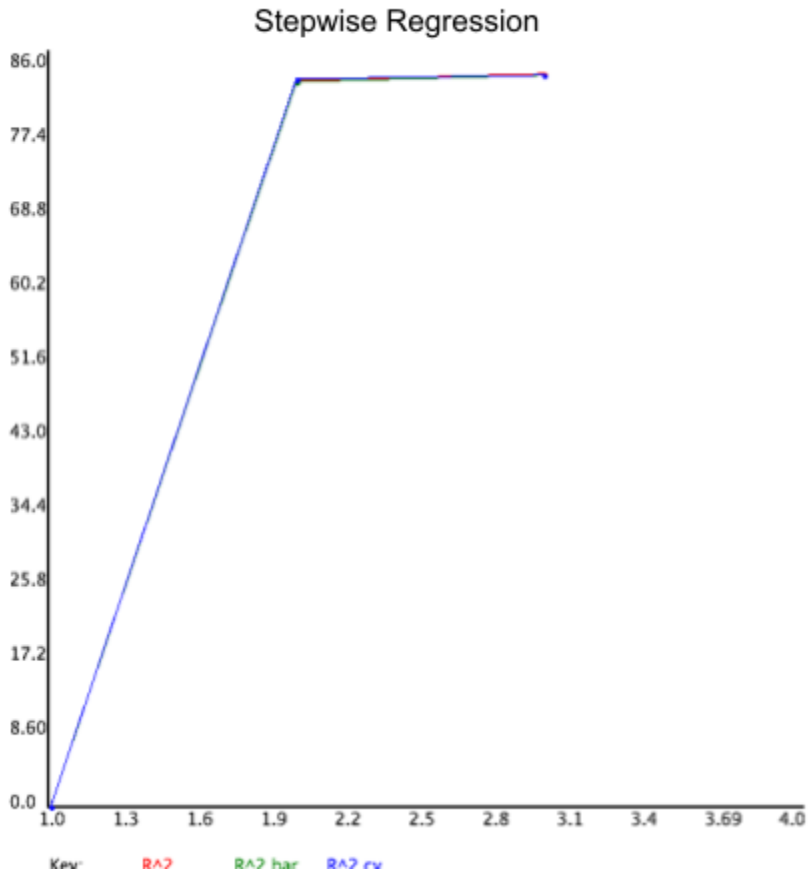


Perceptron:

Perceptron behaves a little differently from regression. $\bar{R}^2$ R^2_{cv} and R^2 have similar values, however, as for learning, Perceptron does not perform better than Regression for auto-mpg datasets. This can be seen from R^2 . Besides, the forward selection behaves even less stable for Perceptron than Regression. And we note that, for stepwise Regression, it only selects 3 features while for Regression, it always selects more than three even though the additional features do not really make the model perform better.

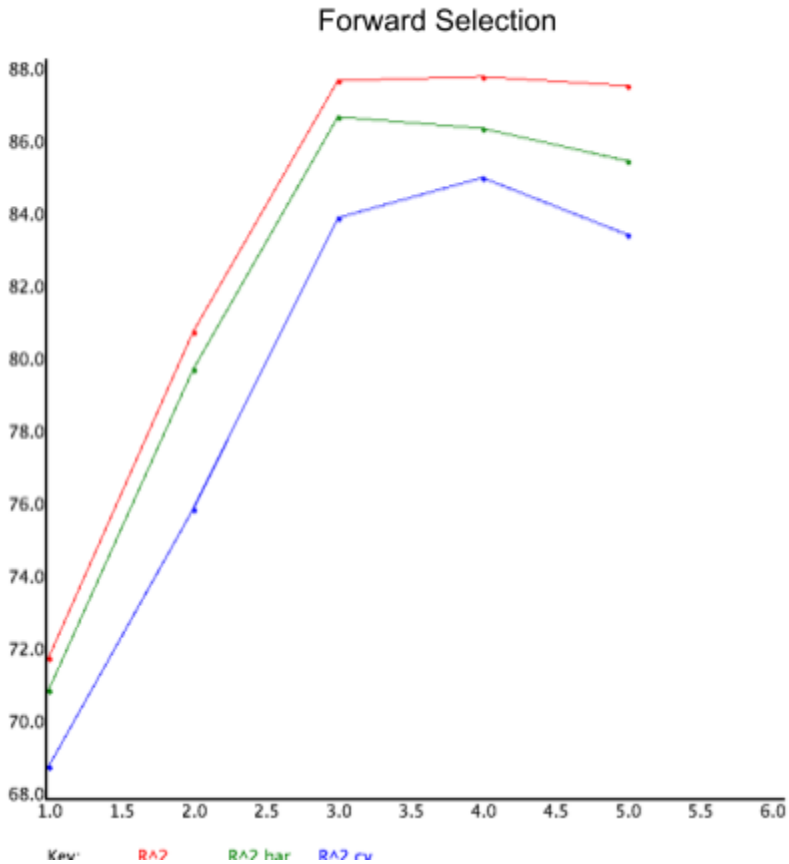

Forward Selection
86.0
84.4
82.8
81.2
79.6
78.0
76.4
74.8
73.2
71.6
70.0
1.0
1.6
2.2
2.8
3.4
4.0
4.6
5.2
5.8
6.39
7.0
Key:
RA2
RA2 bar
RA2 cv

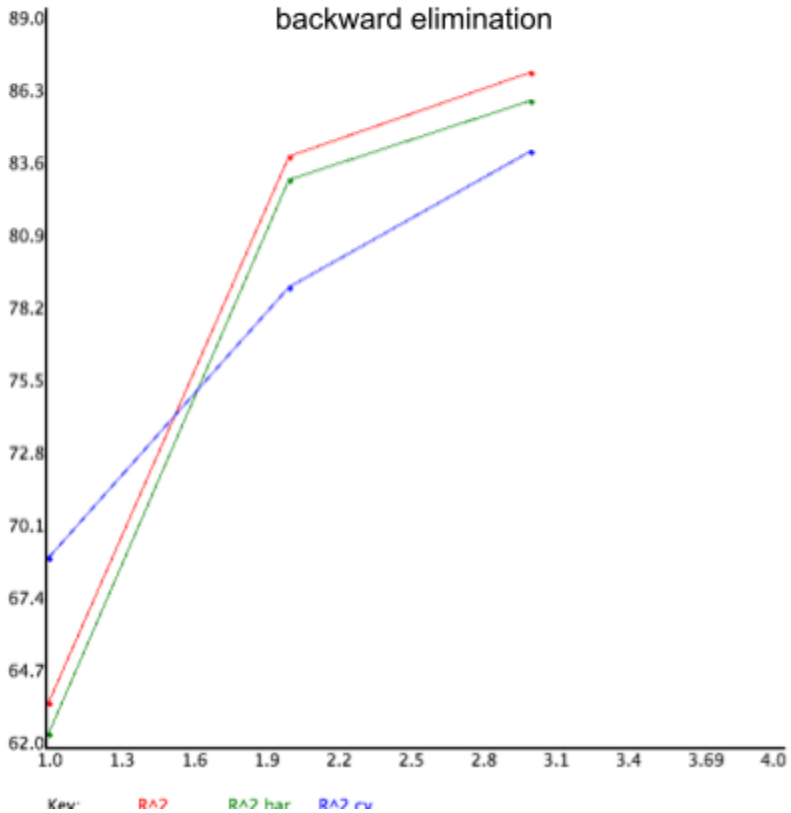

backward elimination
86.0
84.4
82.8
81.2
79.6
78.0
76.4
74.8
73.2
71.6
70.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
RA2
RA2 bar
RA2 cv

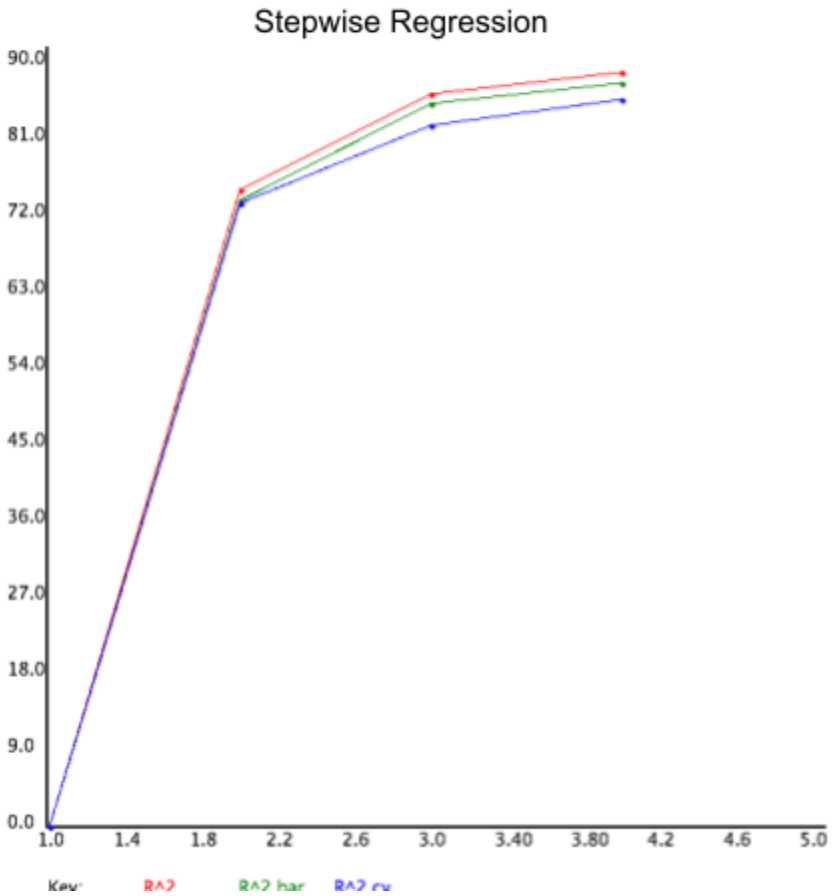


3 layers NeuralNet:

The 3 layers NeuralNet may learn the data slightly better than Perceptron and Regression models. However, when it comes to feature selection, NeuralNet behaves less stable for both forward selection and backward elimination. For different runs, we got slightly different results, the plots belows are just one run. All of $\bar{R}^2$ R^2_{cv} and R^2 could dive down for NeuralNet, not just R^2_{cv}. Stepwise Regression is quite robust, the $\bar{R}^2$ R^2_{cv} and R^2 never really decrease with increasing features. This 3 layers NeuralNet is relatively predictable, as it is not a very complicated model.

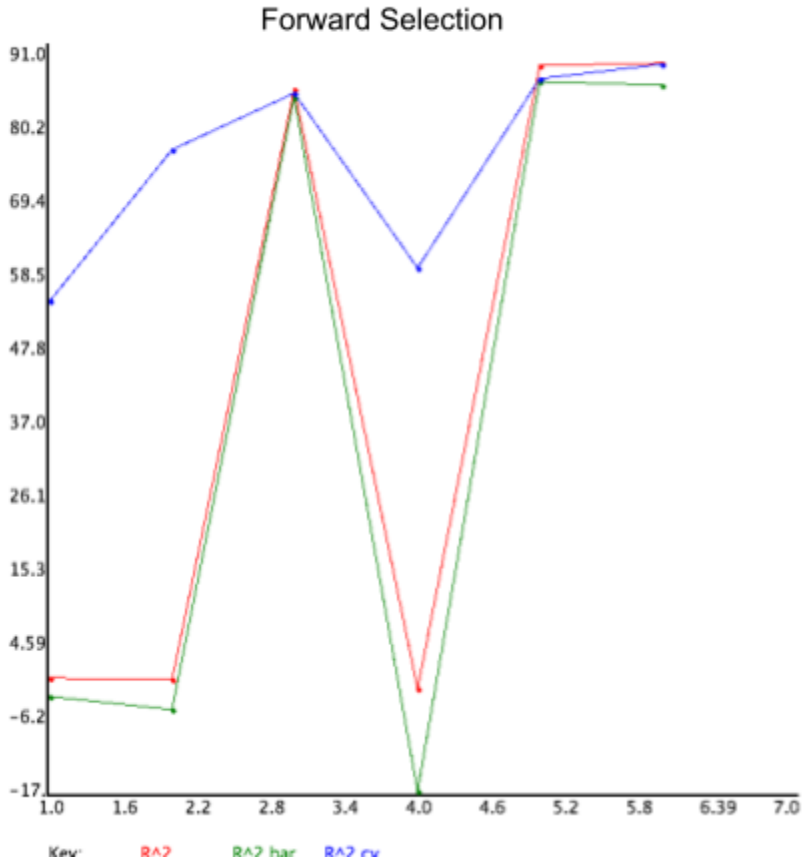

Forward Selection
88.0
86.0
84.0
82.0
80.0
78.0
76.0
74.0
72.0
70.0
68.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
Key:
R^2
R^2 bar
R^2 cv

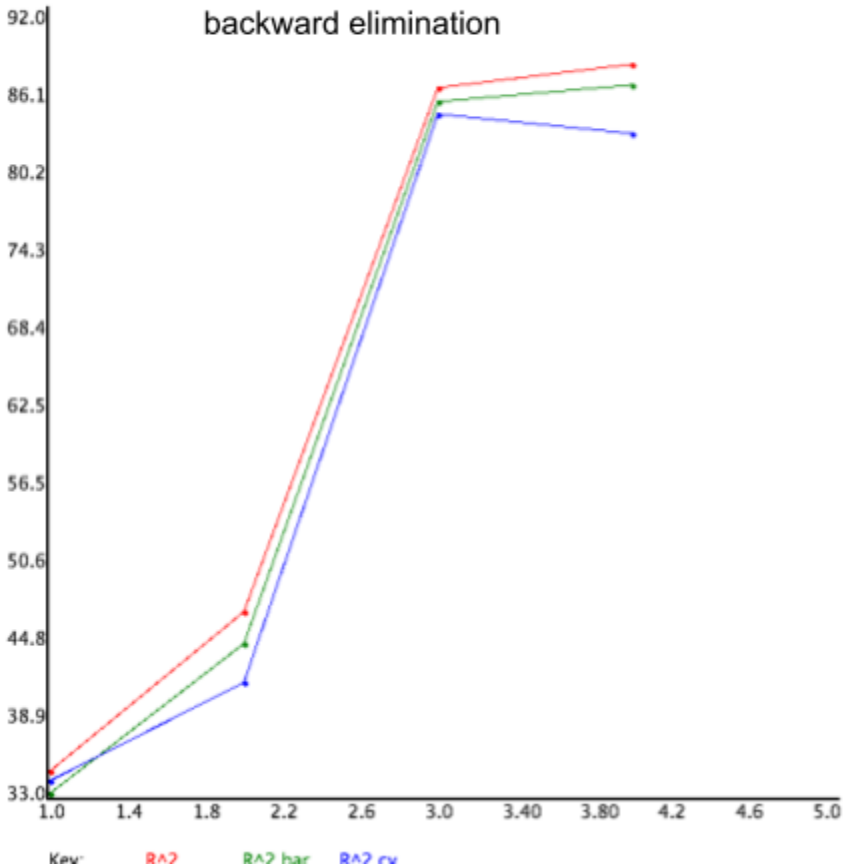

backward elimination
89.0
86.3
83.6
80.9
78.2
75.5
72.8
70.1
67.4
64.7
62.0
1.0
1.3
1.6
1.9
2.2
2.5
2.8
3.1
3.4
3.69
4.0
Key:
R^2
R^2 bar
R^2 cv

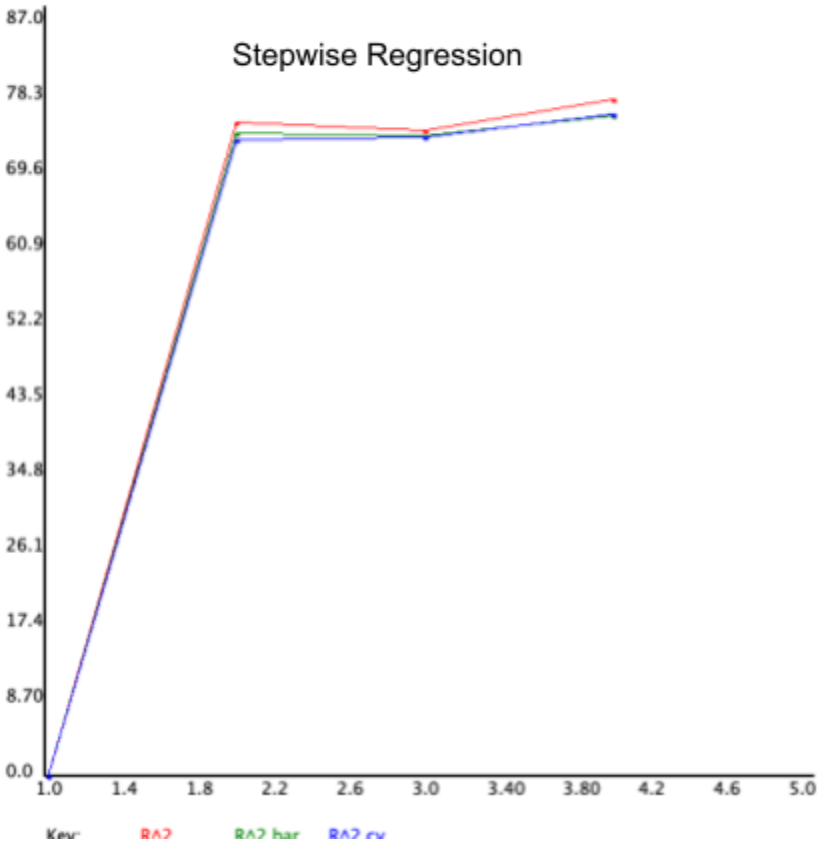


4 layers NeuralNet:

4 layers NeuralNet behaves more unpredictable than 3 layers NeuralNet, although it may get better result at last, based on R^2 . This means that 4 layers NeuralNet is capable of capturing more features of the dataset, but it is also harder to be trained. As stated in 3 layers NeuralNel, all of $\bar{R}^2$ R^2_{cv} and R^2 could dive down for NeuralNet, but they usually come back at the end. Besides, as different runs give slightly different results, some runs did provide pretty looking results, without diving down in the middle. And stepwise Regression is still the most robust one.


Forward Selection
91.0
80.2
69.4
58.5
47.8
37.0
26.1
15.3
4.59
-6.2
-17
1.0
1.6
2.2
2.8
3.4
4.0
4.6
5.2
5.8
6.39
7.0
Key:
RA2
RA2 bar
RA2 cv


backward elimination
92.0
86.1
80.2
74.3
68.4
62.5
56.5
50.6
44.8
38.9
33.0
1.0
1.4
1.8
2.2
2.6
3.0
3.40
3.80
4.2
4.6
5.0
Key:
R^2
R^2 bar
R^2 cv



If there is a large amount of data, Neural Network models would perform better than the traditional Regression models.