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 to Transformed regression Perceptron, Neuralnet-3L and Neuralnet-XL on the following 5 datasets from UCI Machine Learning Repository.
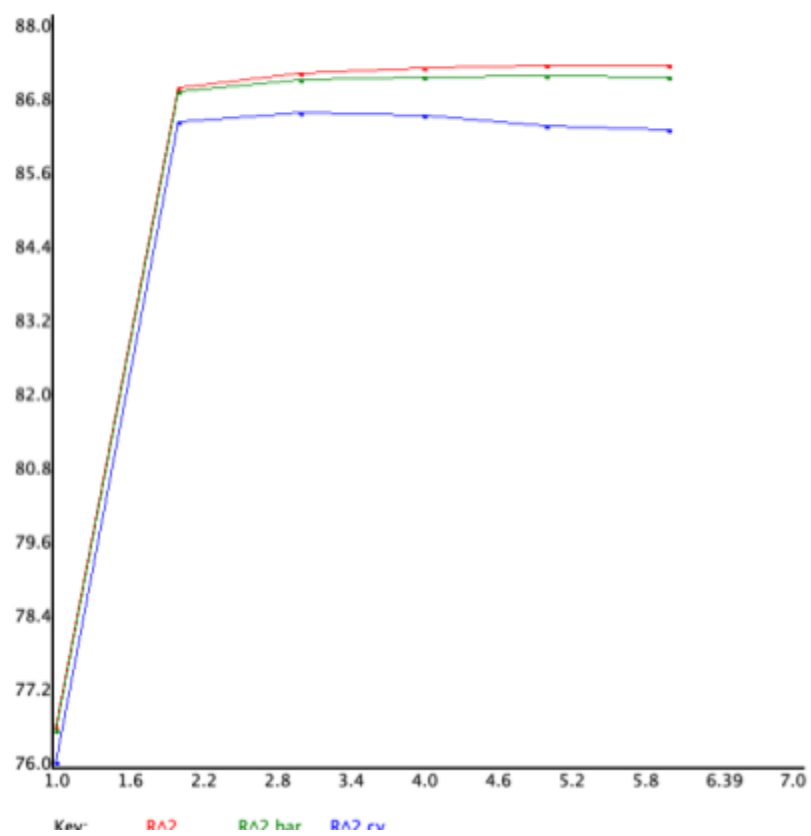
These datasets include a wide range of number of instances and features.

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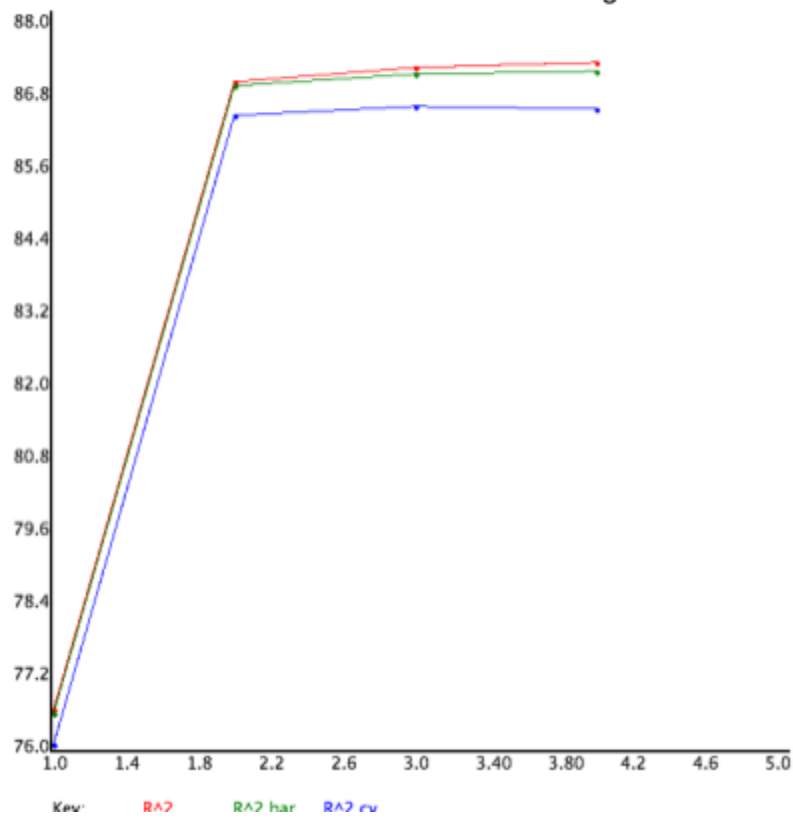3) SkillCraft_1
4) Bias_correction
5) Forest + Fires

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^2 increases, the adjusted coefficient of determination $\bar{R}^2$ follows closely to R^2, and the cross validation R^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.
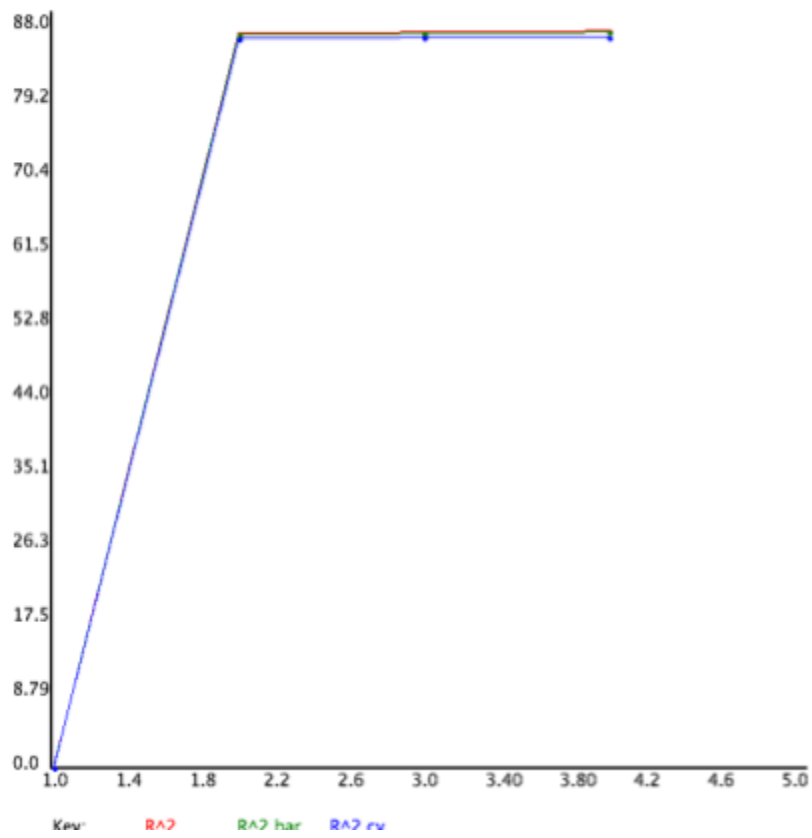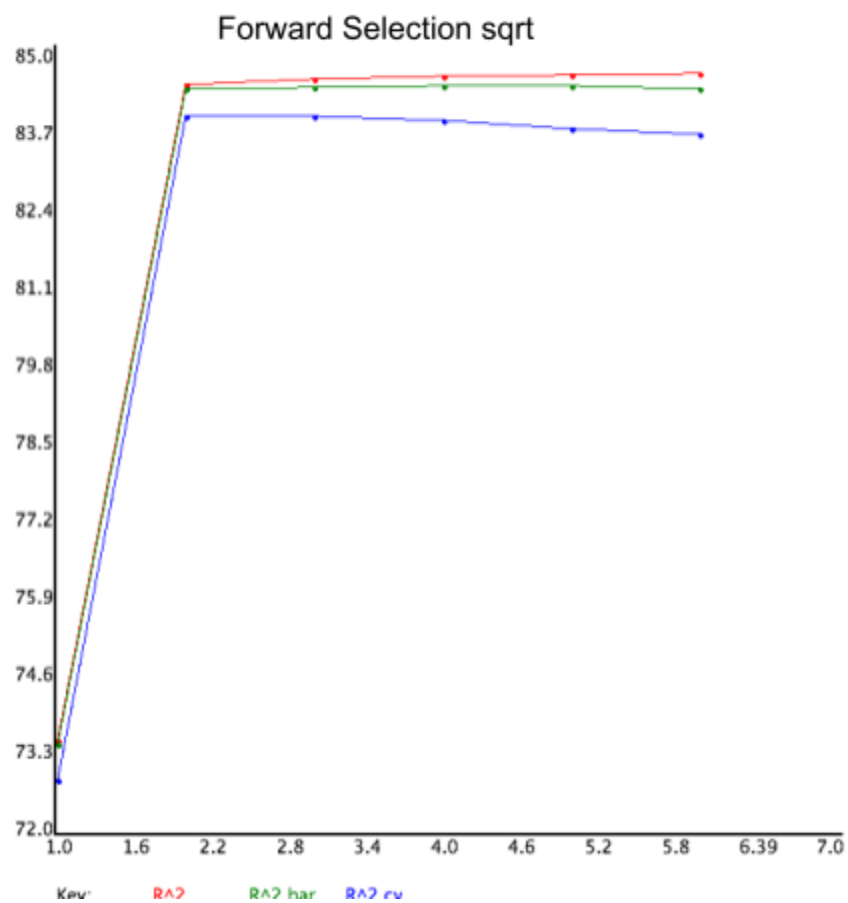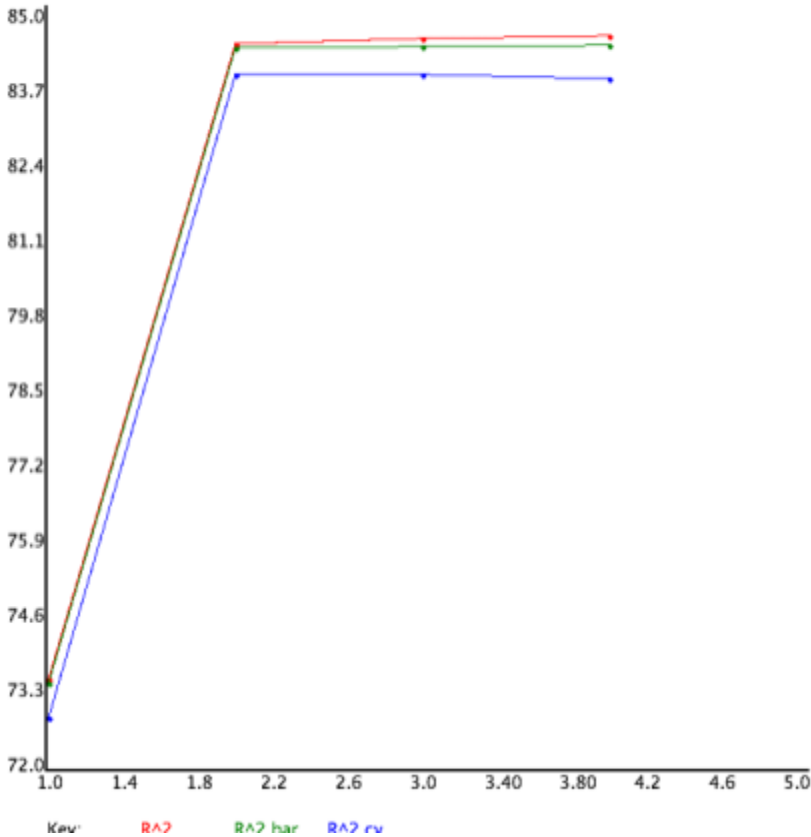
Forward Selection log

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: R^2   R^2 bar   R^2 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: R^2   R^2 bar   R^2 cv

backward elimination sqrt

	1.0	1.4	1.8	2.2	2.6	3.0	3.40	3.80	4.2	4.6	5.0
85.0											
83.7											
82.4											
81.1											
79.8											
78.5											
77.2											
75.9											
74.6											
73.3											
72.0											

Key: R^2   R^2 bar   R^2 cv

Stepwise Regression sqrt

Key: R^2   R^2 bar   R^2 cv

backward elimination rep

Key: R^2   R^2 bar   R^2 cv

Stepwise Regression rep
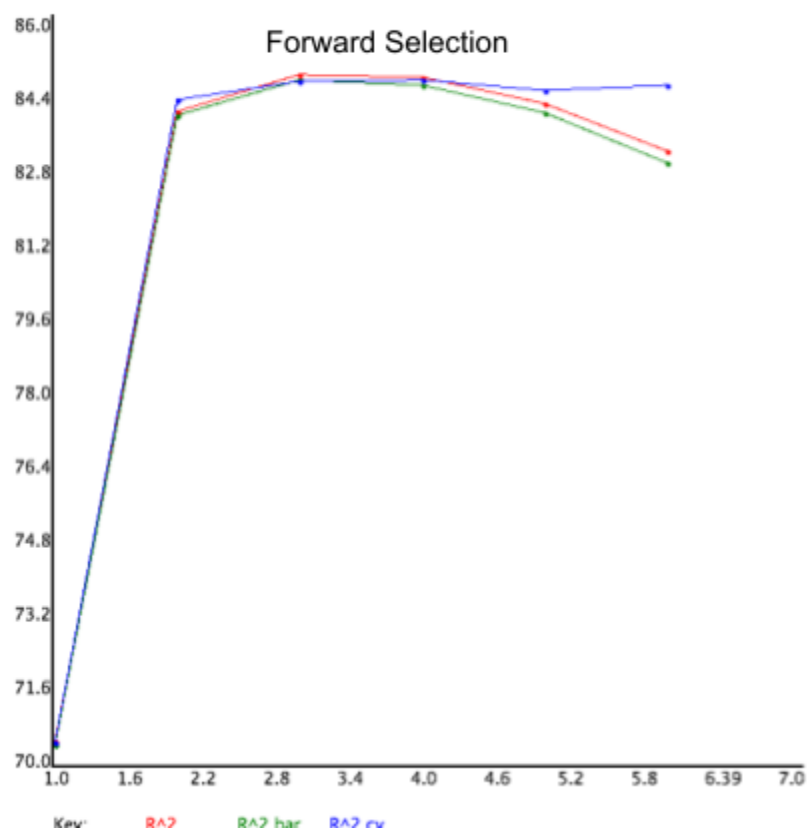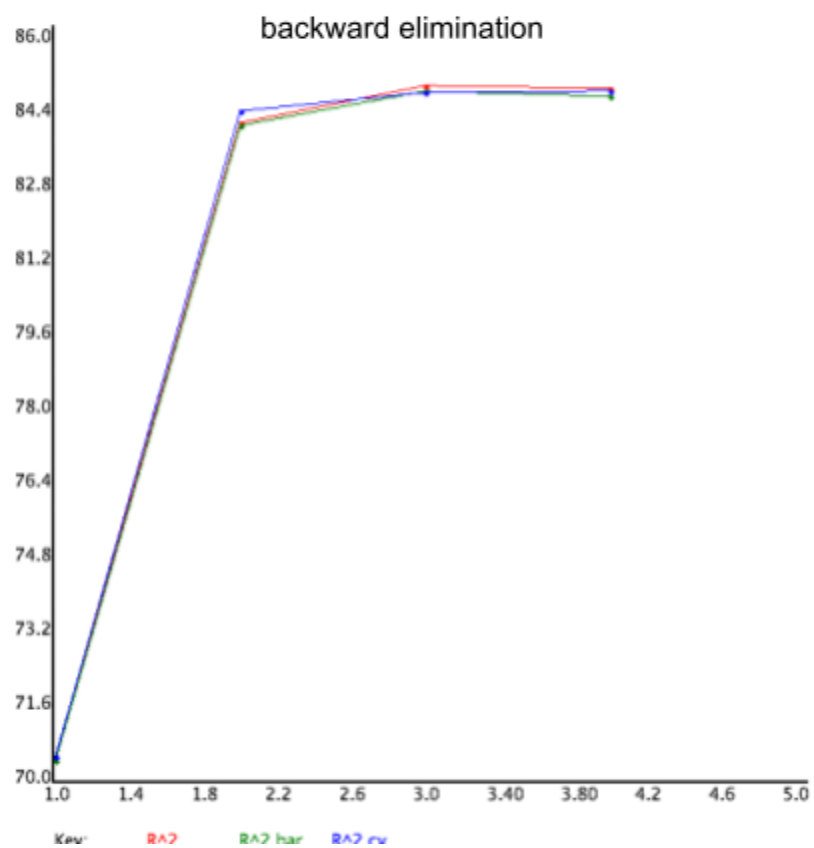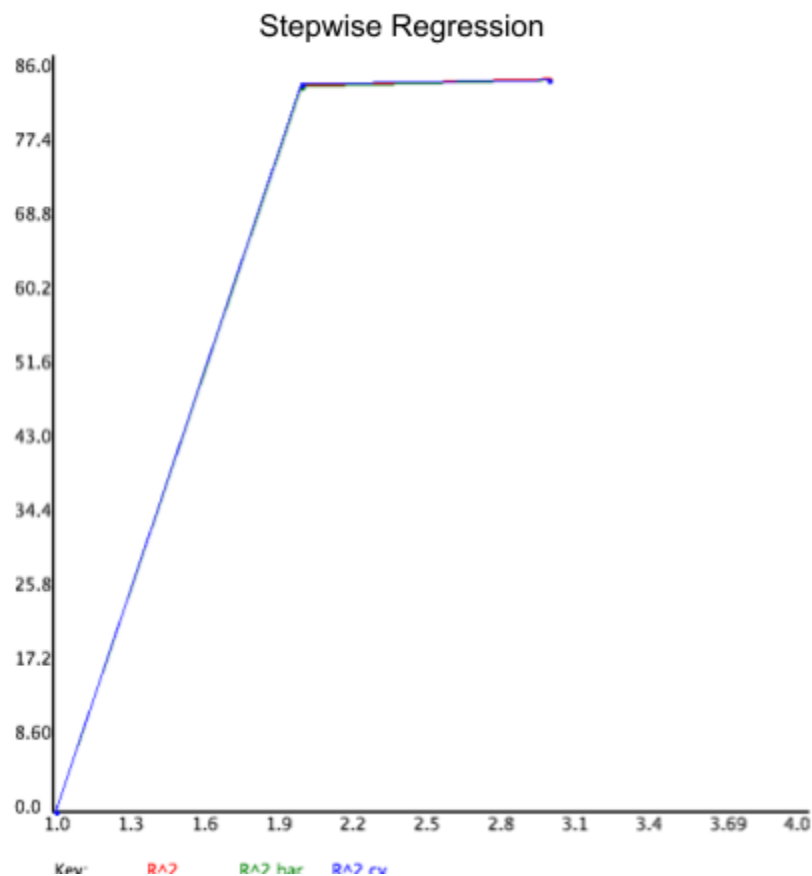
Key: R^2   R^2 bar   R^2 cv

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

Key: R^2   R^2 bar   R^2 cv

backward elimination

Key: R^2   R^2 bar   R^2 cv

Stepwise Regression

86.0
77.4
68.8
60.2
51.6
43.0
34.4
25.8
17.2
8.60
0.0

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

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.
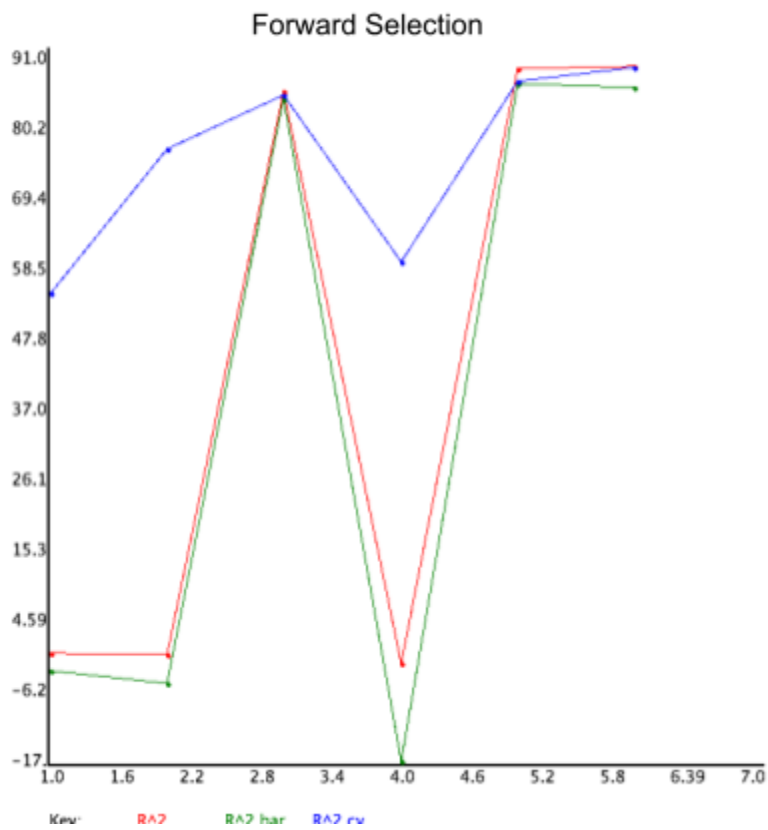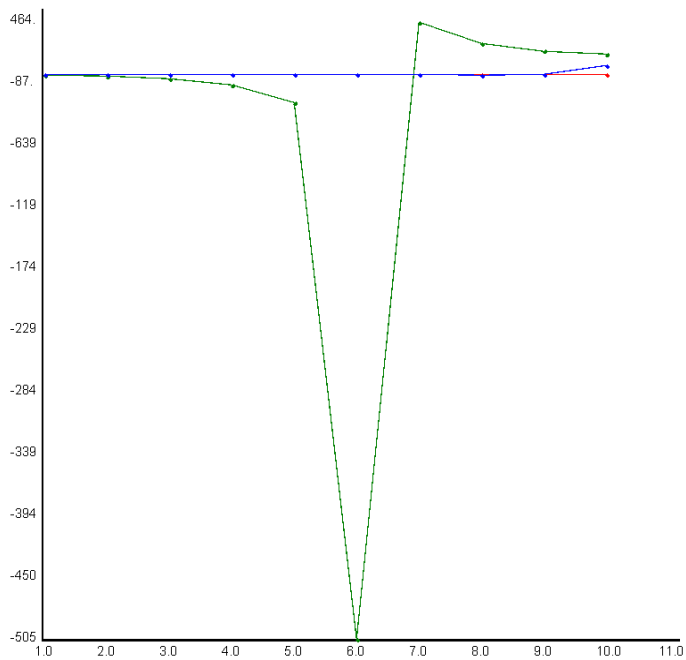
Forward Selection

Key: R^2   R^2 bar   R^2 cv

backward elimination

Key: R^2   R^2 bar   R^2 cv

Stepwise Regression

Key: RA2   RA2 bar   RA2 cv
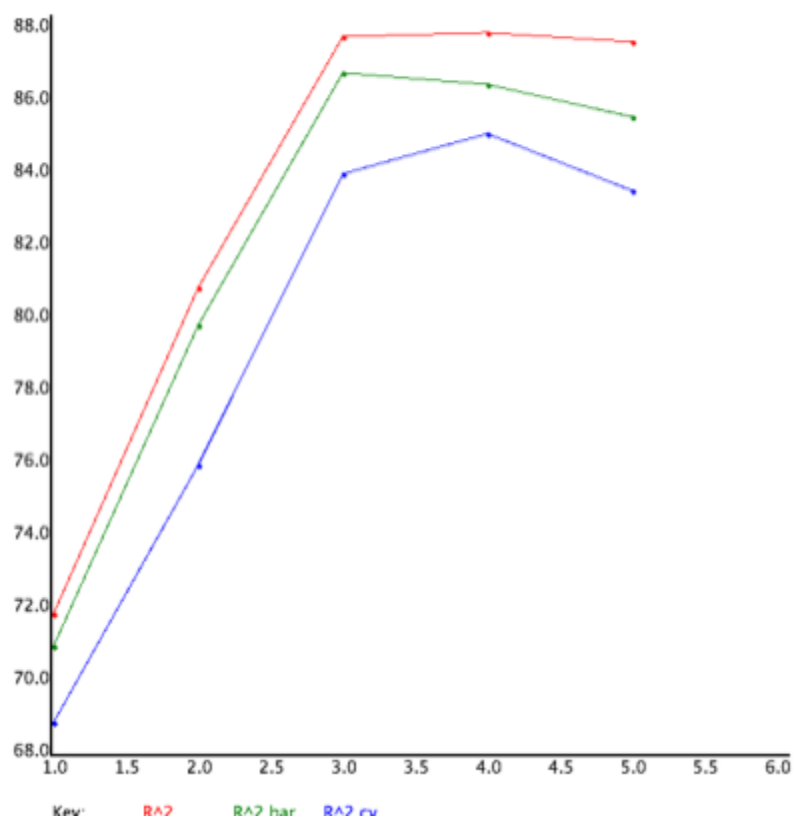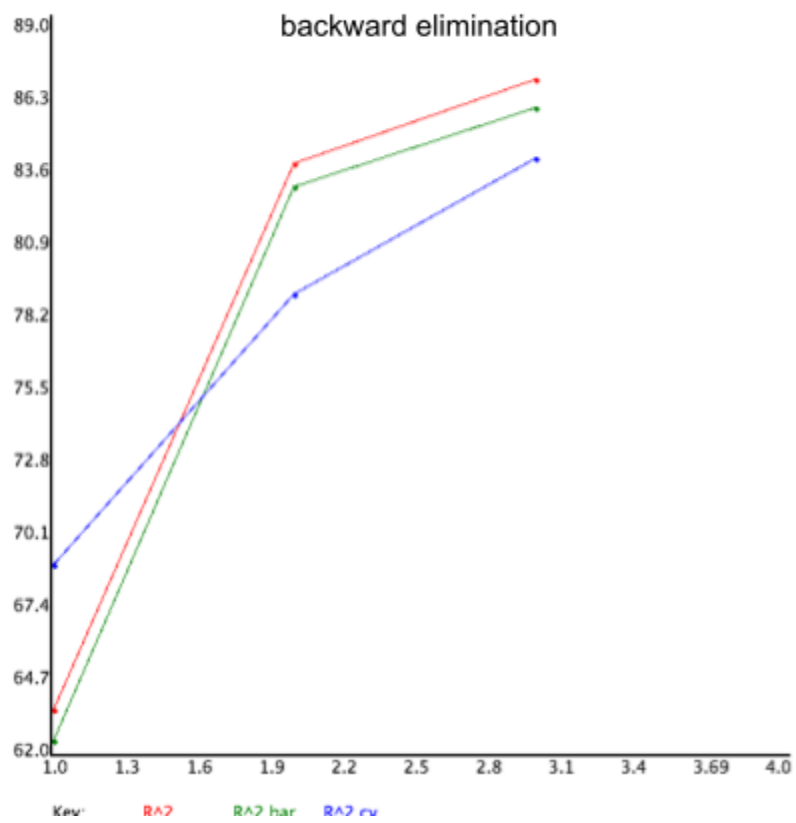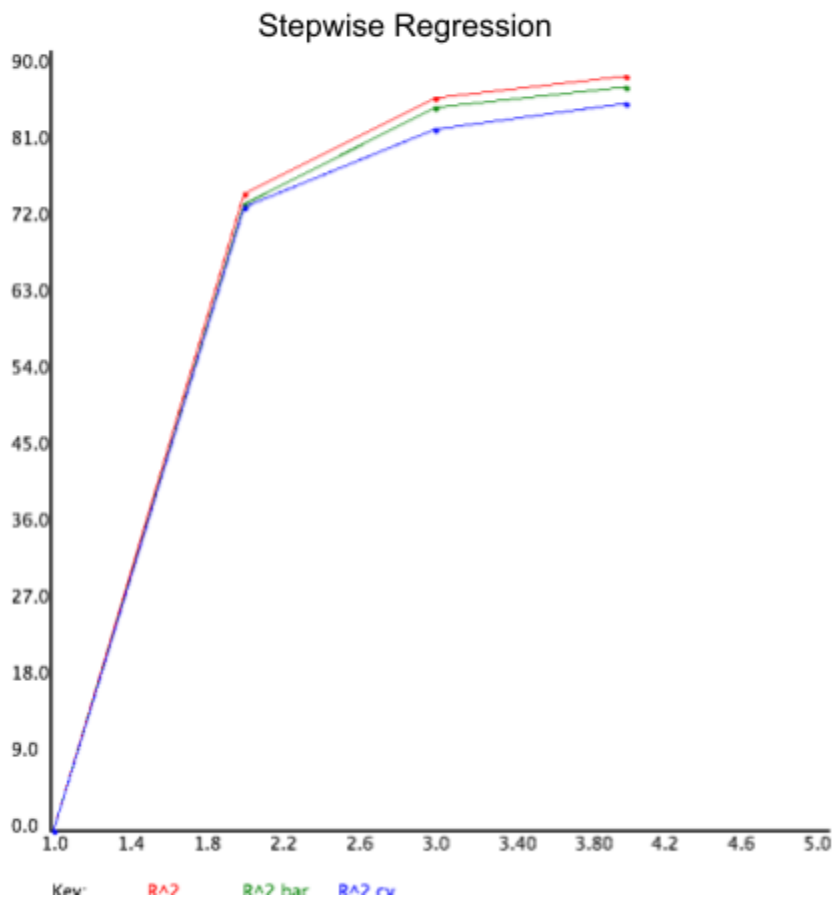
4 layers NeuralNet:

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

Key: RA2   RA2 bar   RA2 cv

Below is another graph for forward selection regression on 4 layers neural networs (with Concrete datasets using scalation):

Which behaves rather strangely, I ran the code multiple times and roughly it has been similar pattern. Nonetheless, comparable to the previous example with forward selection on 4 layers neural networks, there's an obvious "dive down" of the graph line in the middle of the process, and in this specific example, that "jump" is rather drastic.

Stepwise Regression

Key: R^2   R^2 bar   R^2 cv
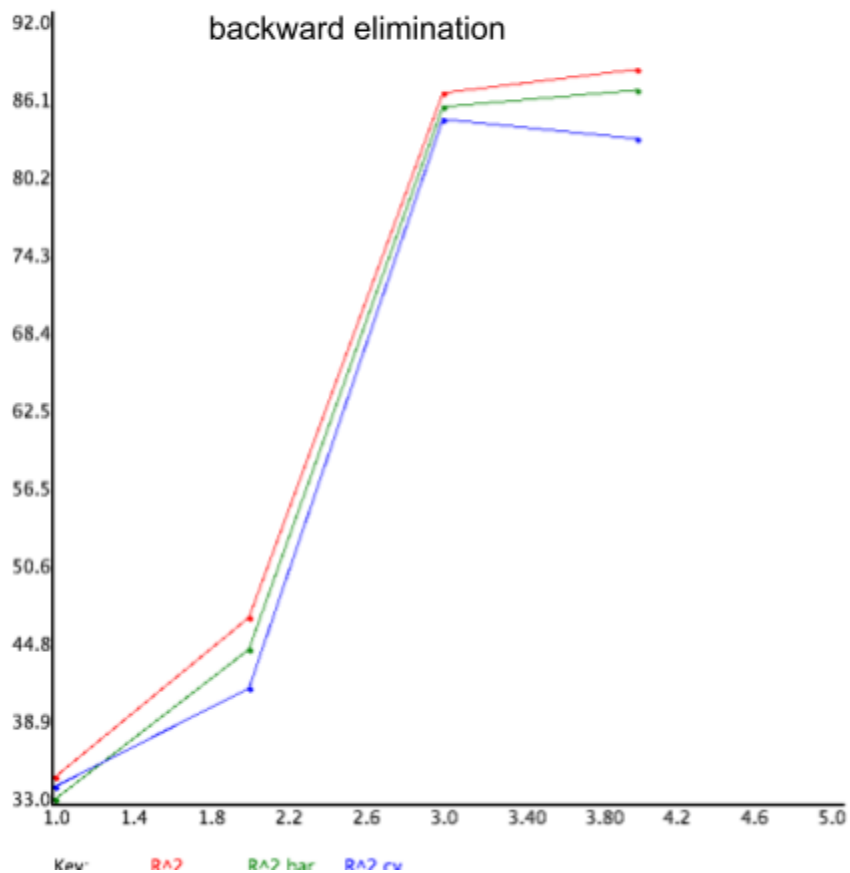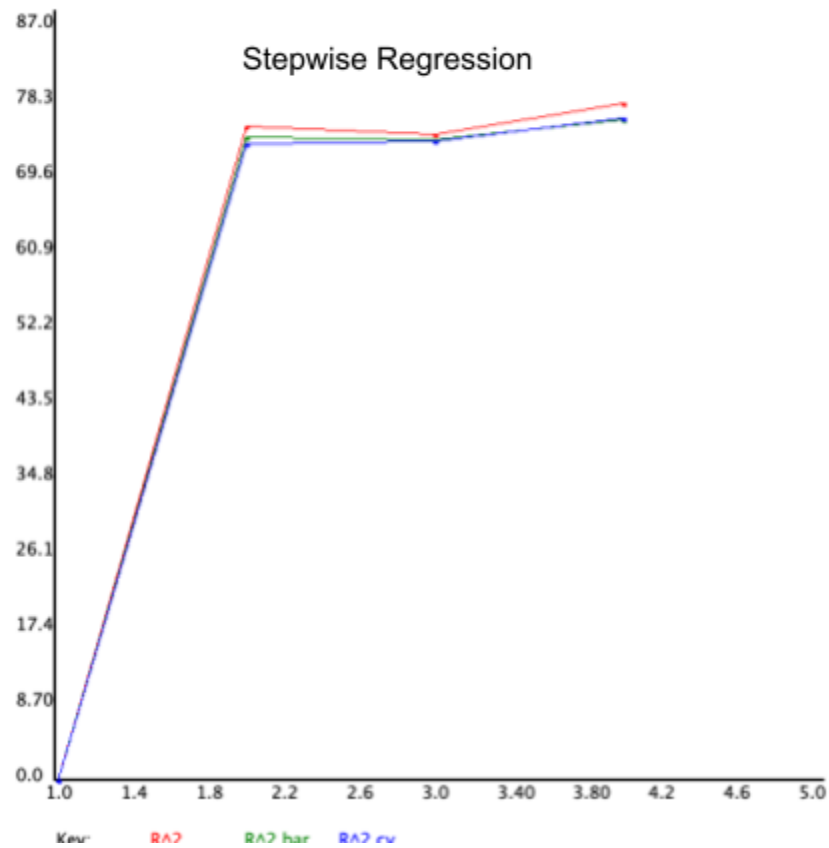
If there is a large amount of data, Neural Network models would perform better than the traditional Regression models.