To begin,

In this project we used Transformed regression and three kinds of Neural Networks along with feature selection methods such as Forward Selection, Backward Elimination, and Stepwise Regression to test how well these models performed when compared to regular polynomial regression used with feature selection. The Neural Networks we used were: Perceptron, no hidden layers, Neural Net 3L, 1 hidden layer, and Neural Net XL, 2 hidden layers.

We measured R^2, Adjusted R^2, Cross Validated R^2, and Akaike Information Criterion (AIC) to for our models, but chose to generally use Adjusted R^2 when evaluating them for feature selection since it adds a penalty for adding more features as opposed to general R^2. AIC is useful in comparing models to see if one model is better than another, as the lower AIC between models would imply a better model, but it is not useful for evaluating a model's actual performance since it doesn't explain how well a model is fit.

Summary:
After using Transformed regression and three kinds of Neural Networks along with feature selection methods such as Forward Selection, Backward Elimination, and Stepwise Regression, we can compare how well these models performed compared to the polynomial regressions used with feature selection. The Neural Networks we used were: Perceptron, no hidden layers, Neural Net 3L, 1 hidden layer, and Neural Net XL, 2 hidden layers.

Transformed Regression

When running with transformed regression, what we saw was a pretty good model. The fit was typically better than that of the multiple linear regression of the last project. Cubic regression still fits to the curve better but that is to be expected. It tended to outperform Lasso and Ridge as well, but not always. There were some instances where Transformed Regression did not perform as well as those from the latter project, typically in datasets with less related variables such as the wine dataset.

Perceptron

Perceptron was an interesting model to use. It was very consistent in its performance. It came in around the middle of the pack on each dataset. This can be useful as although it was not the best model, it was never the worst. We would typically not recommend using perceptrons as the preferred model though, unless the 3 and 4 layer neural nets are both struggling,

3 Layer Neural Network

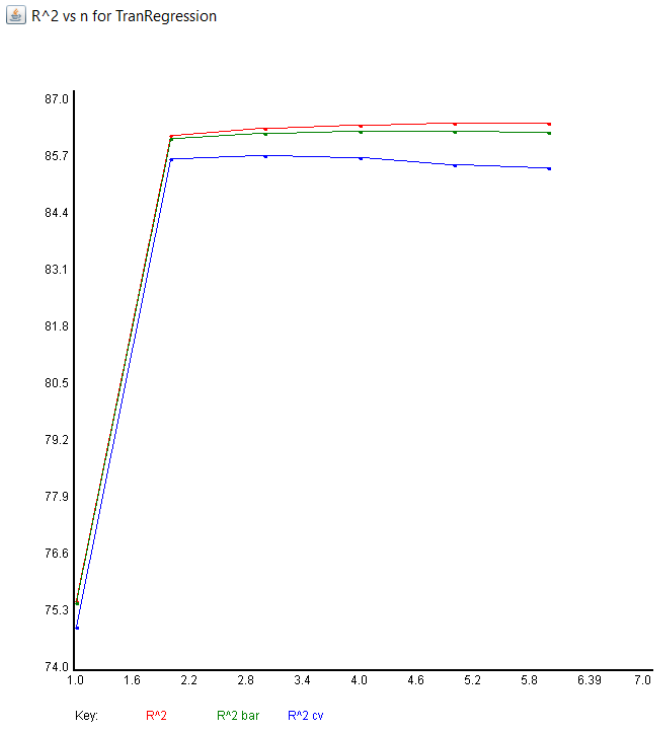
The 3 layer neural network was the worst. It was outperformed consistently and struggled to even compete with other models at times. We would not recommend using this model as it is hard to even think of a redeeming factor that it contributed.

4 Layer Neural Network
The 4 layer Neural Network performed on both extreme ends of the spectrum so it is tough to sing its praise without mentioning its failures. It was not uncommon for it to be the best performing model, but it also performed quite poorly at times to where there was reason for concern. In the right circumstances though, it would be our recommended model.

The most telling plots of our models can be found below.

<u>Dataset: AutoMPG</u>
Our first dataset, the AutoMPG, is a dataset in which there are 398 cars, and based on 6 predictors (cylinders, displacement, horsepower, weight, acceleration and model year), we are attempting to predict the Miles Per Gallon that a given car gets.

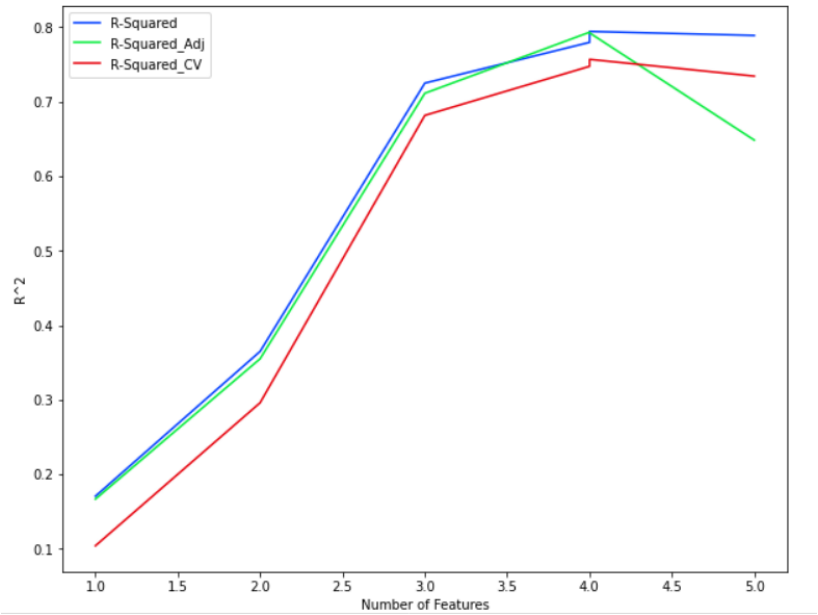
Transformed Regression:

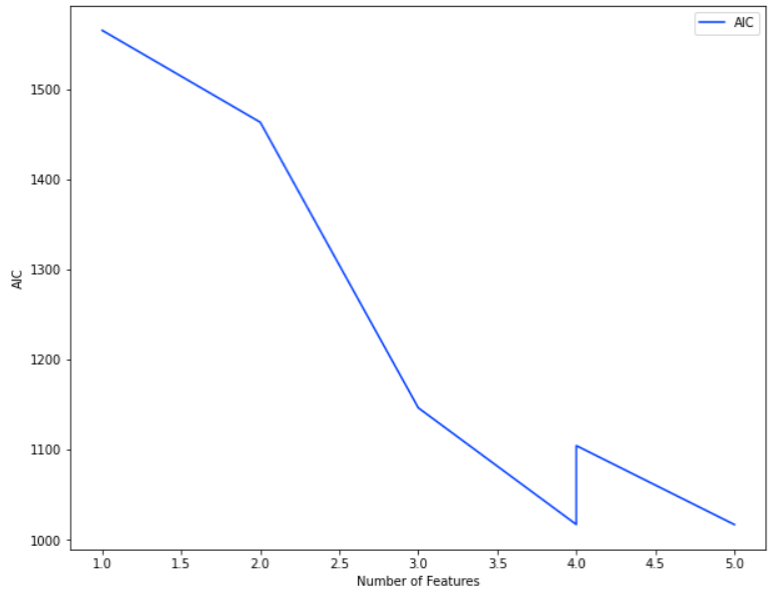


Perceptron:

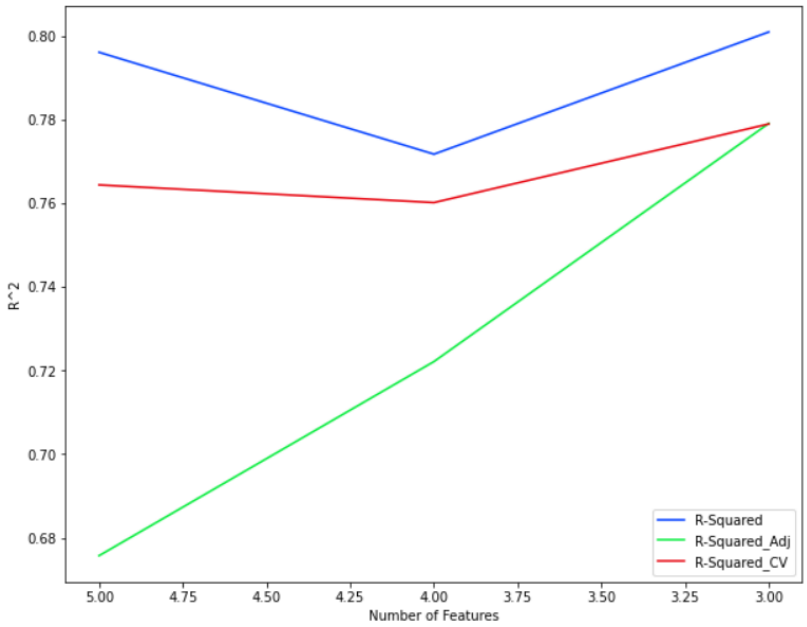
Forward Best Features:

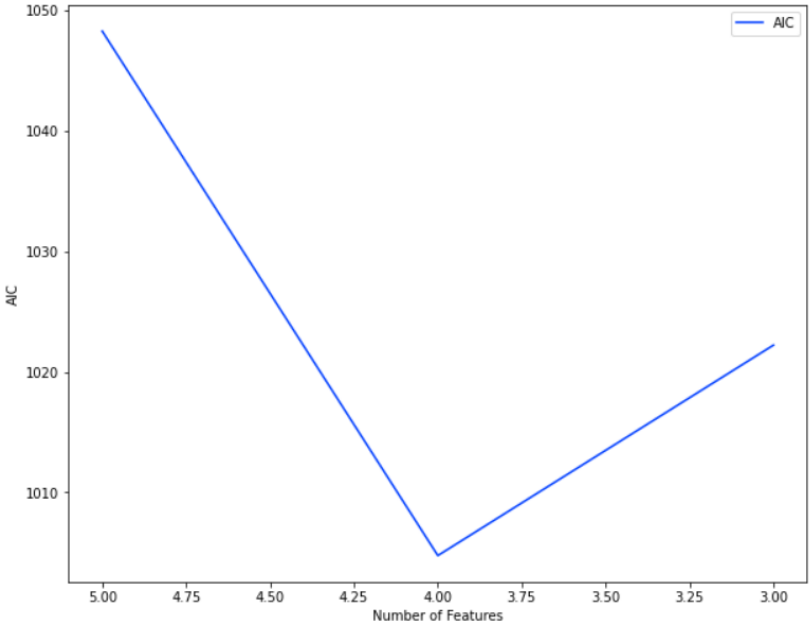
'acceleration', 'displacement', 'model-year', 'weight', and 'cylinders' were incorporated after the forward selection process found them most useful.





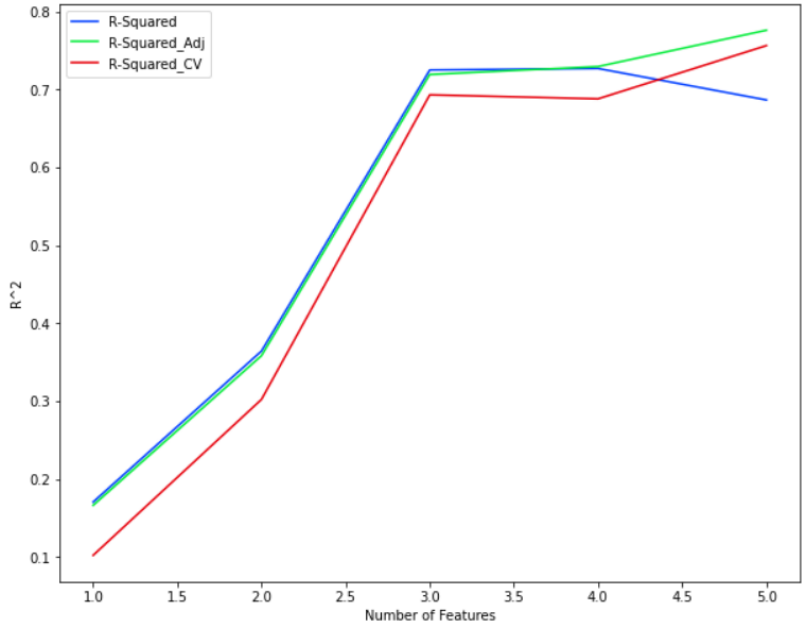
Backward Best Features:
'cylinders', 'weight', and 'model-year' were the best performing variables.

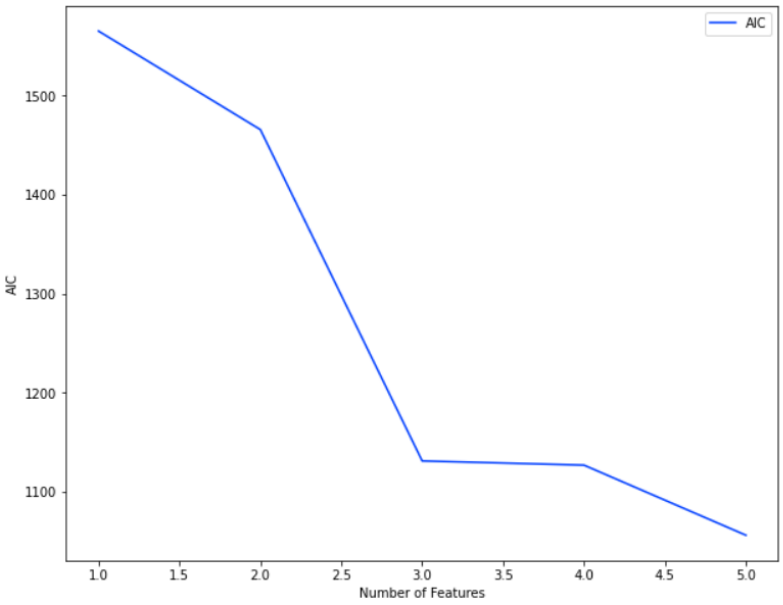




Step Best Features:

'Acceleration', 'displacement', 'model-year', 'horsepower', and 'weight' emerged from the stepwise selection process.





Regularization:

<----------L1 Regularization---------->

R-Squared = 0.7927879989147186, Adjusted R-Squared = 0.7678959061702093, R-Squared CV = 0.7188553995461511, AIC = 1163.489573955536

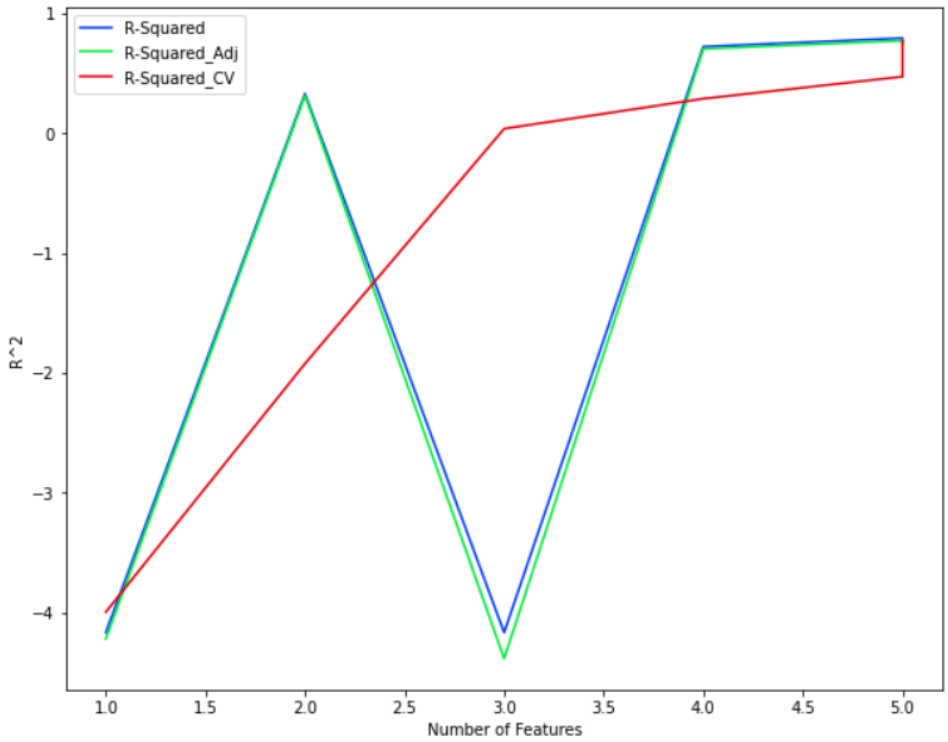
<----------L2 Regularization---------->

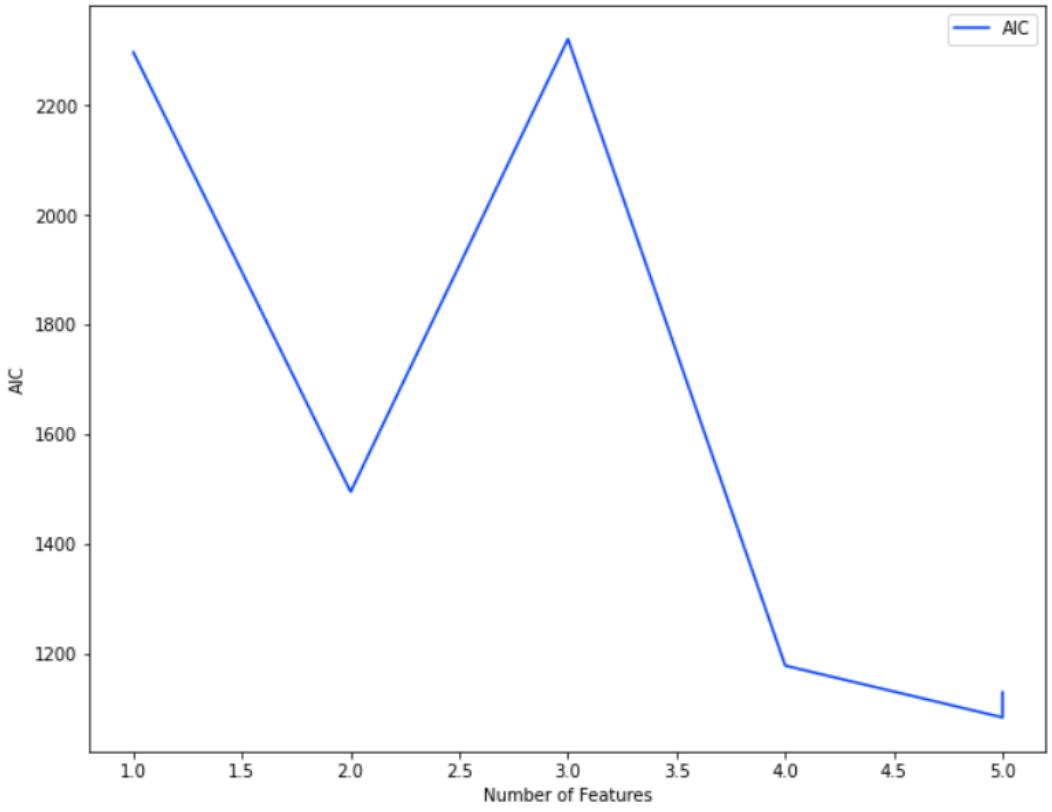
R-Squared = 0.6885439455509186, Adjusted R-Squared = 0.7918968462791198, R-Squared CV = 0.765721457105672, AIC = 1034.3713726997375

NN3l:

Forward

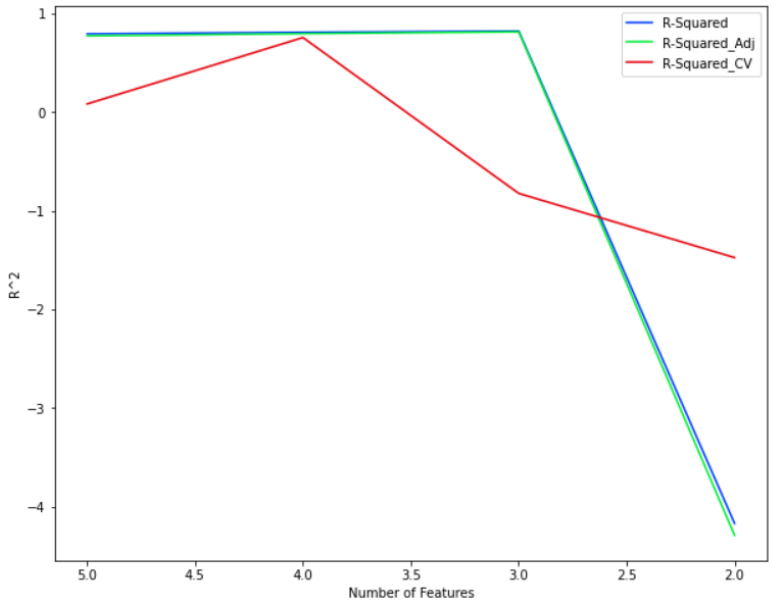
['cylinders', 'acceleration', 'displacement', 'model-year', 'weight']

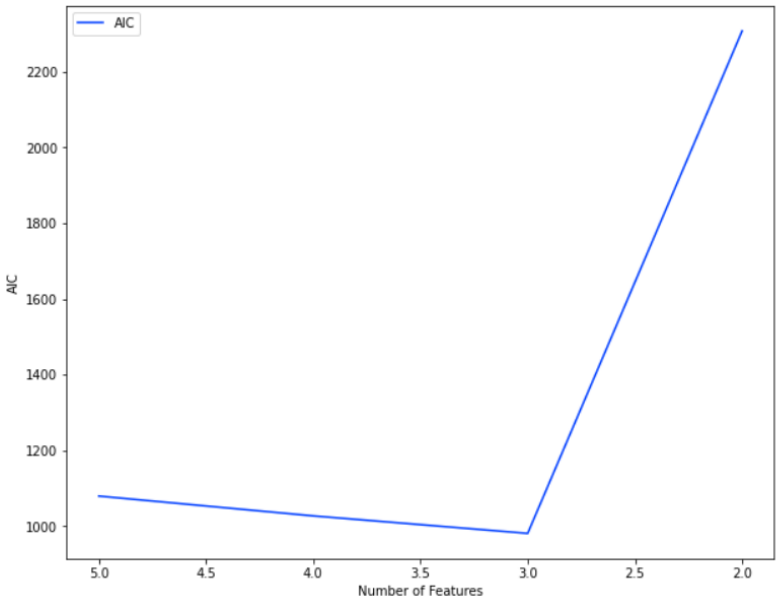




Backward

['weight', 'model-year']





Stepwise

Regularization

<----------L1 Regularization---------->

R-Squared = 0.8446444571018219, Adjusted R-Squared = 0.8275200044629217, R-Squared CV = 0.32997333109970883, AIC = 994.8323588371277

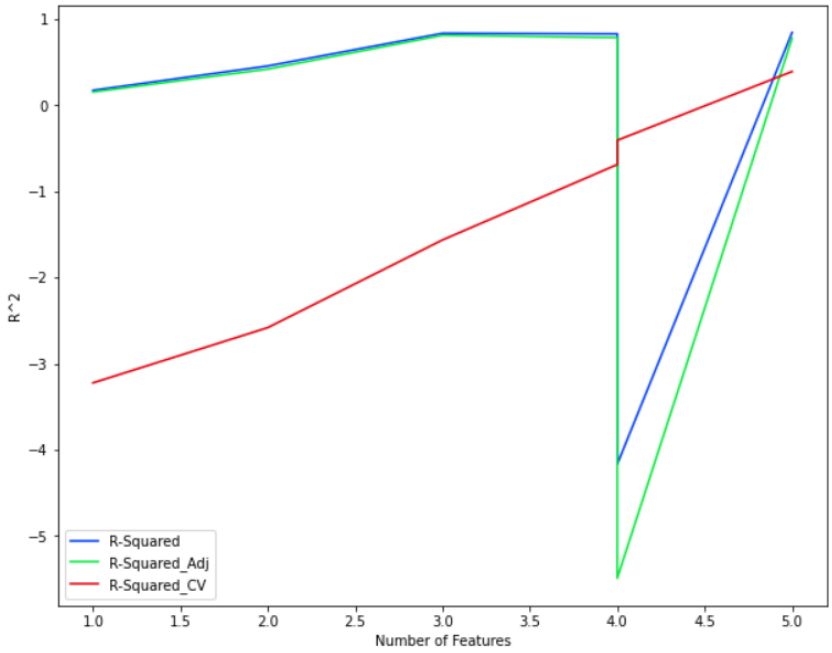
<----------L2 Regularization---------->

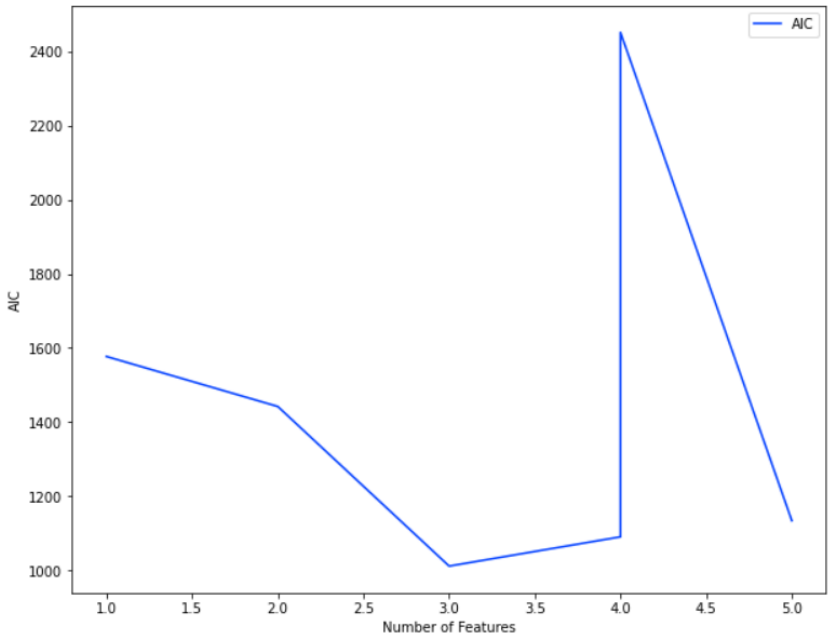
R-Squared = 0.7849325388669968, Adjusted R-Squared = 0.7215868973269545, R-Squared CV = -0.41071297599974255, AIC = 1108.895218372345

NNXL

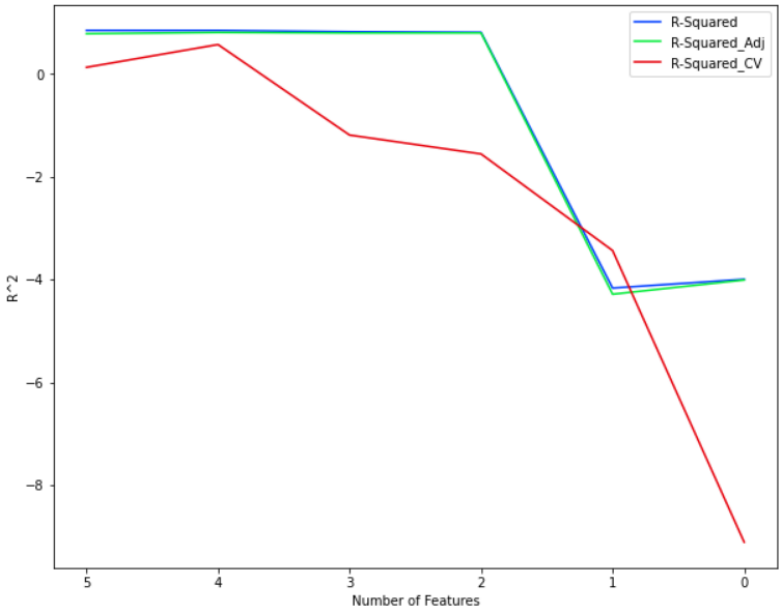
Forward

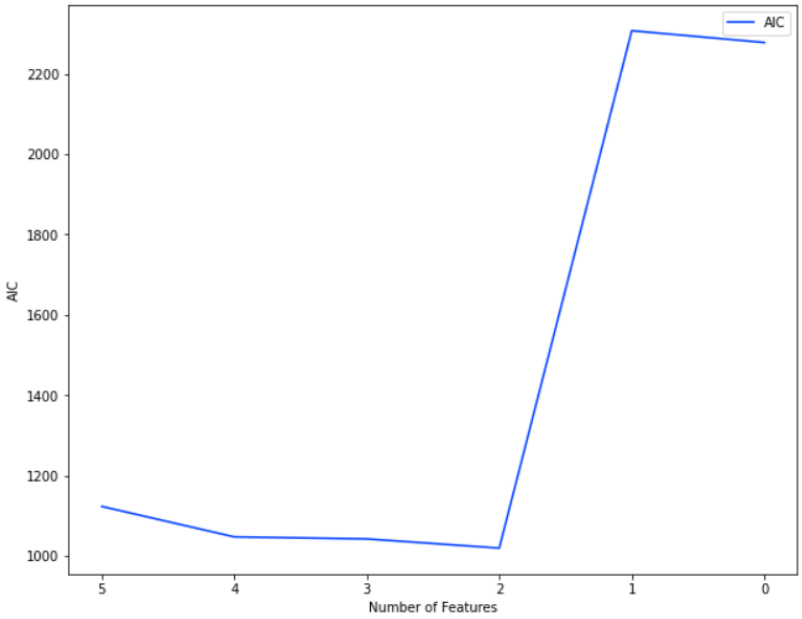
['acceleration', 'weight', 'model-year', 'displacement', 'horsepower']


R^2
Number of Features
R-Squared
R-Squared_Adj
R-Squared_CV


AIC
Number of Features
AIC

Backward


R-Squared
R-Squared_Adj
R-Squared_CV
R^2
0
−2
−4
−6
−8
5
4
3
2
1
0
Number of Features


AIC
AIC
2200
2000
1800
1600
1400
1200
1000
5
4
3
2
1
0
Number of Features

Stepwise

Regularization

<----------L1 Regularization---------->

R-Squared = 0.847817599773407, Adjusted R-Squared = 0.7327745416969584, R-Squared CV = -0.18143512500941372, AIC = 1207.9067730903625

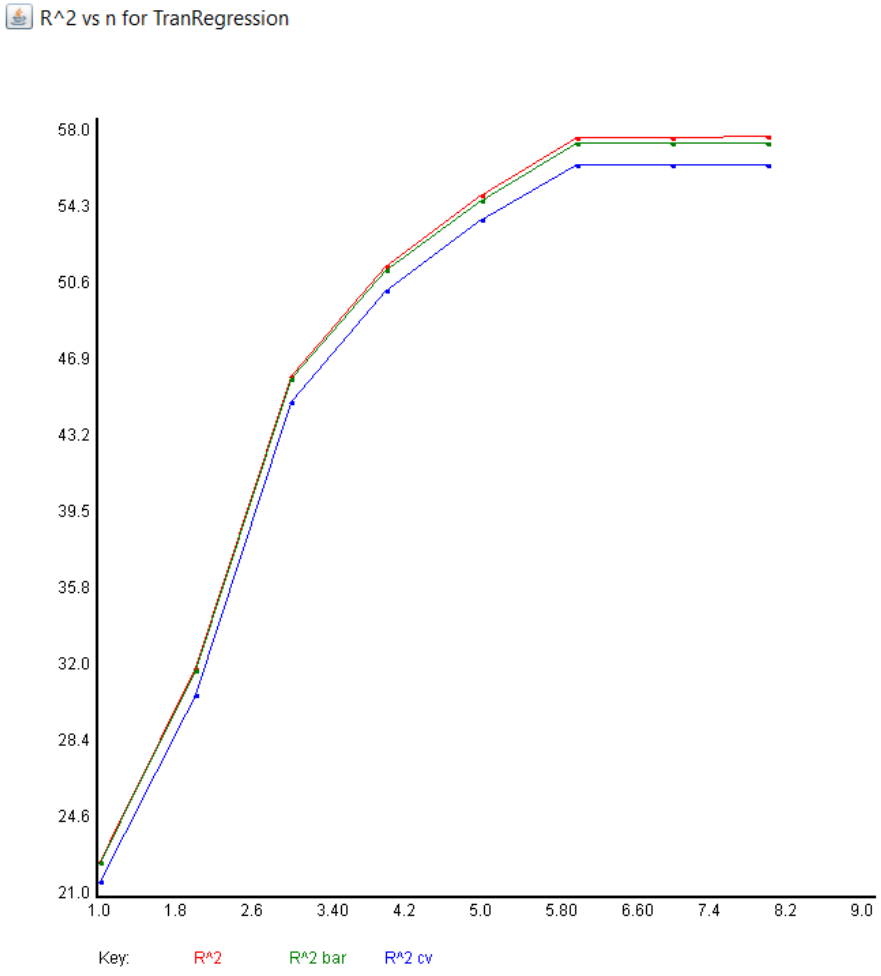
<----------L2 Regularization---------->

R-Squared = 0.8273974806070328, Adjusted R-Squared = 0.71913585851067, R-Squared CV = 0.8202865875420221, AIC = 1216.2821016311646

Dataset: Concrete

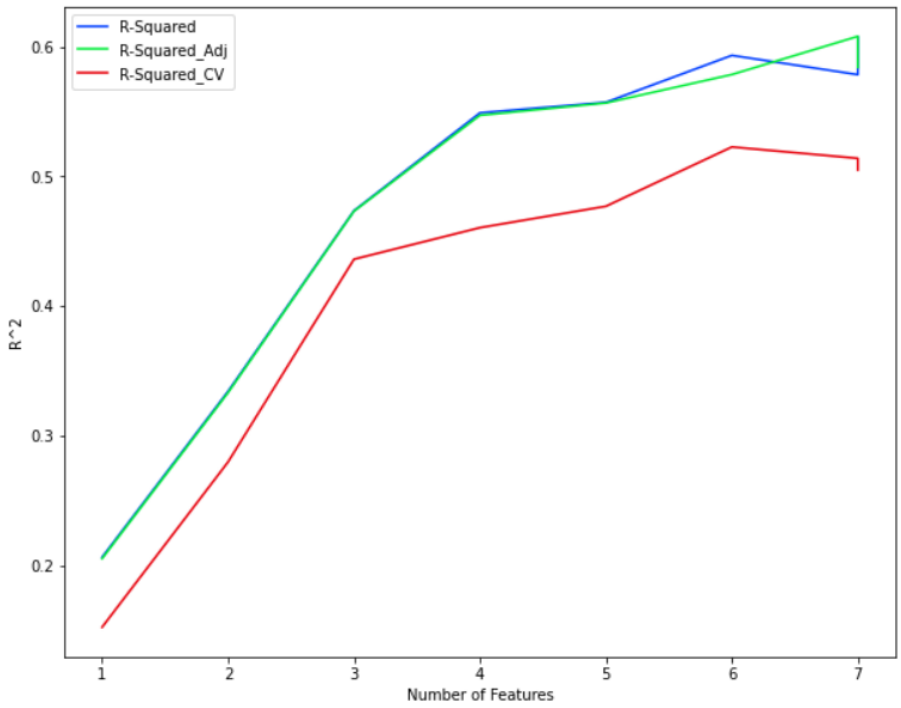
The concrete dataset has 9 predictors and 1030 instances. The dependent variable we were trying to predict was the strength of the concrete.

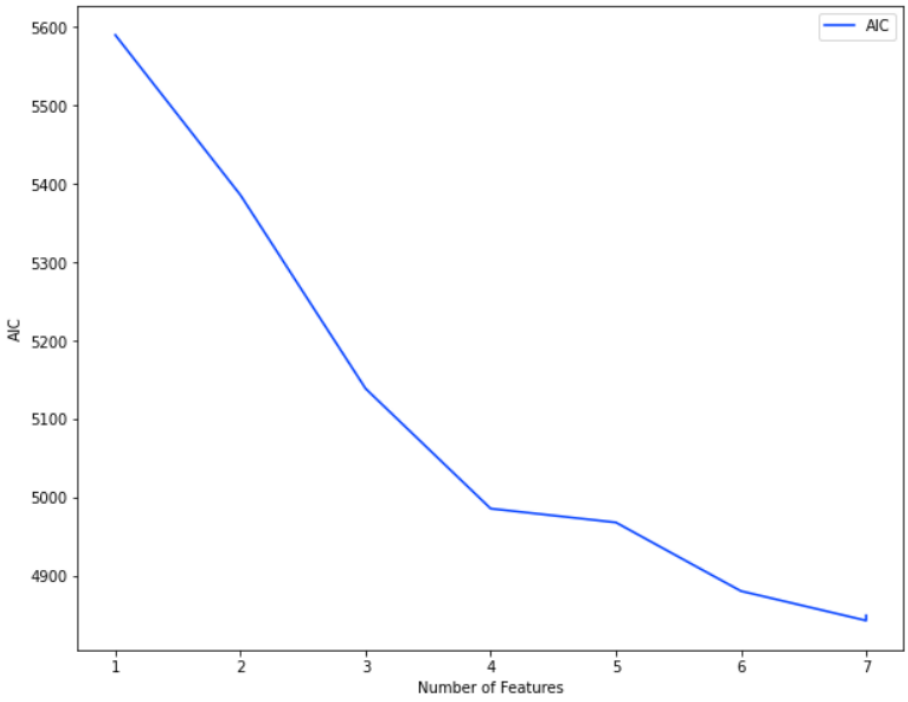
Transformed Regression:



Perceptron:

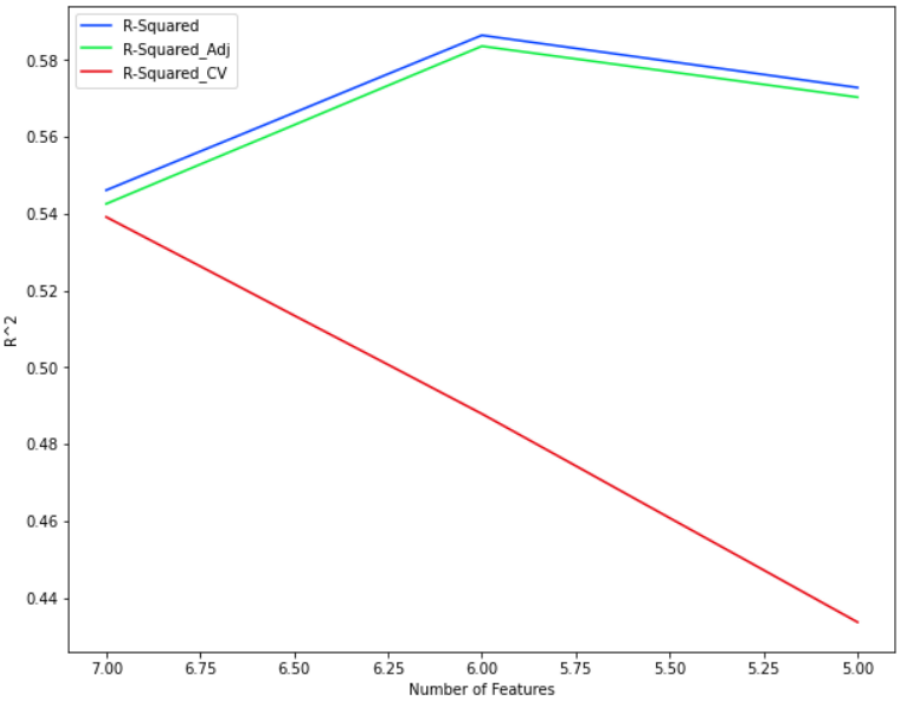
Forward Best Features:
'Cement', 'Superplasticizer', 'Age', 'Blast Furnace Slag', 'Fly Ash ', 'Water', and 'Coarse Aggregate' were found to be the best predictors.

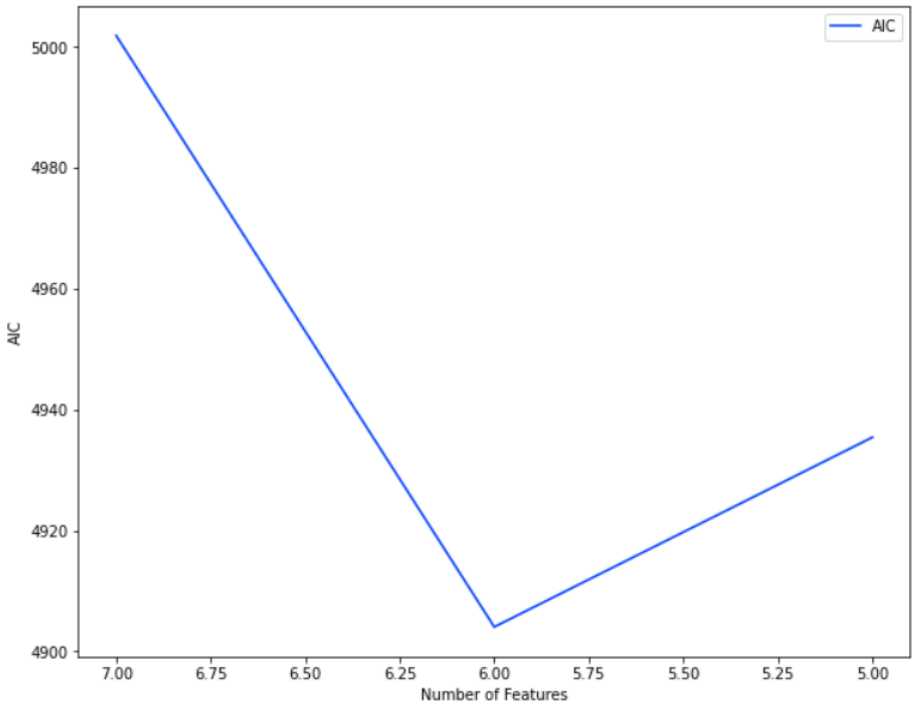




Backward Best Features:

'Cement', 'Blast Furnace Slag', 'Fly Ash', 'Water', and 'Age' were found to be best predictors after using backwards selection.





Regularization:

<----------L1 Regularization---------->

R-Squared = 0.6053773462772369, Adjusted R-Squared = 0.609922301418641, R-Squared CV = 0.5157192948217467, AIC = 4845.312159538269

<----------L2 Regularization---------->

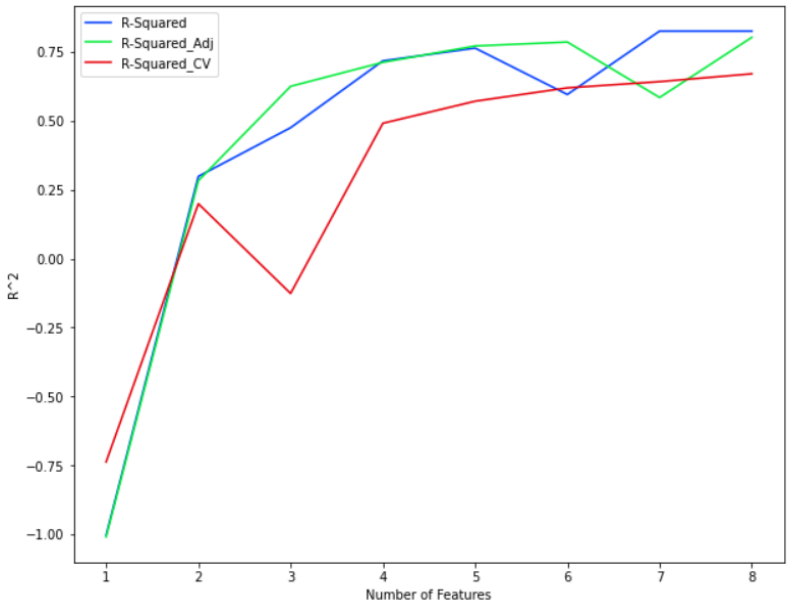
R-Squared = 0.6130619645118713, Adjusted R-Squared = 0.6098895603243042, R-Squared CV = 0.525397651704403, AIC = 4923.562419891357

NN3L:

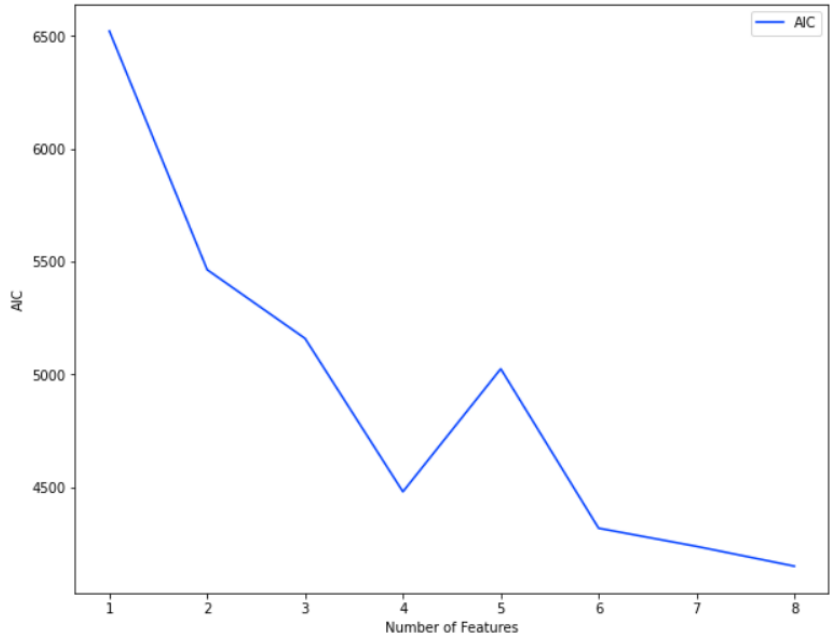
Forward

['Age (day)', 'Superplasticizer (component 5)(kg in a m^3 mixture)', 'Cement (component 1)(kg in a m^3 mixture)', 'Blast Furnace Slag (component 2)(kg in a m^3 mixture)', 'Fly Ash (component 3)(kg in a m^3 mixture)', 'Water (component 4)(kg in a m^3 mixture)', 'Coarse

Aggregate (component 6)(kg in a m^3 mixture)', 'Fine Aggregate (component 7)(kg in a m^3 mixture)']

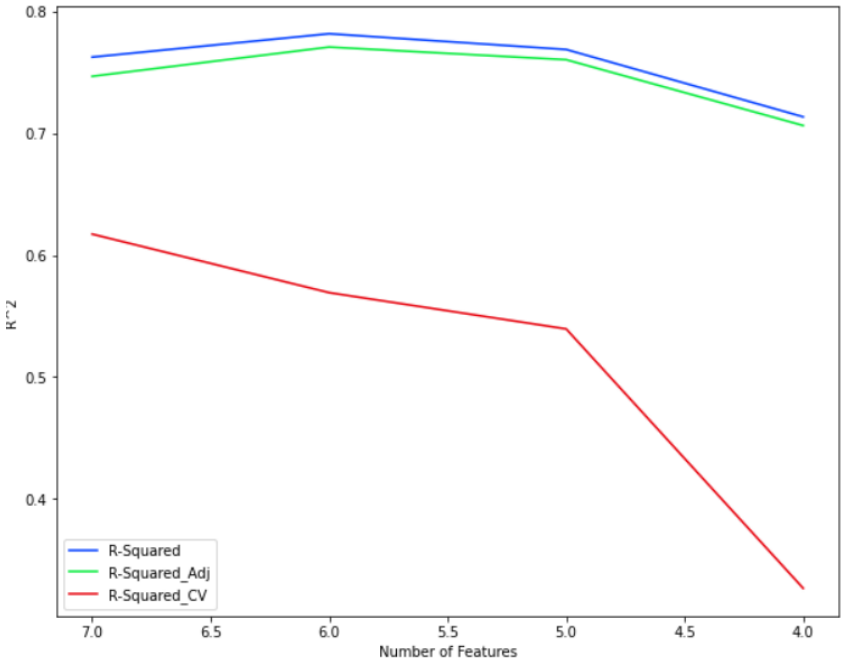


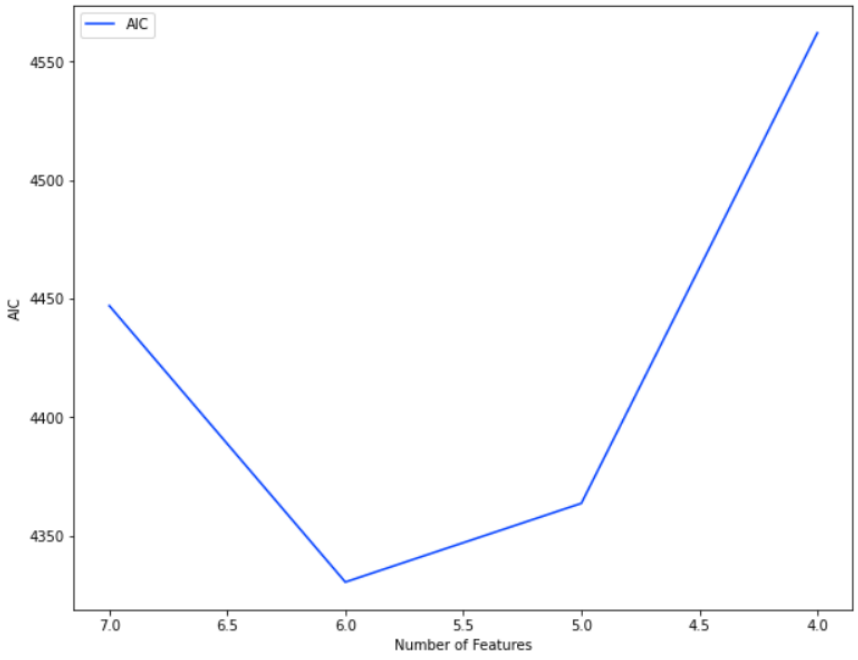
<Figure size 432x288 with 0 Axes>



Backward

['Cement (component 1)(kg in a m^3 mixture)', 'Blast Furnace Slag (component 2)(kg in a m^3 mixture)', 'Water (component 4)(kg in a m^3 mixture)', 'Age (day)']


0.8
0.7
0.6
0.5
0.4
R^2
R-Squared
R-Squared_Adj
R-Squared_CV
7.0
6.5
6.0
5.5
5.0
4.5
4.0
Number of Features


AIC
4550
4500
4450
4400
4350
AIC
7.0
6.5
6.0
5.5
5.0
4.5
4.0
Number of Features

Stepwise

Regularization:

<----------L1 Regularization---------->

R-Squared = 0.8197339624166489, Adjusted R-Squared = 0.8143434629221505, R-Squared CV = 0.58145311858826, AIC = 4247.780644416809

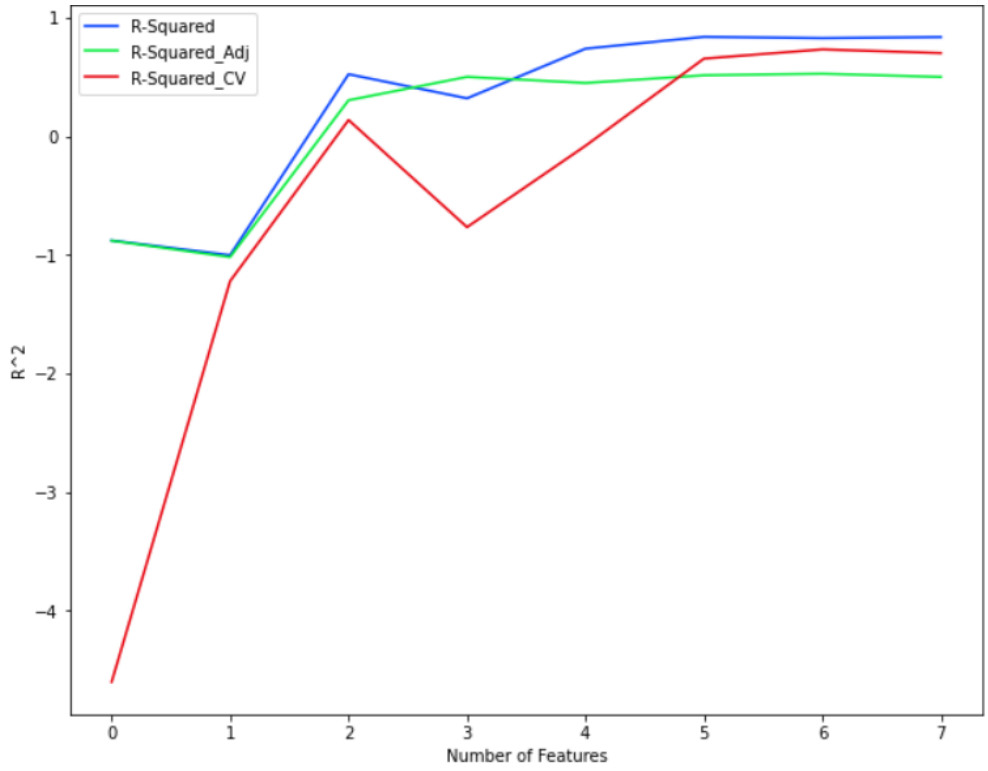
<----------L2 Regularization---------->

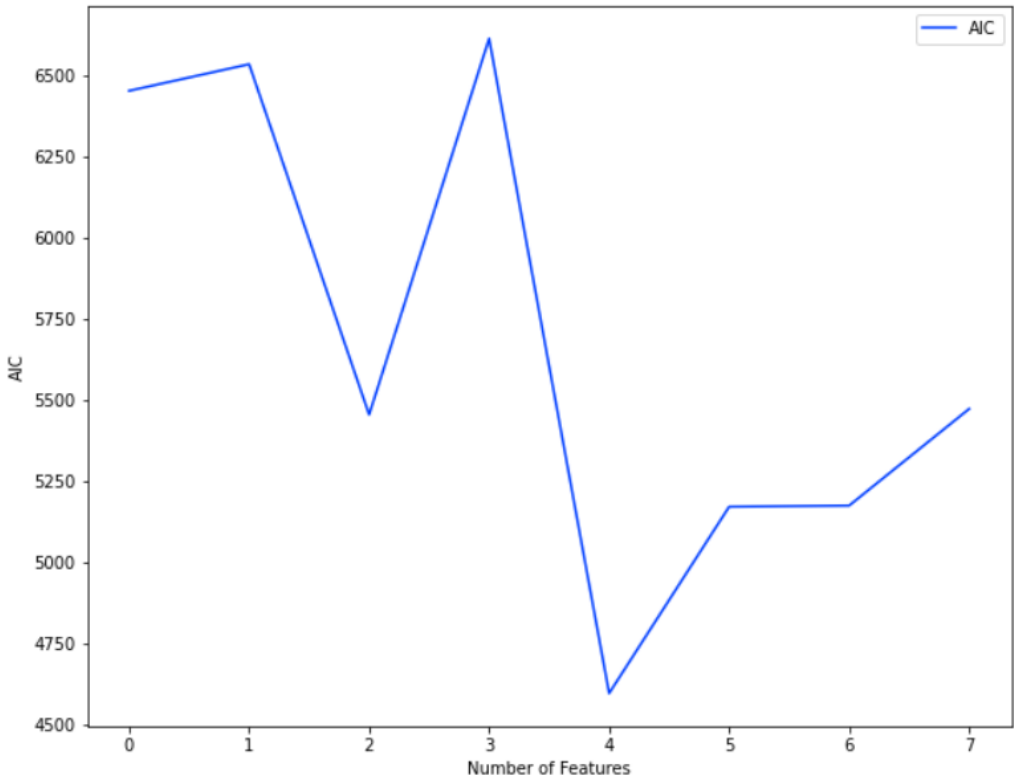
R-Squared = 0.7714501768350601, Adjusted R-Squared = 0.814815463144568, R-Squared CV = 0.6365922566648712, AIC = 4253.533639431

NNXL:

Forward

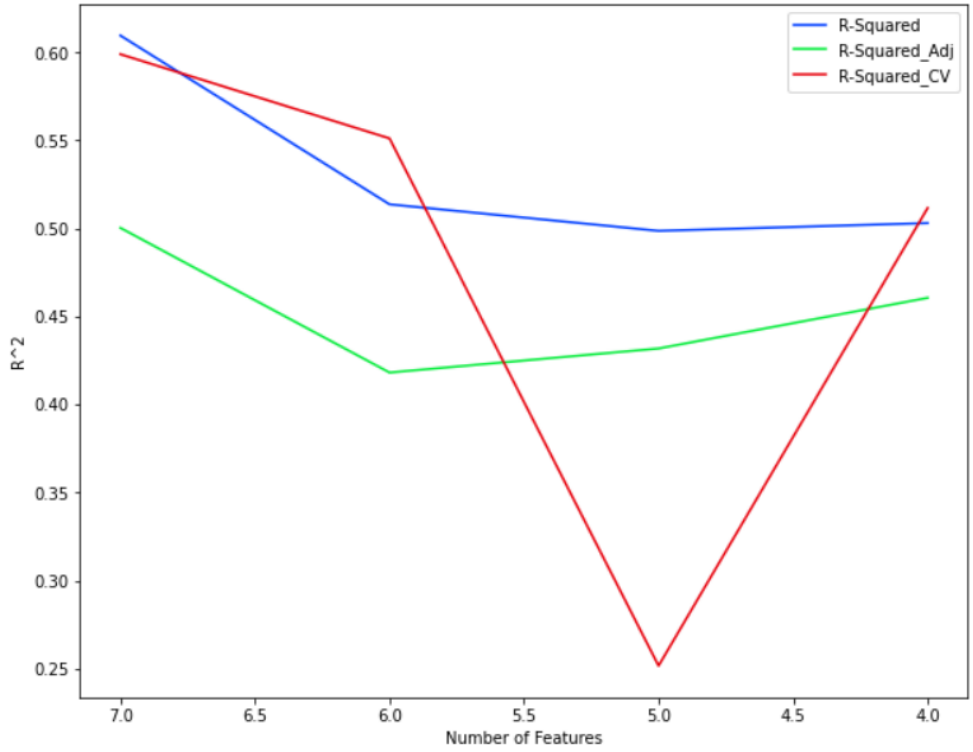
['Cement (component 1)(kg in a m^3 mixture)', 'Age (day)', 'Fly Ash (component 3)(kg in a m^3 mixture)', 'Blast Furnace Slag (component 2)(kg in a m^3 mixture)', 'Water (component 4)(kg in a m^3 mixture)', 'Fine Aggregate (component 7)(kg in a m^3 mixture)', 'Superplasticizer (component 5)(kg in a m^3 mixture)']

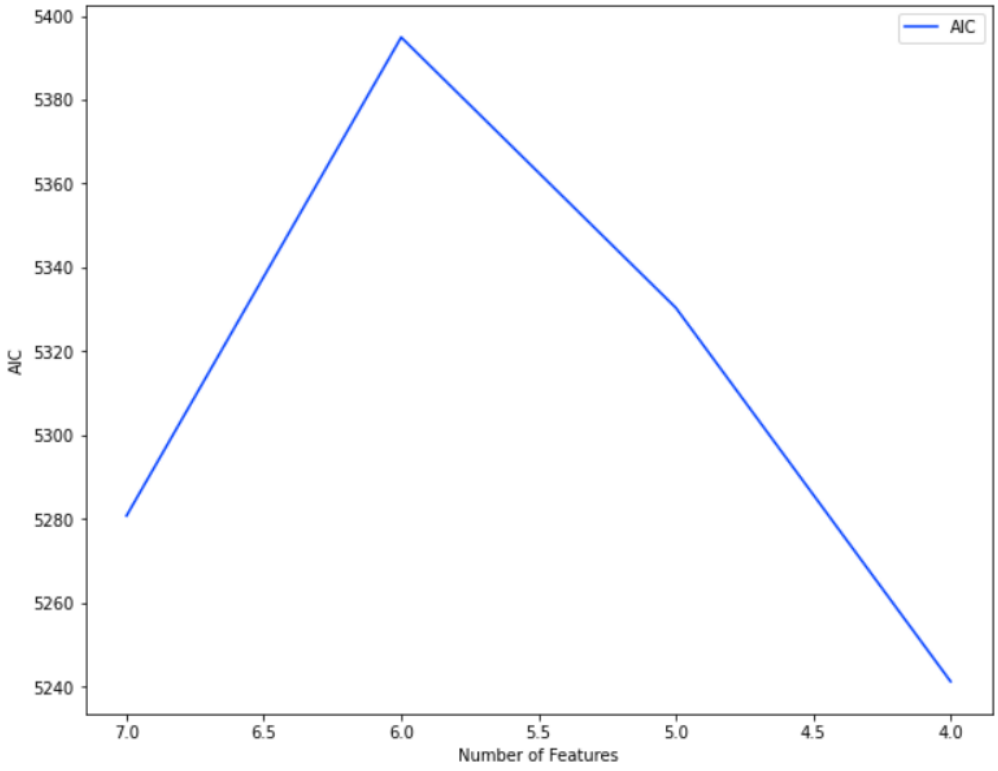

R-Squared
R-Squared_Adj
R-Squared_CV
R^2
Number of Features


AIC
AIC
Number of Features

Backward

['Cement (component 1)(kg in a m^3 mixture)', 'Blast Furnace Slag (component 2)(kg in a m^3 mixture)', 'Fly Ash (component 3)(kg in a m^3 mixture)', 'Age (day)']





Stepwise

Regularization

<----------L1 Regularization---------->

R-Squared = 0.847817599773407, Adjusted R-Squared = 0.7327745416969584, R-Squared CV = -0.18143512500941372, AIC = 1207.9067730903625

<----------L2 Regularization---------->

R-Squared = 0.8273974806070328, Adjusted R-Squared = 0.71913585851067, R-Squared CV = 0.8202865875420221, AIC = 1216.2821016311646

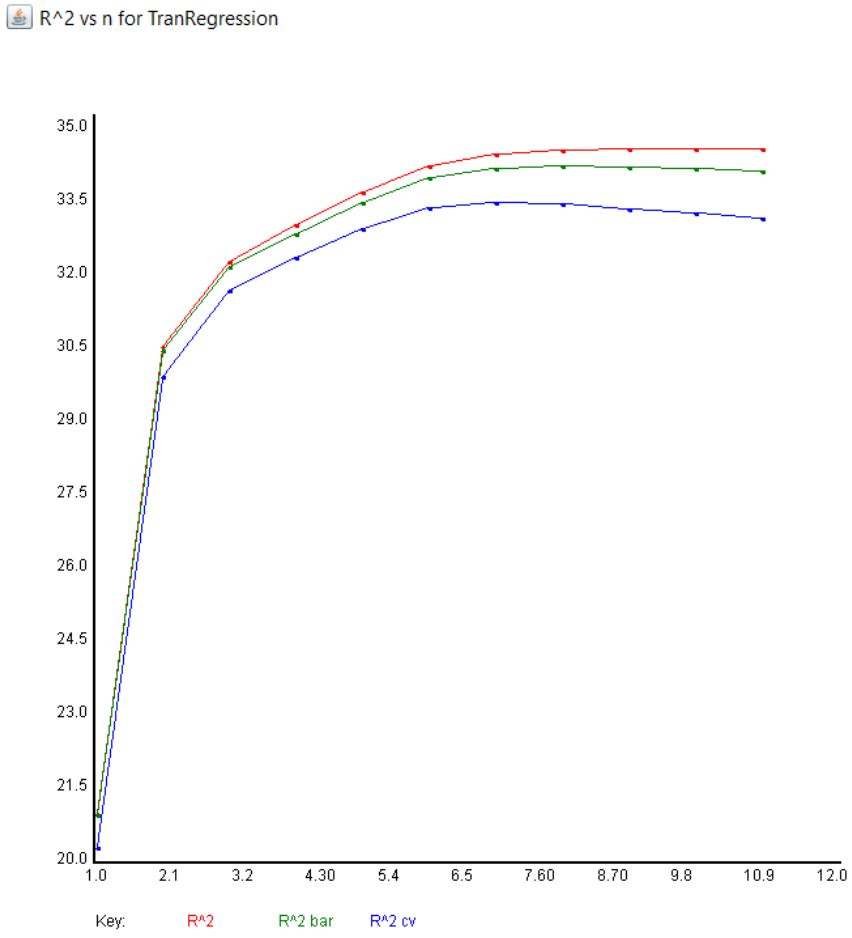
Wine

Dataset: Red Wine Quality

This dataset takes in physical properties of red wines, and attempts to predict their quality ratings.

Perceptron:

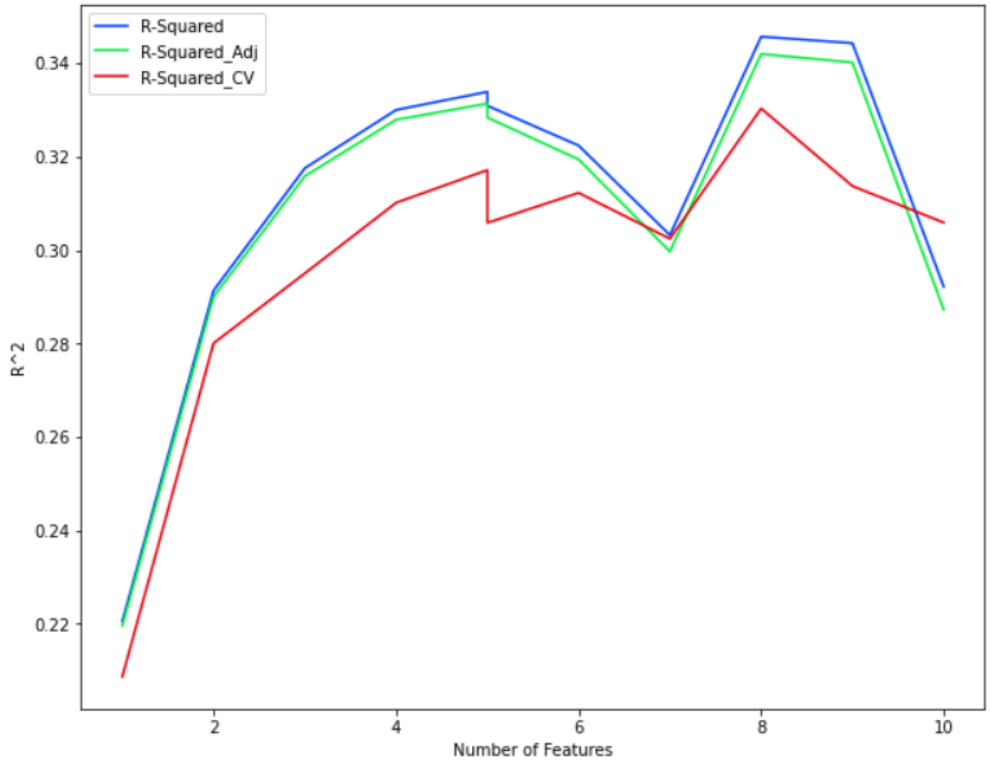
Transformed Regression:

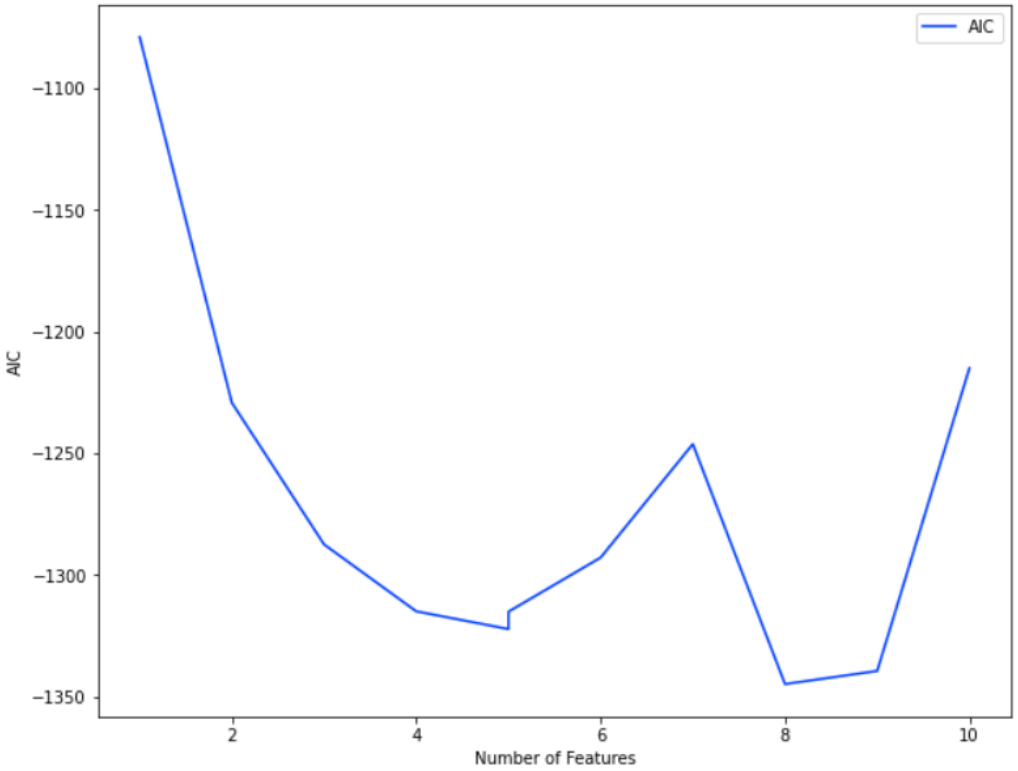


Perceptron

Forward Best Features:

'alcohol', 'volatile acidity', 'sulphates', 'density', 'total sulfur dioxide', 'free sulfur dioxide', 'fixed acidity', 'chlorides', 'citric acid', 'pH' were all used after the forward selection process.





Backward Best Features:

Same as forward.

Step Best Features:
Same as forward as well.

Regularization

<----------L1 Regularization---------->

R-Squared = 0.32452917098999023, Adjusted R-Squared = 0.31693860381127603, R-Squared CV = 0.2895145477287004, AIC = -1286.621961236

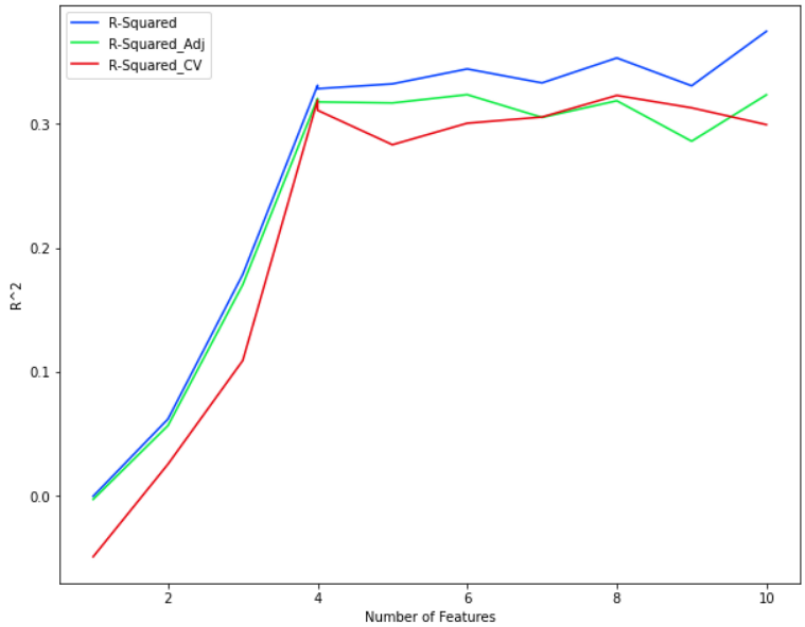
<----------L2 Regularization---------->

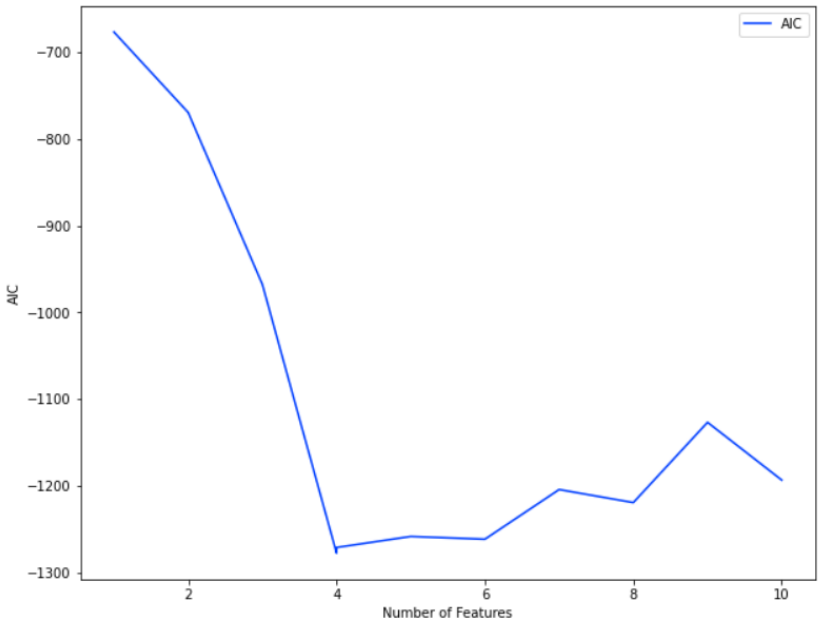
R-Squared = 0.32011914253234863, Adjusted R-Squared = 0.31483479518758095, R-Squared CV = 0.2904179215748426, AIC = -1248.2796224355698

NN3L:

Forward

['sulphates', 'total sulfur dioxide', 'volatile acidity', 'alcohol', 'free sulfur dioxide', 'density', 'pH', 'fixed acidity', 'citric acid', 'residual sugar']





Regularization

<----------L1 Regularization---------->

R-Squared = 0.3330368995666504, Adjusted R-Squared = 0.18442910417729919, R-Squared CV = 0.304811348354157, AIC = -1042.097639143467

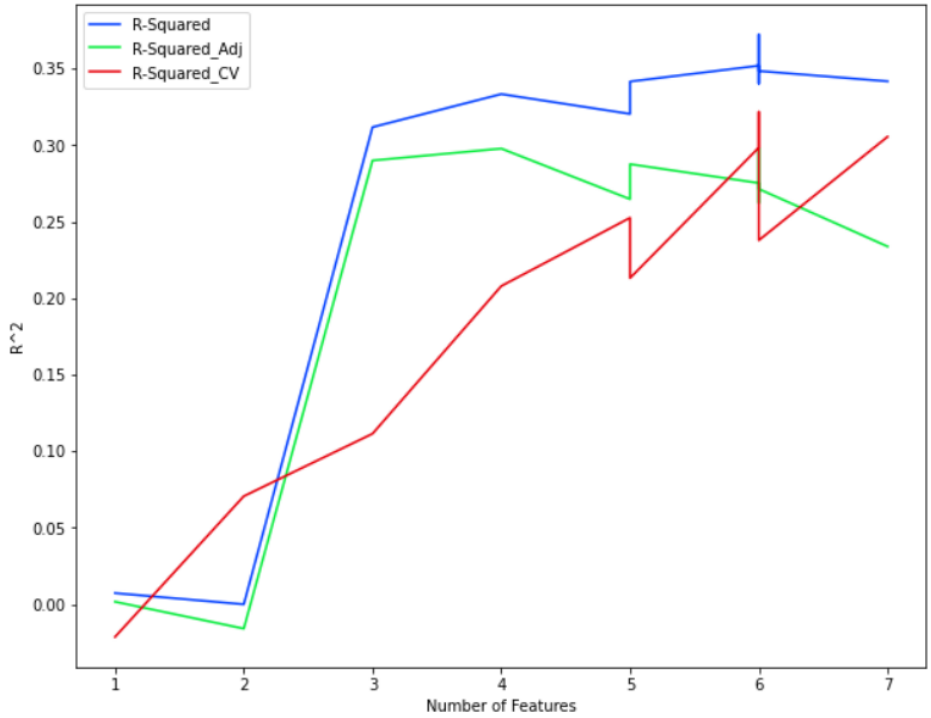
<----------L2 Regularization---------->

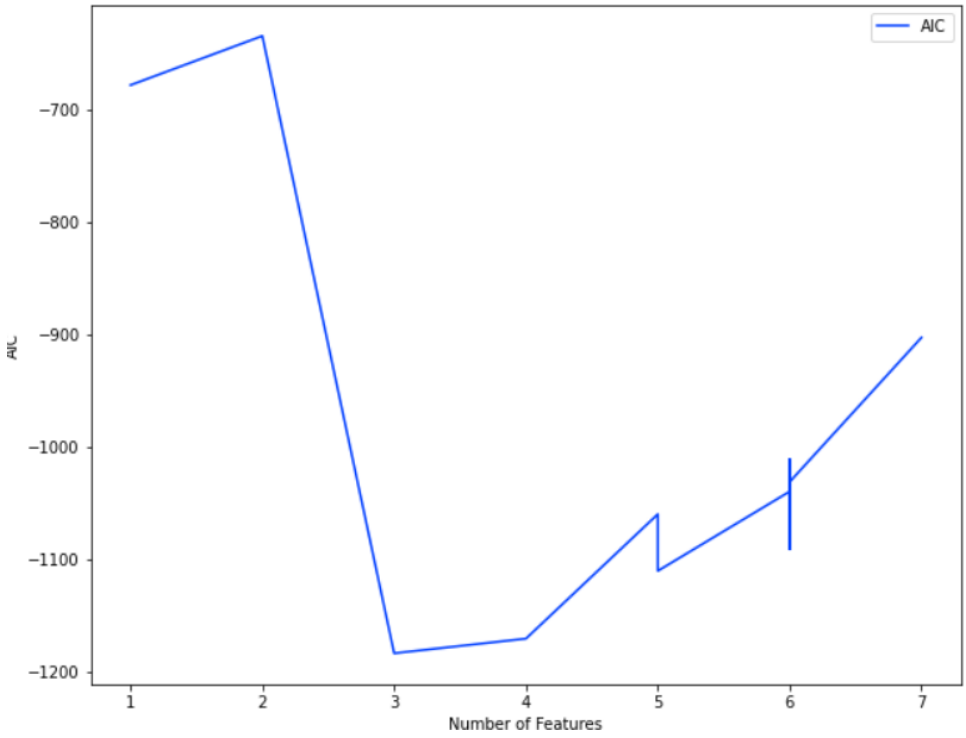
R-Squared = 0.3301594853401184, Adjusted R-Squared = 0.2163644443679873, R-Squared CV = 0.3198029700044748, AIC = -1029.2536704540253

NNXL:

Forward

['chlorides', 'volatile acidity', 'alcohol', 'sulphates', 'free sulfur dioxide', 'total sulfur dioxide', 'density']





Regularization

<----------L1 Regularization---------->

R-Squared = 0.3129616379737854, Adjusted R-Squared = -0.06921965658831986, R-Squared CV = 0.28701772819102733, AIC = -208.13989222049713

<----------L2 Regularization---------->

R-Squared = 0.3210456371307373, Adjusted R-Squared = -0.013257093692740307, R-Squared CV = 0.2740425438906544, AIC = -262.46859407424927

Seoul Bike

Seoul Bike Rental Dataset

This dataset used predictors such as the weather and time to try and predict how many bikes would be rented in a given hour.

Perceptron:

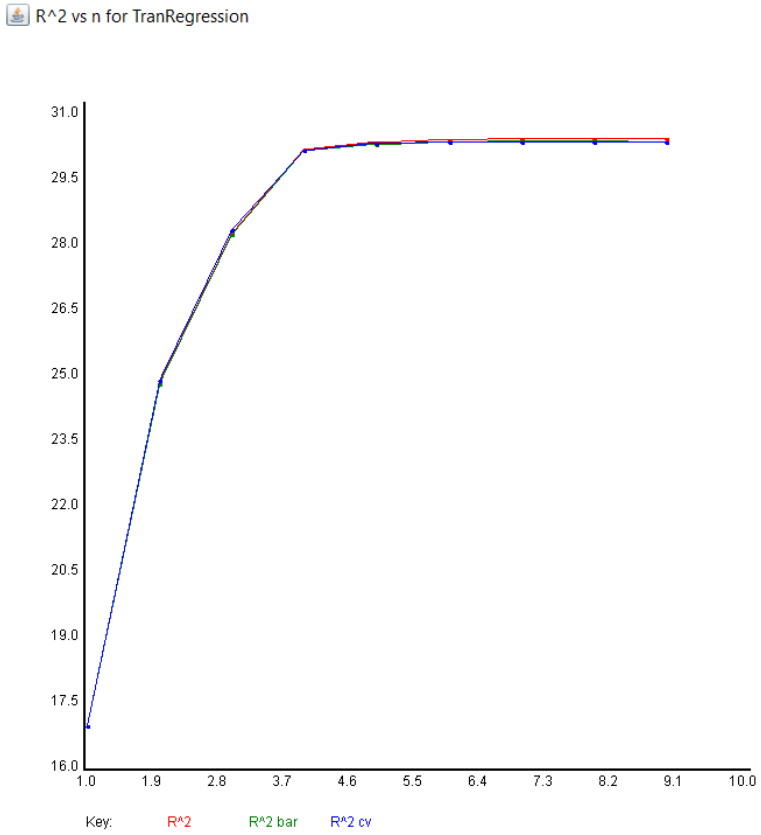
Forward Best Features:

Backward Best Features:

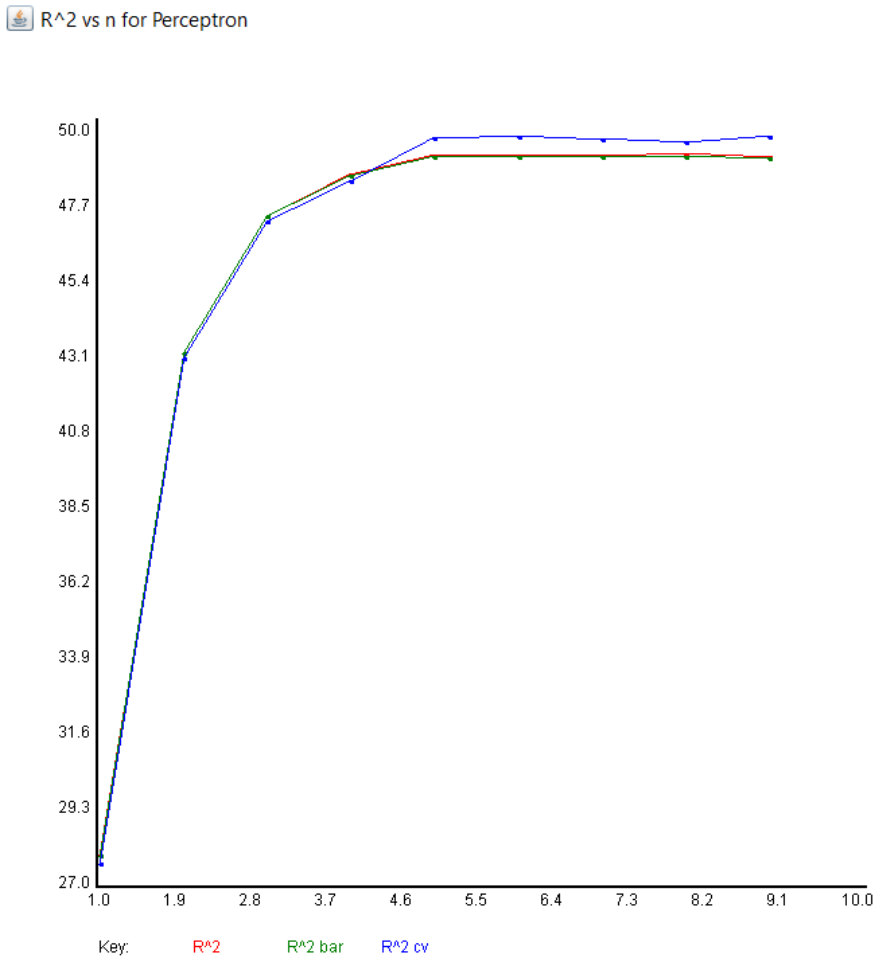
Step Best Features:

3NN:

Transformed Regression:



Perceptron:



Regularization

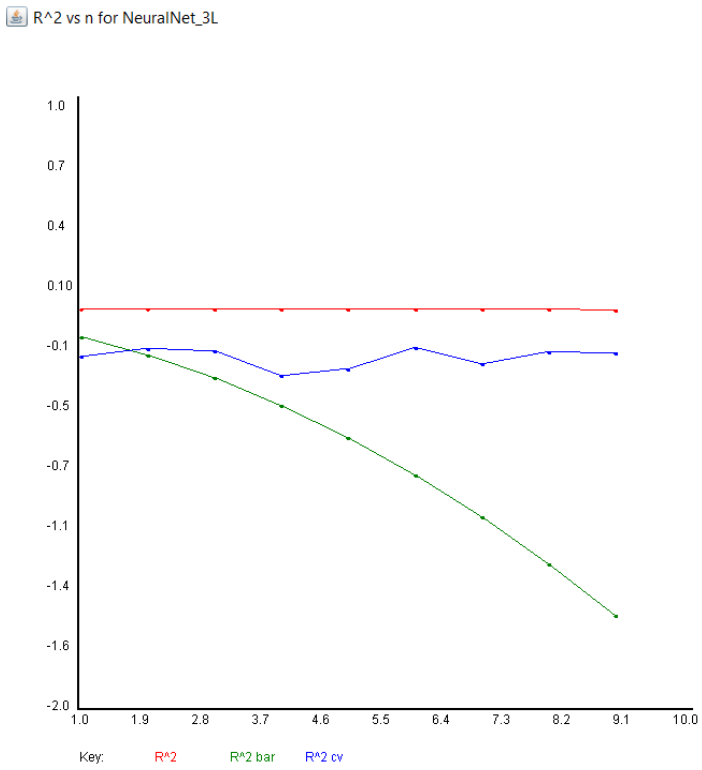
<----------L1 Regularization---------->

R-Squared = 0.45092350244522095, Adjusted R-Squared = 0.45120109100452976, R-Squared CV = 0.40829377947476464, AIC = 108097.93357849121

<----------L2 Regularization---------->

R-Squared = 0.4517557621002197, Adjusted R-Squared = 0.45104080993852036, R-Squared CV = 0.4083505776161759, AIC = 108099.55429077148

NN3L:



Regularization

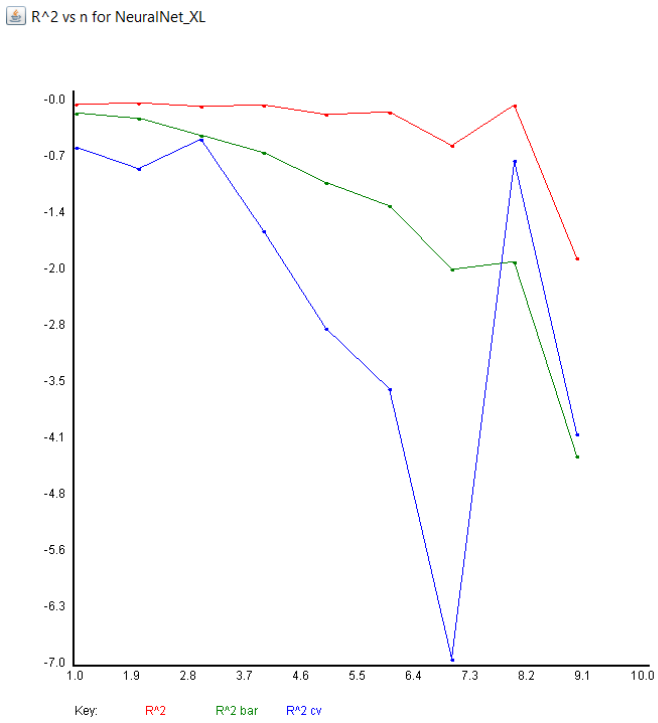
<----------L1 Regularization---------->

R-Squared = 0.5186724960803986, Adjusted R-Squared = 0.6056400701525801, R-Squared CV = 0.4380344128147966, AIC = 106248.20030212402

<----------L2 Regularization---------->

R-Squared = 0.5168326497077942, Adjusted R-Squared = 0.5446029041990257, R-Squared CV = 0.46796157739643574, AIC = 105717.5089263916

NNXL:



Regularization

<----------L1 Regularization---------->

R-Squared = 0.49360716342926025, Adjusted R-Squared = 0.5984309333886496, R-Squared CV = 0.46753312113693946, AIC = 106658.31267547607

<----------L2 Regularization---------->

R-Squared = 0.5148705840110779, Adjusted R-Squared = 0.5823938791973019, R-Squared CV = 0.5380710826617485
AIC = 107274.8433227539

Electric Grid Stability Dataset

Perceptron:

Forward Best Features:

Backward Best Features:

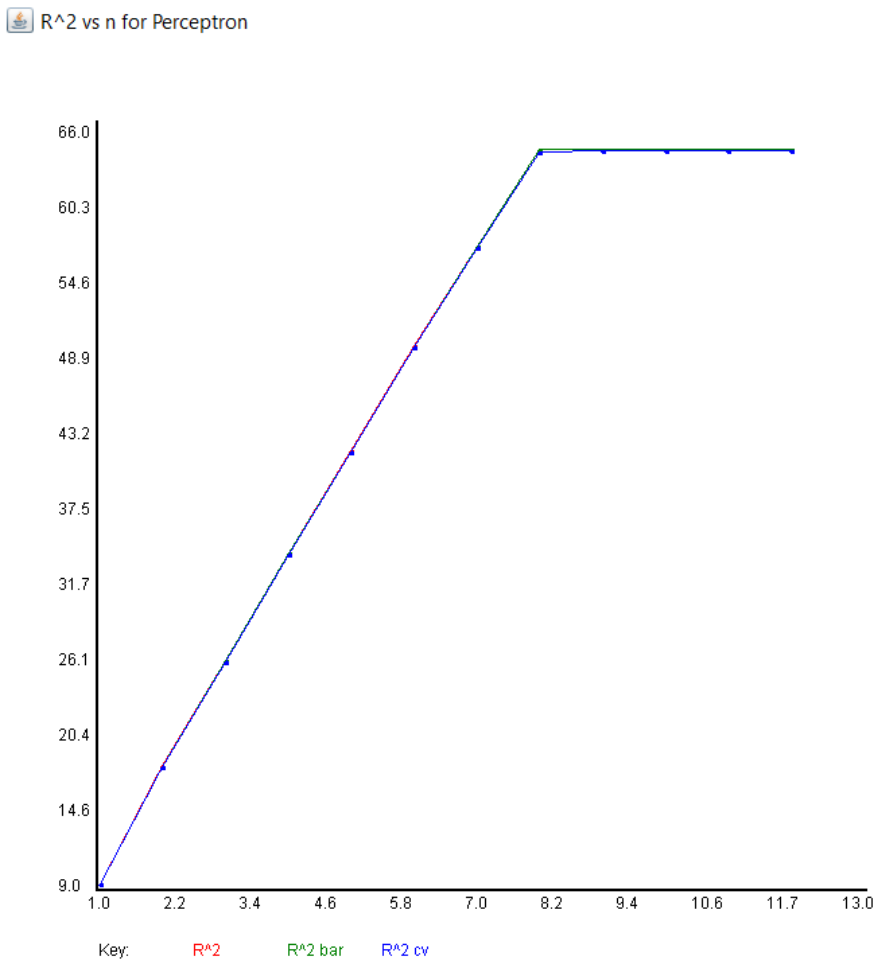
Step Best Features:

3NN:

Transformed Regression

(Could not run transformed regression on the electricity grid dataset as it contains negative y values)

Perceptron:



Regularization

<----------L1 Regularization---------->

R-Squared = 0.26146721839904785, Adjusted R-Squared = 0.2763282576329288, R-Squared CV = 0.505449131462778, AIC = -66471.68352508545

<----------L2 Regularization---------->

R-Squared = 0.6215843558311462, Adjusted R-Squared = 0.6019865755725189, R-Squared CV = 0.4886558408444833, AIC = -72222.44932556152

NN3L

Forward

Backward

Stepwise

Regularization

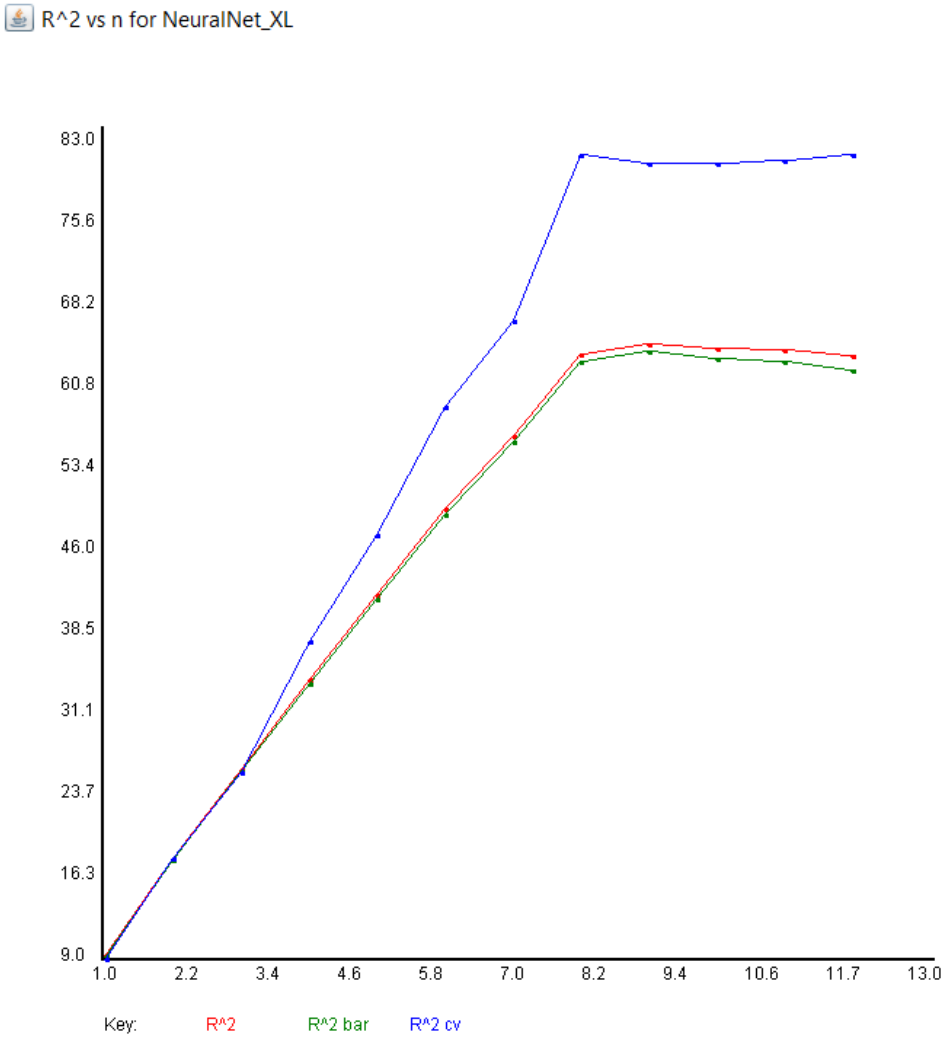
<----------L1 Regularization---------->

R-Squared = -0.007398843765258789, Adjusted R-Squared = -0.019715178485804286, R-Squared CV = 0.8801329702042532, AIC = -65631.59590911865

<----------L2 Regularization---------->

R-Squared = 0.3213427662849426, Adjusted R-Squared = 0.31872646957181117, R-Squared CV = 0.8843687548261558, AIC = -69074.43171691895

NNXL:



Regularization

<----------L1 Regularization---------->
R-Squared = -9.179115295410156e-06, Adjusted R-Squared = -0.07940808585235914,
R-Squared CV = 0.9507357616122436, AIC = -64709.40113067627

<----------L2 Regularization---------->

R-Squared = -0.005037188529968262, Adjusted R-Squared = -0.08534290427105606,
R-Squared CV = 0.9535901389249546, AIC = -64579.0433883667