Data Science Project

Scala Report

Prepared by Group 6

Team members

Bhavana Nare

Vishnu Priya Buggineni

Yaswanth Kumar Palli

Vamshi Krishna B

We have worked on all 6 datasets using both **python** and **scala**.
All the libraries we used in python are mentioned below:
Numpy, Pandas, Matplotlib, Seaborn, Mlxtend, Gplearn, Math, and sklearn

Air Quality Dataset:

We have taken this dataset from UCI repository(We used AirQualityUCI.csv for scala). There are 9358 occurrences of hourly averaged responses from an array of 5 metal oxide chemical sensors integrated in an Air Quality Chemical Multisensor Device in this dataset. This dataset contains 12 attributes, the information about each attribute is mentioned below:

Date- format (DD/MM/YYYY)

Time- format (HH.MM.SS)

CO(GT)- True hourly averaged concentration CO

PT08.S1(CO)- (tin oxide) hourly averaged sensor response

NMHC(GT)- True hourly averaged overall Non Metanic HydroCarbons concentration

C6H6(GT)- True hourly averaged Benzene concentration

PT08.S2(NMHC)- (titania) hourly averaged sensor response

NOx(GT)- True hourly averaged NOx concentration

PT08.S3(NOx)- (tungsten oxide) hourly averaged sensor response

NO2(GT)- True hourly averaged NO2 concentration—-**Response Variable**

PT08.S4(NO2)- (tungsten oxide) hourly averaged sensor response

PT08.S5(O3) -(indium oxide) hourly averaged sensor response

T- Temperature in °C

RH- Relative Humidity (%)

AH- Absolute Humidity

Scala report:

First we imported all the necessary modules.We used matrixD.load() to load the dataset. The date and time are excluded from the airquality dataset. The airquality dataset contains negative values, So we used imputation to replace the -200 values with mean. The code we used for imputation is mentioned below:
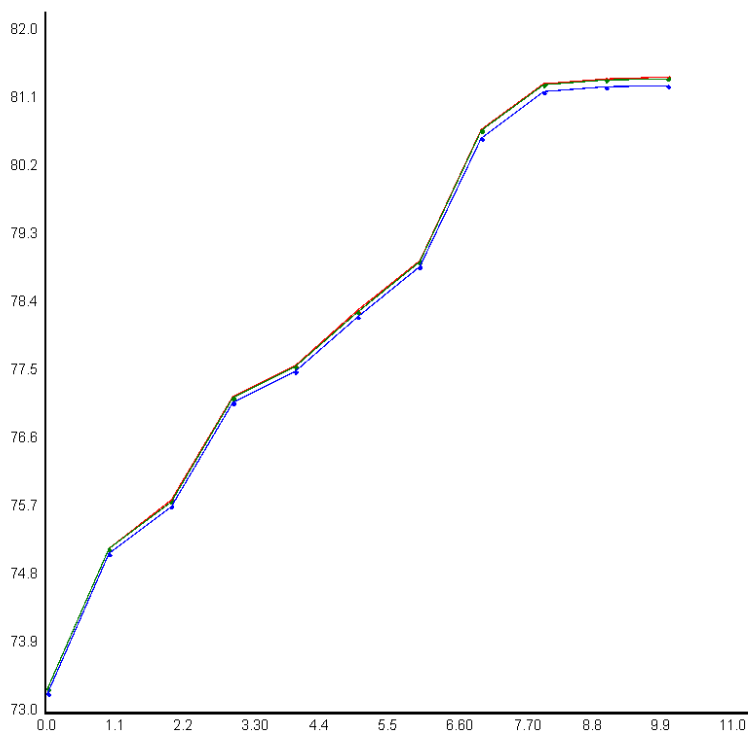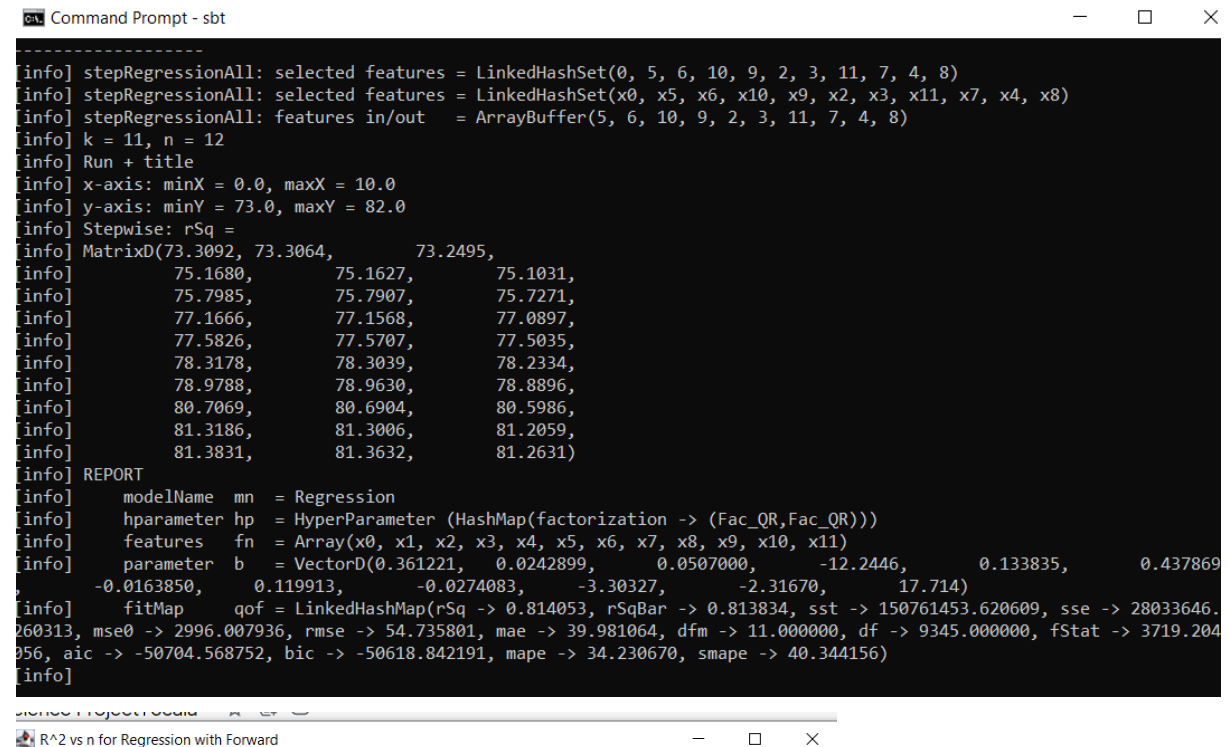
```scala
@main def RegressionTest12 (): Unit =
{
    val airqual = MatrixD.load("AirQualityUCI.csv",1)
    println(airqual.dim)
    println(airqual.dim2)
    val n = airqual.dim2 - 1
    val (x, y) = (airqual.not(?, n), airqual(?, n))
    for (i <- (0 until airqual.dim2) )
        {
        val z = airqual(?, i)
        val x2= x.copy
        val x3 = x.copy
        //var iv = ( -200.000,-200.000)
        val iv = ImputeMean.impute (z)
        println(iv._2)
        println(z.length)

        for (j <- (0 until z.length) ){
            if (airqual(j, i) == -200.000){
                airqual(j, i) = iv._2
            }
        }

    }
    airqual.write("test.csv")
```

Linear Regression:

```
------------------
[info] stepRegressionAll: selected features = LinkedHashSet(0, 5, 6, 10, 9, 2, 3, 11, 7, 4, 8)
[info] stepRegressionAll: selected features = LinkedHashSet(x0, x5, x6, x10, x9, x2, x3, x11, x7, x4, x8)
[info] stepRegressionAll: features in/out   = ArrayBuffer(5, 6, 10, 9, 2, 3, 11, 7, 4, 8)
[info] k = 11, n = 12
[info] Run + title
[info] x-axis: minX = 0.0, maxX = 10.0
[info] y-axis: minY = 73.0, maxY = 82.0
[info] Stepwise: rSq =
[info] MatrixD(73.3092, 73.3064,      73.2495,
[info]         75.1680,       75.1627,       75.1031,
[info]         75.7985,       75.7907,       75.7271,
[info]         77.1666,       77.1568,       77.0897,
[info]         77.5826,       77.5707,       77.5035,
[info]         78.3178,       78.3039,       78.2334,
[info]         78.9788,       78.9630,       78.8896,
[info]         80.7069,       80.6904,       80.5986,
[info]         81.3186,       81.3006,       81.2059,
[info]         81.3831,       81.3632,       81.2631)
[info] REPORT
[info]     modelName   mn  = Regression
[info]     hparameter  hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]     features    fn  = Array(x0, x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11)
[info]     parameter   b   = VectorD(0.361221,   0.0242899,      0.0507000,      -12.2446,      0.133835,       0.437869
,     -0.0163850,     0.119913,      -0.0274083,     -3.30327,     -2.31670,     17.714)
[info]       fitMap    qof = LinkedHashMap(rSq -> 0.814053, rSqBar -> 0.813834, sst -> 150761453.620609, sse -> 28033646.
260313, mse0 -> 2996.007936, rmse -> 54.735801, mae -> 39.981064, dfm -> 11.000000, df -> 9345.000000, fStat -> 3719.204
056, aic -> -50704.568752, bic -> -50618.842191, mape -> 34.230670, smape -> 40.344156)
[info]
```


R^2 vs n for Regression with Forward

R2: 0.814053
Adj R2: 0.813834

Ridge Regression:

Command Prompt - sbt

3.000, 193.000,   185.000,   147.000,   117.000,   112.000,   101.000,   93.0000,   6
.0000, 61.0000,   54.0000,   64.0000,   80.0000,   133.000,   170.000,   139.000,   1
24.000, 118.000,   108.000,   100.000,   91.0000,   89.0000,   102.000,   114.000,   1
24.000, 136.000,   118.000,   93.0000,   72.0000,   69.0000,   61.0000,   50.0000,   3
.0000, 29.0000,   41.0000,   66.0000,   89.0000,   150.000,   150.000,   127.000,   1
20.000, 123.000,   115.000,   122.000,   111.000,   119.000,   137.000,   147.000,   1
64.000, 160.000,   134.000,   110.000,   99.0000,   127.000,   174.000,   127.000,   8
.0000, 70.0000,   49.0000,   44.0000,   107.000,   141.000,   121.000,   103.000,   1
2.000, 95.0000,   93.0000,   95.0000,   80.0000,   81.0000,   106.000,   113.000,   1
7.000, 124.000,   122.000,   87.0000,   79.0000,   93.0000,   102.000,   86.0000,   7
.0000, 88.0000,   64.0000,   46.0000,   55.0000,   84.0000,   89.0000,   118.000,   1
3.000, 115.000,   124.000,   86.0000,   87.0000,   100.000,   132.000,   156.000,   1
8.000, 181.000,   187.000,   158.000,   153.000,   128.000,   93.0000,   58.0000,   5
.0000, 51.0000,   43.0000,   53.0000,   93.0000,   155.000,   174.000,   187.000,   1
0.000, 179.000,   175.000,   156.000,   168.00)
[info] ------------------
[info] | Optimize lambda |
[info] ------------------
[info] findLambda2 = -0.0
[info] REPORT
[info]   modelName mn = RidgeRegression
[info]   hparameter hp = HyperParameter (HashMap(lambda -> (1.0,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky))
[info]   features   fn = Array(x0, x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11)
[info]   parameter b   = VectorD(0.361227,   0.0242961,   0.0507017,   -12.2433,   0.133810,   0.437866
[info]   -0.0163796,   0.119900,   -0.0274074,   -3.30269,   -2.31651,   17.712)
[info]   fitMap   qof = LinkedHashMap(rSq -> 0.814053, rSqBar -> 0.813814, sst -> 150761453.620609, sse -> 28033646.
315129, mse0 -> 2996.007942, rmse -> 54.735801, mae -> 39.981120, dfm -> 12.000000, df -> 9344.000000, fStat -> 3408.905
654, aic -> -50702.568761, bic -> -50609.698321, mape -> 34.230603, smape -> 40.343996)
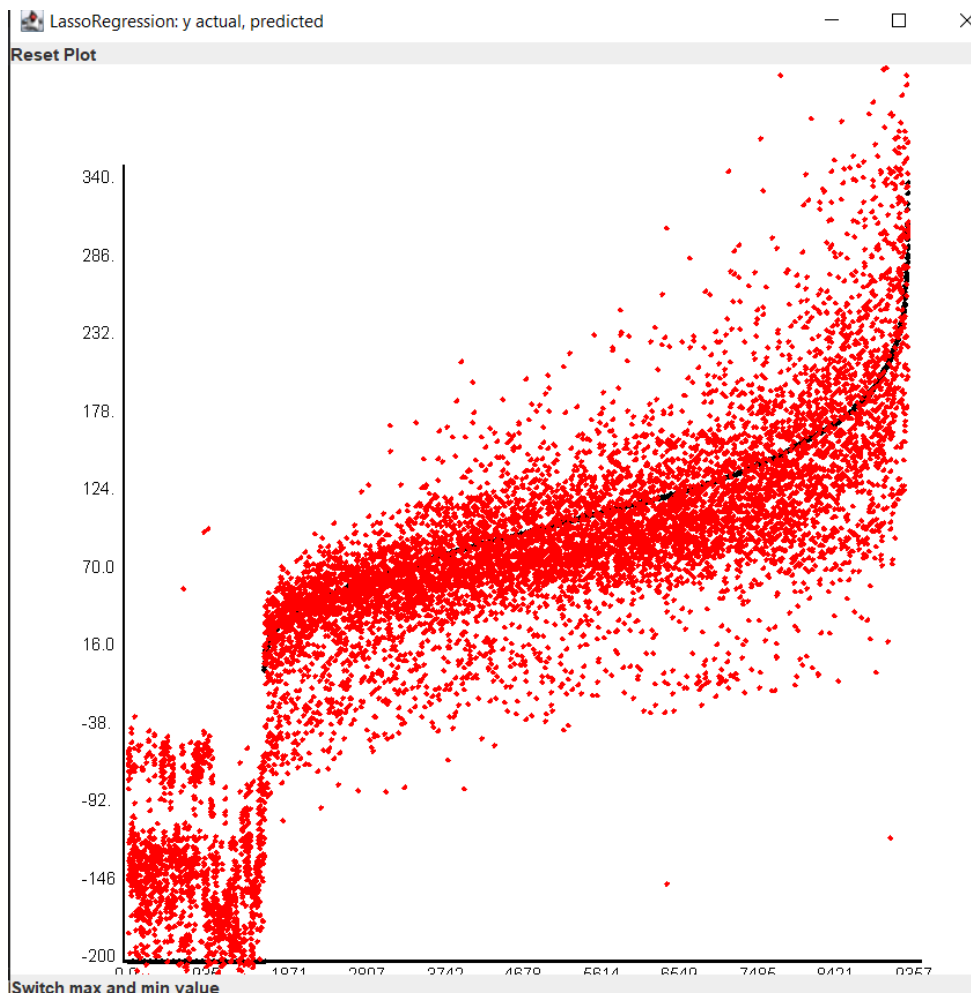[info]

Reset Plot
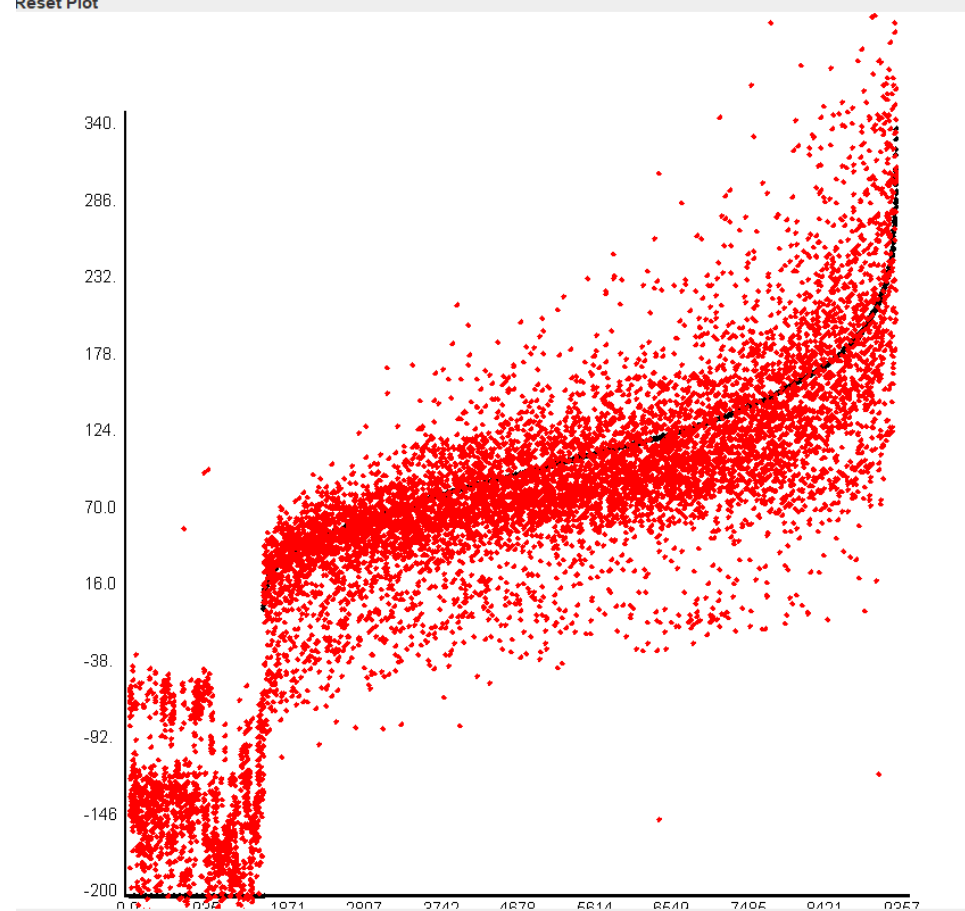
R2: 0.814053
Adj R2: 0.813814

Lasso Regression:

149.99, 64.4204,   53.6079,   1.00000,   9355.00,   27048.0,   -10444.4,   -10433.3,   4
9.6001, 59.490)
[info] DEBUG @ Predictor.crossValidate: fold 3: test set size = 1871
[info] DEBUG @ Predictor.validate: test set size = 1871
[info] DEBUG @ LassoRegression.train: LassoAdmm estimates parameter b = VectorD(0.413841,   0.36228)
[info] DEBUG @ Predictor.validate: parameters b = VectorD(0.413841,   0.36228)
[info] DEBUG @ Predictor.crossValidate: fold 3: qof = VectorD(0.711937, 0.711906,   2.94996e+07,   8.49774e+06,   4
541.82, 67.3930,   55.5291,   1.00000,   9355.00,   23120.5,   -10531.8,   -10520.7,   5
0.6385, 62.118)
[info] DEBUG @ Predictor.crossValidate: fold 4: test set size = 1871
[info] DEBUG @ Predictor.validate: test set size = 1871
[info] DEBUG @ LassoRegression.train: LassoAdmm estimates parameter b = VectorD(0.399677,   0.36656)
[info] DEBUG @ Predictor.validate: parameters b = VectorD(0.399677,   0.36656)
[info] DEBUG @ Predictor.crossValidate: fold 4: qof = VectorD(0.732676, 0.732647,   3.02237e+07,   8.07953e+06,   4
318.29, 65.7137,   53.9803,   1.00000,   9355.00,   25640.0,   -10481.9,   -10470.9,   4
8.8367, 58.883)
[info] --
[info] | backwardElimAll: (l = 10) REMOVE variable (7, x7) => cols = LinkedHashSet(0, 5) @ 0.733063524968985 |
[info] --
[info] REPORT
[info]   modelName mn = LassoRegression
[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]   features   fn = Array(x0, x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11)
[info]   parameter b   = VectorD(0.361221,   0.0242899,   0.0507000,   -12.2446,   0.133835,   0.437869
,   -0.0163850,   0.119913,   -0.0274083,   -3.30327,   -2.31670,   17.714)
[info]   fitMap   qof = LinkedHashMap(rSq -> 0.814053, rSqBar -> 0.813834, sst -> 150761453.620609, sse -> 28033646.
260312, mse0 -> 2996.007936, rmse -> 54.735801, mae -> 39.981064, dfm -> 11.000000, df -> 9345.000000, fStat -> 3719.204
056, aic -> -50704.568752, bic -> -50618.842191, mape -> 34.230670, smape -> 40.344156)
[info]

LassoRegression: y actual, predicted

R2: 0.814053
Adj R2: 0.813834

Quadratic Regression:


```
[info] ------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ------------------------------------------------
[info] |         name | num |         min |          max |         mean |       stdev |     interval |
[info] ------------------------------------------------
[info] |          rSq |   5 |       0.576 |        0.638 |        0.608 |       0.022 |        0.028 |
[info] |       rSqBar |   5 |       0.576 |        0.638 |        0.608 |       0.022 |        0.028 |
[info] |          sst |   5 | 82832113.983 | 87920724.853 | 84690782.380 | 2002528.300 | 2486954.372 |
[info] |          sse |   5 | 30805228.308 | 35319171.311 | 33205180.373 | 1753807.782 | 2178066.563 |
[info] |         mse0 |   5 |   16464.580 |    18877.163 |    17747.290 |     937.364 |     1164.119 |
[info] |         rmse |   5 |     128.314 |      137.394 |      133.182 |       3.530 |        4.384 |
[info] |          mae |   5 |      94.416 |      100.455 |       97.315 |       2.297 |        2.853 |
[info] |          dfm |   5 |       2.000 |        2.000 |        2.000 |       0.000 |        0.000 |
[info] |           df |   5 |    9354.000 |     9354.000 |     9354.000 |       0.000 |        0.000 |
[info] |        fStat |   5 |    6359.958 |     8242.418 |     7278.074 |     679.657 |      844.071 |
[info] |          aic |   5 |  -11861.350 |   -11734.099 |   -11801.755 |      49.441 |       61.401 |
[info] |          bic |   5 |  -11844.747 |   -11717.496 |   -11785.152 |      49.441 |       61.401 |
[info] |         mape |   5 |       8.750 |        9.294 |        9.033 |       0.258 |        0.320 |
[info] |        smape |   5 |       8.554 |        9.100 |        8.820 |       0.228 |        0.283 |
[info] ------------------------------------------------
[info] REPORT
[info]      modelName  mn  = PolyRegression
[info]      hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_Cholesky,Fac_Cholesky)))
[info]      features   fn  = Array(x0, x1, x2)
[info]      parameter  b   = VectorD(784.874,    164.252,        -6.1798)
[info]      fitMap     qof = LinkedHashMap(rSq -> 0.608218, rSqBar -> 0.608134, sst -> 423642445.749750, sse -> 165975457
.961019, mse0 -> 17738.106013, rmse -> 133.184481, mae -> 97.271774, dfm -> 2.000000, df -> 9354.000000, fStat -> 7260.7
63228, aic -> -59042.974549, bic -> -59021.542909, mape -> 9.024178, smape -> 8.817180)
[info]
```
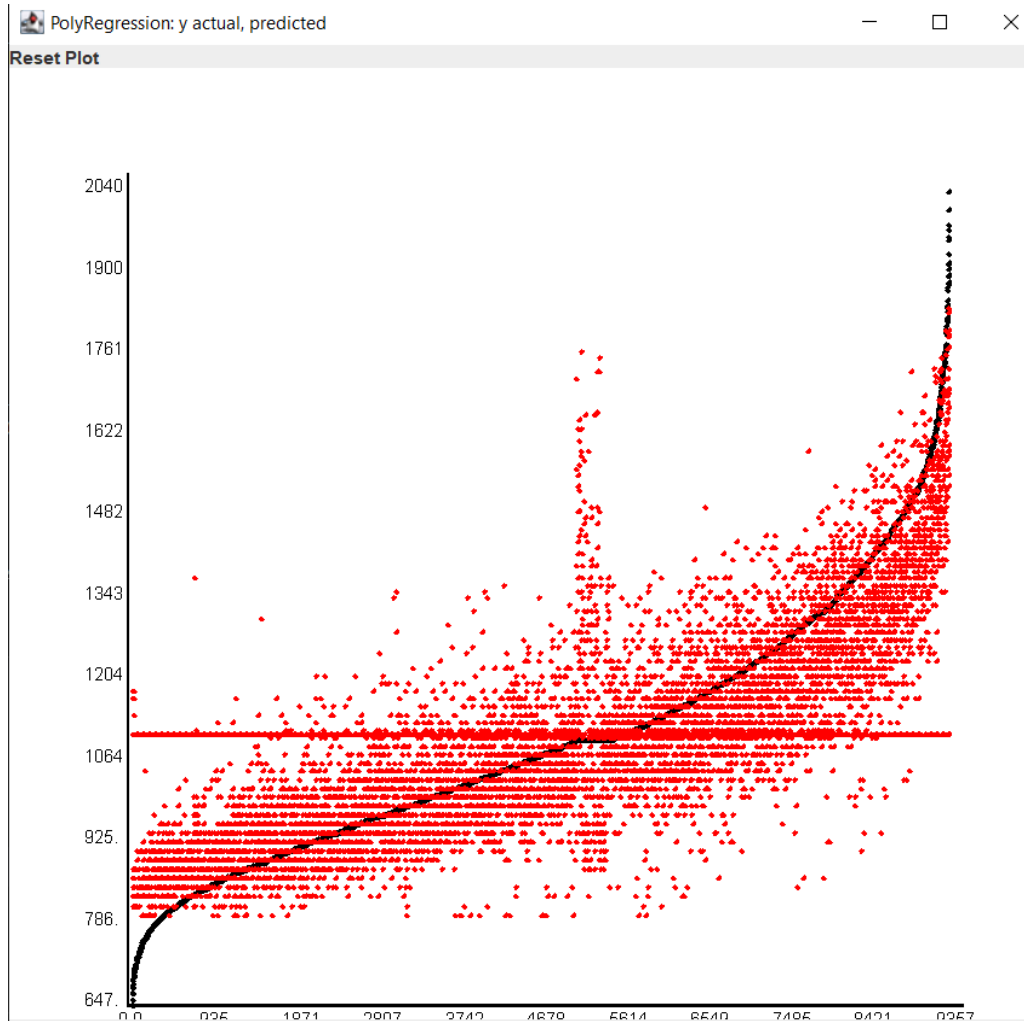
R2: 0.608218
Adj R2: 0.608134

Symbolic Regression:

Command Prompt - sbt

```
[info]      hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]      features   fn  = Array(CO(GT), PT08.S1(CO), NMHC(GT), C6H6(GT), PT08.S2(NMHC), NOx(GT), PT08.S3(NOx), PT08.S4
(NO2), PT08.S5(O3), T, RH, AH, PT08.S1(CO)_CO(GT), NMHC(GT)_CO(GT), NMHC(GT)_PT08.S1(CO), C6H6(GT)_CO(GT), C6H6(GT)_PT08
.S1(CO), C6H6(GT)_NMHC(GT), PT08.S2(NMHC)_CO(GT), PT08.S2(NMHC)_PT08.S1(CO), PT08.S2(NMHC)_NMHC(GT), PT08.S2(NMHC)_C6H6(
GT), NOx(GT)_CO(GT), NOx(GT)_PT08.S1(CO), NOx(GT)_NMHC(GT), NOx(GT)_C6H6(GT), NOx(GT)_PT08.S2(NMHC), PT08.S3(NOx)_CO(GT)
, PT08.S3(NOx)_PT08.S1(CO), PT08.S3(NOx)_NMHC(GT), PT08.S3(NOx)_C6H6(GT), PT08.S3(NOx)_PT08.S2(NMHC), PT08.S3(NOx)_NOx(G
T), PT08.S4(NO2)_CO(GT), PT08.S4(NO2)_PT08.S1(CO), PT08.S4(NO2)_NMHC(GT), PT08.S4(NO2)_C6H6(GT), PT08.S4(NO2)_PT08.S2(NM
HC), PT08.S4(NO2)_NOx(GT), PT08.S4(NO2)_PT08.S3(NOx), PT08.S5(O3)_CO(GT), PT08.S5(O3)_PT08.S1(CO), PT08.S5(O3)_NMHC(GT),
 PT08.S5(O3)_C6H6(GT), PT08.S5(O3)_PT08.S2(NMHC), PT08.S5(O3)_NOx(GT), PT08.S5(O3)_PT08.S3(NOx), PT08.S5(O3)_PT08.S4(NO2
), T_CO(GT), T_PT08.S1(CO), T_NMHC(GT), T_C6H6(GT), T_PT08.S2(NMHC), T_NOx(GT), T_PT08.S3(NOx), T_PT08.S4(NO2), T_PT08.S
5(O3), RH_CO(GT), RH_PT08.S1(CO), RH_NMHC(GT), RH_C6H6(GT), RH_PT08.S2(NMHC), RH_NOx(GT), RH_PT08.S3(NOx), RH_PT08.S4(NO
2), RH_PT08.S5(O3), RH_T, AH_CO(GT), AH_PT08.S1(CO), AH_NMHC(GT), AH_C6H6(GT), AH_PT08.S2(NMHC), AH_NOx(GT), AH_PT08.S3(
NOx), AH_PT08.S4(NO2), AH_PT08.S5(O3), AH_T, AH_RH)
[info]      parameter  b  = VectorD(-0.976034,   -0.135022,     0.462958,      -28.6695,      0.102849,       1.96163,
-0.0412362,    -0.105958,     0.0463079,     1.02112,       2.96906,       -180.036,      0.000858809,   -0.00027
7666,   -0.000101399,   -0.0143979,    -0.000115742,   0.0117626,     0.000463475,   0.000450867,    -0.000200400,  0
.00754315,    -0.000986516,  -0.000361759,   0.000254492,   0.0280100,     -0.00152005,    0.000147738,   0.000157
151,    -0.000162943,  0.00692277,    -0.000342181,   -5.68531e-05,   0.000307364,    -0.000300042,   -0.000222607,  -
0.0194473,     0.000616826,   0.000244363,   -4.10545e-06,   -0.000178948,   0.000131338,    -4.27895e-05,   0.011017
6,      -0.000412535,   9.07251e-06,   3.73794e-05,    1.87043e-05,   -0.0170942,    -0.00165674,    0.00775525,    0
.379995,      -0.00991272,   0.00261205,    0.00422417,    0.00297687,    -5.06375e-05,   -0.00565141,    0.001627
02,     0.00307199,    0.173256,       -0.00440764,   -0.00201795,    0.000933011,    -0.000643902,   -0.000815811,  0
.0683510,     0.0273173,     -0.0535103,    -0.0195080,    -3.89541,      -0.0526628,    -0.0184314,     -0.06017
77,     0.100579,      0.0312307,     2.67829,       -0.39155)
[info]      fitMap     qof = LinkedHashMap(rSq -> 0.952488, rSqBar -> 0.952093, sst -> 150761453.620609, sse -> 7163045.5
14921, mse0 -> 765.528002, rmse -> 27.668177, mae -> 19.072004, dfm -> 77.000000, df -> 9279.000000, fStat -> 2415.80771
7, aic -> -44188.894925, bic -> -43631.672284, mape -> 17.038205, smape -> 17.521613)
[info]
[info] Run + title
```

SymbolicRegressionX: y actual, predicted

Reset Plot

Switch max and min value
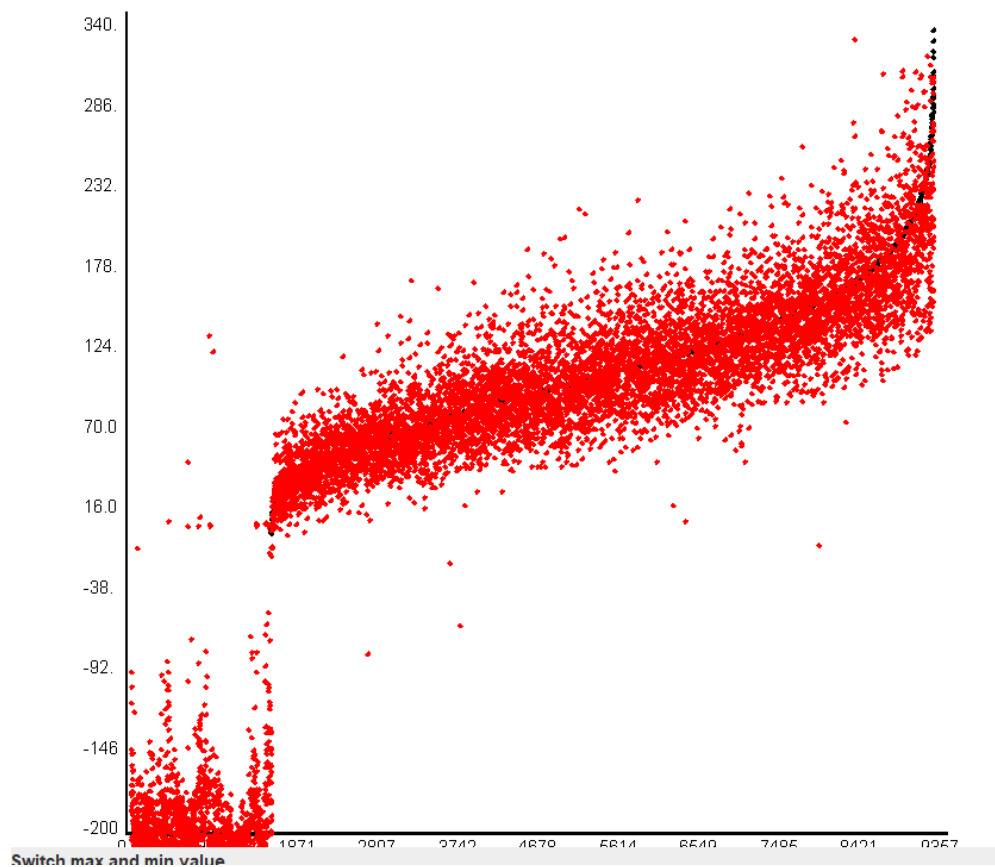
R2: 0.952488
Adj R2: 0.952093

Symbolic Ridge Regression:


```
-07,      -8.84040e-07,    -4.47320e-05,    1.05508e-06,     3.31231e-07,     -3.24656e-07,    -2.23063e-07,    -5.16524e-06,    6
.60529e-07,     2.18323e-08,    -6.23063e-07,    5.80493e-07,     4.88642e-07,     6.45662e-05,     -2.38438e-06,    -4.18862
e-06,      -9.92365e-07,    5.04494e-07,     -2.14042e-07,    -9.32604e-06,    3.66260e-08,     5.82433e-07,     6.78548e-07,    -
7.12718e-06,    5.53967e-08,     4.55389e-07,     2.27909e-07,     1.64509e-07,     -8.30708e-06,    6.10063e-08,     1.67320e
-07,      3.52666e-07,     -6.32627e-07,    -2.21246e-07,    1.83245e-05,     2.80966e-08,     -5.14840e-07,    -1.52283e-08,    -
7.21392e-05,    -0.000139480,    1.70203e-05,     0.000214924,     -0.000545448,    0.00218834,      -6.84406e-05,    1.94100e
-05,      4.67947e-05,     0.00139561,      5.99093e-05,     2.31865e-05,     -6.00460e-05,    0.00111925,      -1.91348e-05,    2
.97470e-05,     5.64197e-06,     -4.86790e-06,    -7.47441e-05,    5.01671e-06,     -9.73478e-06,    4.71226e-05,     4.14501e
-06,      -6.19314e-05,    -0.000990908,    -1.55594e-05,    3.17897e-06,     -5.64466e-06,    2.32544e-05,     -1.92580e-05,    -
9.34792e-06,    0.000585389,     -2.66625e-05,    -1.51099e-05,    1.55506e-06,     2.92118e-05,     -1.75075e-05,    -5.53138
e-05,      4.73336e-06,     -0.000190178,    -0.000407511,    0.00101097,      -2.94065e-05,    1.09492e-05,     1.48063e-05,     0
.000498114,    1.97546e-05,     8.48472e-06,     -4.79723e-06,    0.000429315,     -4.51327e-06,    6.23019e-06,     -6.57924
e-06,      2.14542e-06,     -0.000116261,    3.04644e-06,     5.50514e-07,     2.13532e-05,     1.70614e-06,     -2.44311e-05,    -
0.000288588,    -5.14263e-06,    -4.12850e-06,    2.07116e-06,     8.52461e-06,     -8.02412e-06,    -7.25291e-06,    0.000149
962,      -6.67296e-06,    -1.67216e-06,    2.15586e-06,     9.47900e-06,     0.000765924,     -0.000667760,    0.000136034,    0
.00577858,      0.000282976,     -0.000247622,    -7.26225e-05,    0.000204668,     0.000346179,     0.00166702,      0.000421
589,      -0.000900229,    0.0184341,       0.0228037,       -0.0416961,      -0.000513640,    -0.000885666,    0.00112785,      -
0.0208040,      -0.000212215,    -0.000703863,    0.000115796,     -0.00754324,     -0.000150840,    -0.000964173,    0.001181
26,      2.26409e-05,     0.0141106,       -0.000496938,    0.000109460,     0.000192733,     -1.54618e-05,    -0.000487668,    0
.0316075,       -0.000397677,    4.85638e-05,     -0.000247306,    -0.000747009,    0.00108636,      0.000379286,     -0.01588
68,      -0.000183161,    0.000538771,     -0.000302495,    -0.000464679,    -0.00960941,     -0.0292680,      0.0246750,       -
0.251069,      0.0330521,       0.00572265,      -0.00691699,     0.0105749,       -0.000753156,    -0.00953489,     -0.00083
2638,      0.0135597,       0.0504302,       0.00554654,      0.000653495,     -0.000675914,    -0.000802532,    -0.00396353,    0
.15478)
[info]      fitMap      qof = LinkedHashMap(rSq -> 0.963405, rSqBar -> 0.962474, sst -> 150761453.620609, sse -> 5517160.0
84849, mse0 -> 589.629164, rmse -> 24.282281, mae -> 15.939130, dfm -> 232.000000, df -> 9124.000000, fStat -> 1035.3347
85, aic -> -42657.469613, bic -> -40992.945572, mape -> 14.316041, smape -> 14.720222)
[info]
[info] Run + title
```
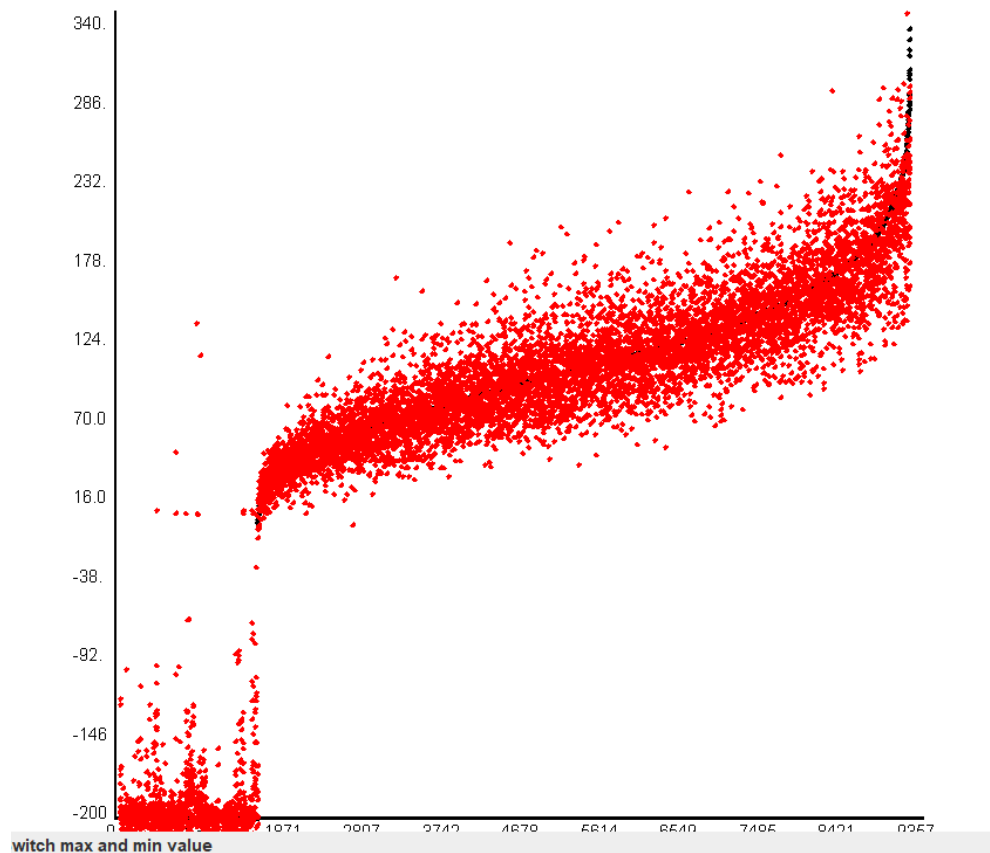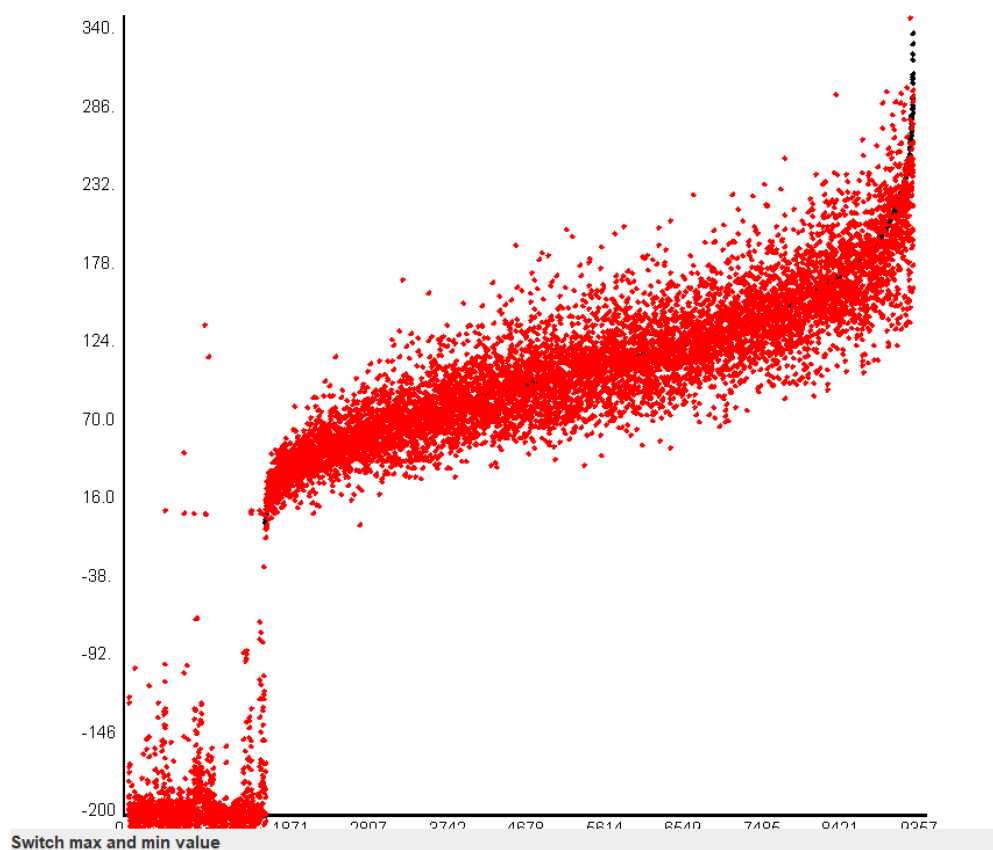
R2: 0.963405
Adj R2: 0.962474

Symbolic Lasso Regression:


```
-07,      -8.77799e-07,    -4.46690e-05,     1.05184e-06,     3.30803e-07,     -3.24808e-07,    -2.16698e-07,    -5.46940e-06,    6
.63980e-07,     2.30039e-08,     -6.22065e-07,    5.83632e-07,     4.88396e-07,     6.45258e-05,     -2.23250e-06,    -4.11175
e-06,    -9.93039e-07,    5.01678e-07,     -2.19406e-07,    -9.41814e-06,    3.56195e-08,     5.81606e-07,     6.79823e-07,     -
7.20755e-06,     5.82372e-08,     4.56029e-07,     2.27780e-07,     1.64346e-07,     -8.43369e-06,    6.25641e-08,     1.67635e
-07,     3.54090e-07,     -6.34009e-07,    -2.18273e-07,    1.82878e-06,     3.05591e-08,     -5.14233e-07,    -1.35755e-08,    -
7.19943e-05,    -0.000139347,    1.74097e-05,     0.000226635,     -0.000566950,    0.00221320,      -6.87668e-05,    1.96829e
-05,     4.63924e-05,     0.00140633,      5.98926e-05,     2.32231e-05,     -6.01596e-05,     0.00111664,      -1.90033e-05,    2
.97596e-05,     5.75446e-06,     -4.87193e-06,    -9.97852e-05,    5.47455e-06,     -9.77991e-06,    4.72374e-05,     4.07326e
-06,     -6.22309e-05,    -0.000993276,    -1.54053e-05,    3.14698e-06,     -5.71838e-06,    2.32148e-05,     -1.91948e-05,    -
9.31932e-06,    0.000591679,     -2.66479e-05,    -1.52117e-05,    1.53781e-06,     2.91178e-05,     -1.74777e-05,    -5.53214
e-05,     4.89859e-06,     -0.000187005,    -0.000412601,    0.00101726,      -2.94200e-05,    1.09948e-05,     1.47590e-05,     0
.000503741,     1.97540e-05,     8.51693e-06,     -4.88290e-06,    0.000429105,     -4.51304e-06,    6.24593e-06,     -6.50875
e-06,     2.15407e-06,     -0.000123619,    3.19027e-06,     5.38604e-07,     2.13523e-05,     1.67259e-06,     -2.45383e-05,    -
0.000291689,    -5.06880e-06,    -4.12896e-06,    2.02488e-06,     8.50945e-06,     -7.98094e-06,    -7.26080e-06,    0.000151
088,     -6.66459e-06,    -1.70537e-06,    2.14015e-06,     9.45243e-06,     0.000768489,     -0.000666137,    0.000142847,     0
.00607473,      0.000279200,     -0.000247359,    -7.17844e-05,    0.000201927,     0.000347427,     0.00166534,      0.000419
087,     -0.000908427,    0.0183996,      0.0235226,      -0.0420502,      -0.000514440,    -0.000901235,    0.00113020,      -
0.0212816,      -0.000212232,    -0.000705894,    0.000116226,     -0.00755183,      -0.000150652,    -0.000965714,    0.001178
80,      1.67862e-05,     0.0148371,      -0.000509955,    0.000109538,     0.000192757,     -1.28924e-05,    -0.000485353,    0
.0317467,      -0.000401686,    4.90522e-05,     -0.000243230,    -0.000747462,    0.00108668,      0.000378544,     -0.01597
27,      -0.000184830,    0.000540278,     -0.000303063,    -0.000463209,    -0.00957925,      -0.0292176,      0.0249894,      -
0.249361,      0.0331281,      0.00574261,      -0.00698223,      0.0105074,      -0.000743263,    -0.00954361,      -0.00083
4453,     0.0135724,      0.0486696,      0.00561981,      0.000644675,     -0.000694131,    -0.000791242,    -0.00398140,     0
.15309)
[info]     fitMap     qof = LinkedHashMap(rSq -> 0.963405, rSqBar -> 0.962478, sst -> 150761453.620609, sse -> 5517138.2
70910, mse0 -> 589.626832, rmse -> 24.282233, mae -> 15.939459, dfm -> 231.000000, df -> 9125.000000, fStat -> 1039.9349
87, aic -> -42659.451115, bic -> -41002.070954, mape -> 14.316473, smape -> 14.720733)
[info]
[info] Run + title
```


R2: 0.963405
Adj R2: 0.962478

Summary: Both Symbolic Ridge Regression and Symbolic Lasso Regression are ideal for this model.

ForestFires Dataset:
We have taken this dataset from the UCI repository. Our aim is to predict the burned area of the forest. There are 518 instances in the dataset.

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMC - FFMC index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84
(this output variable is very skewed towards 0.0, thus it may make
sense to model with the logarithm transform).

Linear Regression:
modelName mn = Linear Regression
[info]   hparameter hp = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]   features   fn = Array(intercept, region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels)
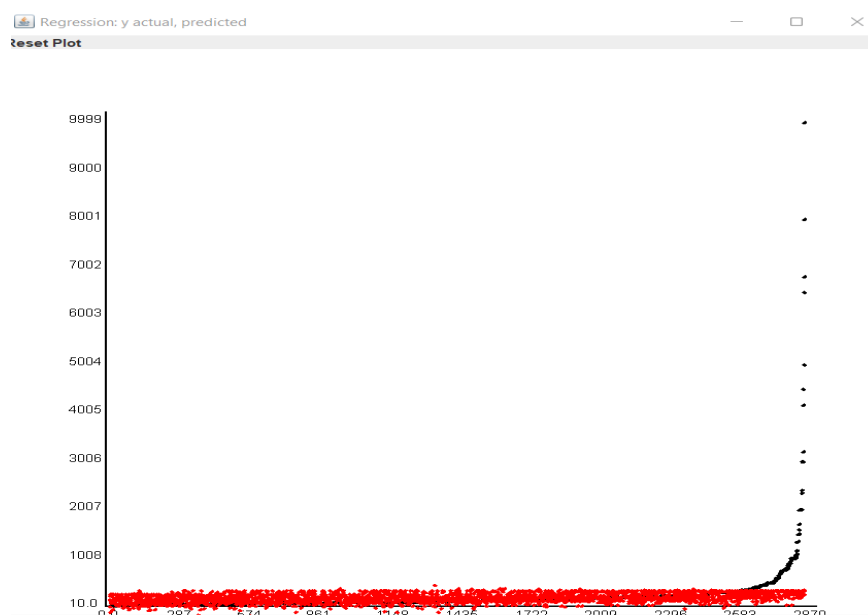[info]   parameter b = VectorD(291.366,   -72.0484,   -2.09576,   116.166, 0.122837,   -0.111580,   -22.5625,   -0.49375)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.048879, rSqBar -> 0.046552, sst -> 473344895.742509, sse -> 450208477.787416, mse0 -> 156867.065431, rmse -> 396.064471, mae -> 138.951817**, dfm -> 7.000000, df -> 2862.000000, fStat -> 21.011354, aic -> -21223.479593, bic -> -21175.783054, mape -> 107.846900, smape -> 63.717290)

```
[info] fname = Array(intercept, region, accommodation_type, yearly_availability, minimum_nights, nu
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var      Estimate     Std. Error      t value         Pr(>|t|)        VIF
[info] --------------------------------------------------------------------------------
[info]     x0       291.365759   27.477024       10.603978       0.000000        NA
[info]     x1       -72.048363   9.269240        -7.772845       0.000000        1.058020
[info]     x2       -2.095756    10.805352       -0.193955       0.846211        1.187625
[info]     x3       116.166033   17.014840       6.827336        0.000000        1.320475
[info]     x4       0.122837     0.199793        0.614823        0.538672        1.049737
[info]     x5       -0.111580    0.281115        -0.396920       0.691427        1.520657
[info]     x6       -22.562545   7.349080        -3.070118       0.002140        1.620195
[info]     x7       -0.493754    0.285724        -1.728079       0.083974        1.093946
[info]     Residual standard error: 396.617634 on 2862.0 degrees of freedom
[info]     Multiple R-squared:  0.048879,       Adjusted R-squared:  0.046552
[info]     F-statistic: 21.011354089903588 on 7.0 and 2862.0 DF,  p-value: 0.0
[info] --------------------------------------------------------------------------------
[info]
```

Cross-Validation

```
[info] --------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] --------------------------------------------------------------
[info] |       name | num |       min |       max |       mean |     stdev |   interval |
[info] --------------------------------------------------------------
[info] |       rSq |  5 |     0.037 |     0.061 |     0.048 |     0.010 |     0.013 |
[info] |     rSqBar |  5 |     0.035 |     0.059 |     0.045 |     0.010 |     0.013 |
[info] |       sst |  5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |       sse |  5 | 61458971.265 | 150647766.310 | 90394318.546 | 38469064.711 | 47775009.558 |
[info] |       mse0 |  5 | 107071.379 | 262452.555 | 157481.391 | 67019.276 | 83231.724 |
[info] |       rmse |  5 |   327.218 |   512.301 |   390.295 |    80.246 |    99.658 |
[info] |       mae |  5 |   135.323 |   150.927 |   139.585 |     6.492 |     8.062 |
[info] |       dfm |  5 |     7.000 |     7.000 |     7.000 |     0.000 |     0.000 |
[info] |        df |  5 |  2862.000 |  2862.000 |  2862.000 |     0.000 |     0.000 |
[info] |     fStat |  5 |    15.819 |    26.699 |    20.504 |     4.732 |     5.877 |
[info] |       aic |  5 | -4425.072 | -4140.791 | -4233.020 |   122.617 |   152.279 |
[info] |       bic |  5 | -4390.251 | -4105.970 | -4198.199 |   122.617 |   152.279 |
[info] |       mape |  5 |   101.210 |   115.381 |   108.296 |     5.461 |     6.782 |
[info] |     smape |  5 |    61.940 |    66.447 |    64.042 |     1.729 |     2.148 |
[info] --------------------------------------------------------------
```

Lasso Regression:

[info] REPORT
[info]   modelName mn = LassoRegression
[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]   features   fn = Array(x0, x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11)
[info]   parameter b   = VectorD(1.91525,   0.228364,   2.82000,   1.33261, -0.268104,   0.101702,   -0.0307685,   -0.654885,   0.891355,   -0.206911,   1.14547, -3.1241)
[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.025264, rSqBar -> 0.004032, sst -> 2090864.624009, sse -> 2038040.487307, mse0 -> 3942.051233, rmse -> 62.785757, mae -> 19.527744**, dfm -> 11.000000, df -> 505.000000, fStat -> 1.189921, aic -> -2849.830723, bic -> -2798.854208, mape -> Infinity, smape -> 153.844345)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var      Estimate     Std. Error       t value        Pr(>|t|)        VIF
[info]     ----------------------------------------------------------------------------
[info]     x0     1.915253      1.447138         1.323476       0.185677        NA
[info]     x1     0.228364      2.716092         0.084078       0.932994        1.426780
[info]     x2     2.819998      2.775993         1.015852       0.309700        5.103943
[info]     x3     1.332613      1.333233         0.999535       0.317535        1.014415
[info]     x4     -0.268104     0.275146         -0.974408      0.329854        0.294951
[info]     x5     0.101702      0.068039         1.494770       0.134974        2.427904
[info]     x6     -0.030769     0.031967         -0.962505      0.335796        8.040274
[info]     x7     -0.654885     0.703758         -0.930555      0.352084        1.316450
[info]     x8     0.891355      0.760351         1.172295       0.241079        2.492321
[info]     x9     -0.206911     0.207833         -0.995563      0.319462        1.470498
[info]     x10    1.145472      1.675054         0.683842       0.494075        1.151578
[info]     x11    -3.124122     9.614787         -0.324929      0.745235        1.035312
[info]     Residual standard error: 63.527346 on 505.0 degrees of freedom
[info]     Multiple R-squared:  0.025264,       Adjusted R-squared:  0.004032
[info]     F-statistic: 1.1899214511000655 on 11.0 and 505.0 DF,  p-value: 0.0
[info]     ----------------------------------------------------------------------------
[info]
```

Cross-validation:

```
[info] -----------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -----------------------------------------------
[info] |      name | num |       min |       max |      mean |      stdev |   interval |
[info] ---------------------------------------------------------------------------------
[info] |       rSq |   5 |    -1.223 |     0.022 |    -0.271 |      0.534 |      0.663 |
[info] |    rSqBar |   5 |    -1.271 |     0.001 |    -0.299 |      0.546 |      0.678 |
[info] |       sst |   5 | 22086.277 | 1319497.171 | 412859.609 | 559096.891 | 694346.471 |
[info] |       sse |   5 | 49089.909 | 1339122.427 | 421127.800 | 560221.257 | 695742.829 |
[info] |      mse0 |   5 |   476.601 | 13001.189 |  4088.619 |   5439.041 |   6754.785 |
[info] |      rmse |   5 |    21.831 |   114.023 |    52.451 |     40.888 |     50.779 |
[info] |       mae |   5 |    15.959 |    28.604 |    20.336 |      5.465 |      6.787 |
[info] |       dfm |   5 |    11.000 |    11.000 |    11.000 |      0.000 |      0.000 |
[info] |        df |   5 |   505.000 |   505.000 |   505.000 |      0.000 |      0.000 |
[info] |     fStat |   5 |   -25.254 |     1.049 |    -6.170 |     10.861 |     13.489 |
[info] |       aic |   5 |  -666.894 |  -503.269 |  -550.457 |     71.057 |     88.246 |
[info] |       bic |   5 |  -635.277 |  -471.652 |  -518.841 |     71.057 |     88.246 |
[info] |      mape |   5 | 179769313486231570000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
[info] |     smape |   5 |   145.773 |   164.441 |   156.293 |      6.881 |      8.546 |
[info] ---------------------------------------------------------------------------------
```
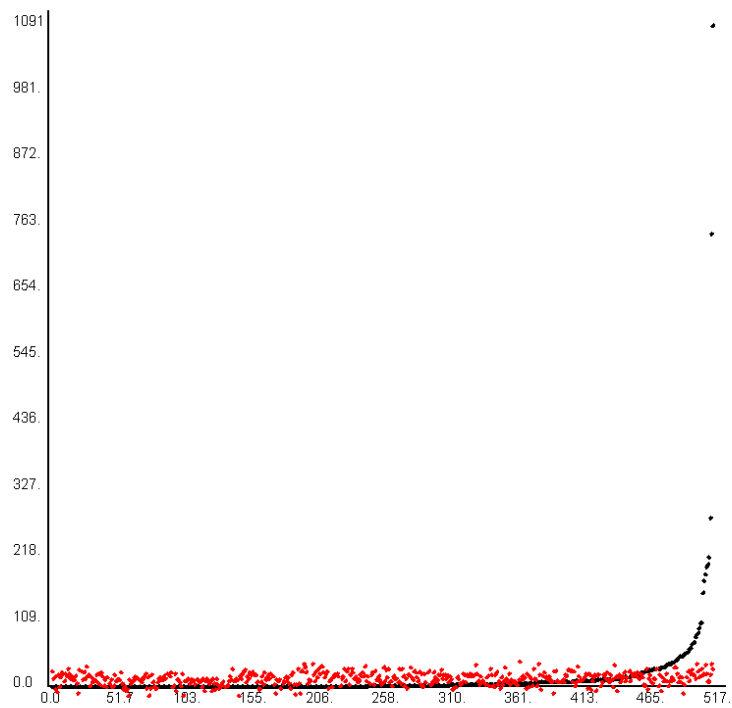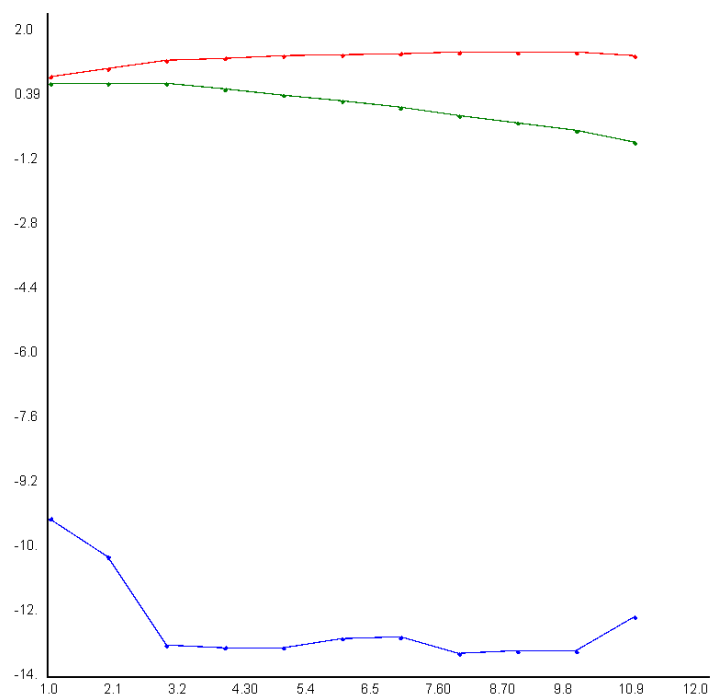
Ridge Regression:

[info] REPORT

[info]   modelName mn = RidgeRegression

[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky)))

[info]   features   fn = Array(X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain)

[info]   parameter b = VectorD(1.90078,   0.343908,   2.88462,   1.36215, -0.109384,   0.0967938,   -0.0314179,   -0.737874,   0.953676,   -0.174952, 1.22324,   -3.4021)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.025398, rSqBar -> 0.002193, sst -> 2090864.624009, sse -> 2037760.331537, mse0 -> 3941.509345, rmse -> 62.781441, mae -> 19.550098**, dfm -> 12.000000, df -> 504.000000, fStat -> 1.094525, aic -> -2847.795186, bic -> -2792.570629, mape -> 155.410179, smape -> 138.169308)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var      Estimate     Std. Error       t value        Pr(>|t|)          VIF
[info] -------------------------------------------------------------------------------
[info]     x0     1.900779       1.449515        1.311321        0.189749          NA
[info]     x1     0.343908       2.753794        0.124885        0.900615          1.463962
[info]     x2     2.884616       2.789390        1.034139        0.301071          5.143826
[info]     x3     1.362152       1.339174        1.017158        0.309078          1.021589
[info]     x4    -0.109384       0.662902       -0.165007        0.868938          1.708917
[info]     x5     0.096794       0.070608        1.370858        0.170419          2.609922
[info]     x6    -0.031418       0.032092       -0.979004        0.327578          8.088132
[info]     x7    -0.737874       0.771744       -0.956112        0.339016          1.580170
[info]     x8     0.953676       0.797038        1.196525        0.231492          2.733585
[info]     x9    -0.174952       0.240871       -0.726330        0.467637          1.971519
[info]     x10    1.223244       1.702437        0.718525        0.472433          1.187344
[info]     x11   -3.402107       9.681662       -0.351397        0.725291          1.047829
[info]     Residual standard error: 63.585967 on 504.0 degrees of freedom
[info]     Multiple R-squared:  0.025398,       Adjusted R-squared:  0.002193
[info]     F-statistic: 1.0945253223939044 on 12.0 and 504.0 DF,  p-value: 0.0
[info] -------------------------------------------------------------------------------
```

Cross-validation:

```
[info] ---------------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ---------------------------------------------------------------------
[info] |       name | num |       min |        max |       mean |      stdev |  interval |
[info] ---------------------------------------------------------------------
[info] |        rSq |   5 |    -1.153 |      0.021 |     -0.252 |      0.506 |     0.628 |
[info] |     rSqBar |   5 |    -1.204 |     -0.003 |     -0.282 |      0.518 |     0.643 |
[info] |        sst |   5 | 22086.277 | 1319497.171 | 412859.609 | 559096.891 | 694346.471 |
[info] |        sse |   5 | 47552.425 | 1327165.759 | 418344.845 | 555856.981 | 690322.803 |
[info] |       mse0 |   5 |   461.674 |  12885.104 |   4061.600 |   5396.670 |  6702.163 |
[info] |       rmse |   5 |    21.487 |    113.513 |     52.243 |     40.808 |    50.680 |
[info] |        mae |   5 |    15.509 |     29.866 |     20.342 |      6.169 |     7.662 |
[info] |        dfm |   5 |    12.000 |     12.000 |     12.000 |      0.000 |     0.000 |
[info] |         df |   5 |   504.000 |    504.000 |    504.000 |      0.000 |     0.000 |
[info] |      fStat |   5 |   -22.493 |      0.892 |     -5.312 |      9.767 |    12.130 |
[info] |        aic |   5 |  -663.393 |   -501.068 |   -548.105 |     70.513 |    87.571 |
[info] |        bic |   5 |  -629.142 |   -466.816 |   -513.853 |     70.513 |    87.571 |
[info] |       mape |   5 |   113.445 |    296.321 |    175.665 |     70.923 |    88.080 |
[info] |      smape |   5 |   126.783 |    156.154 |    140.360 |     12.063 |    14.981 |
[info] ---------------------------------------------------------------------
```
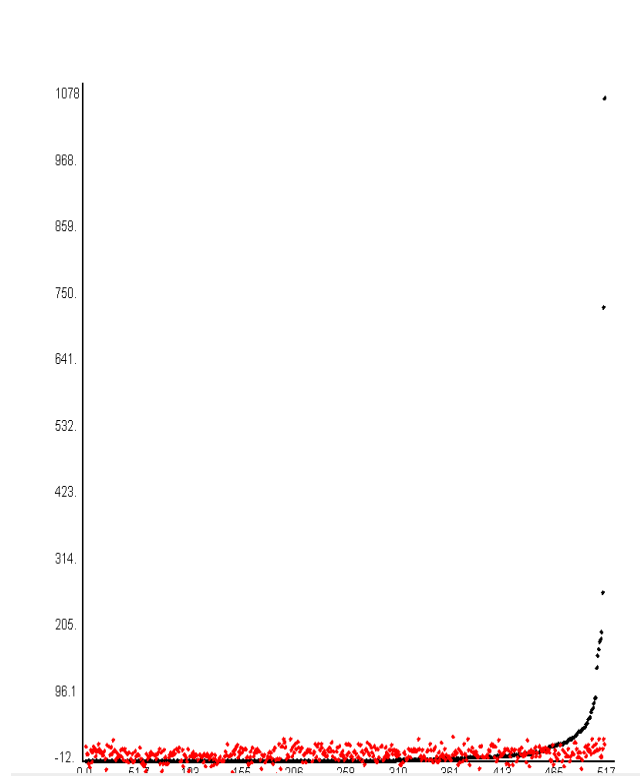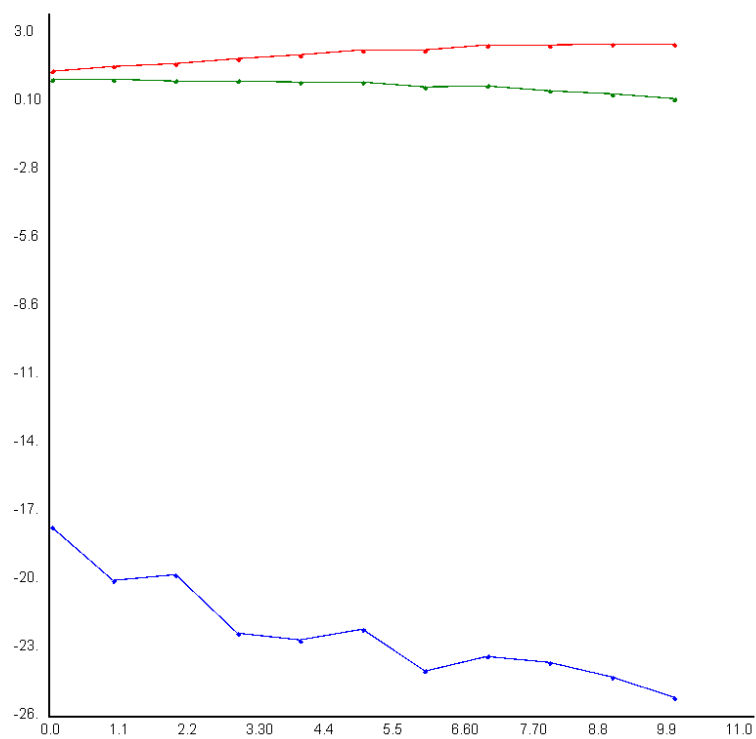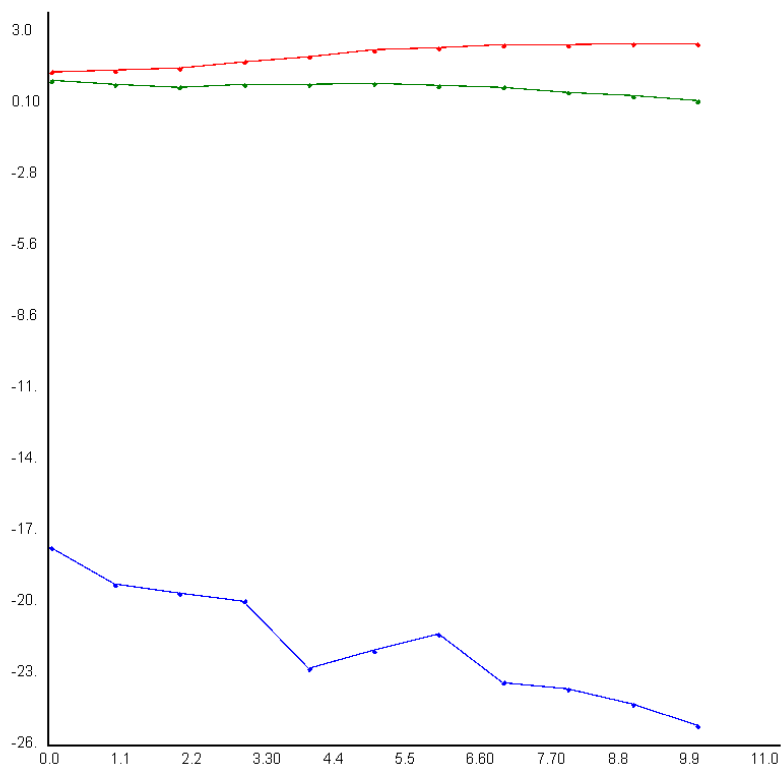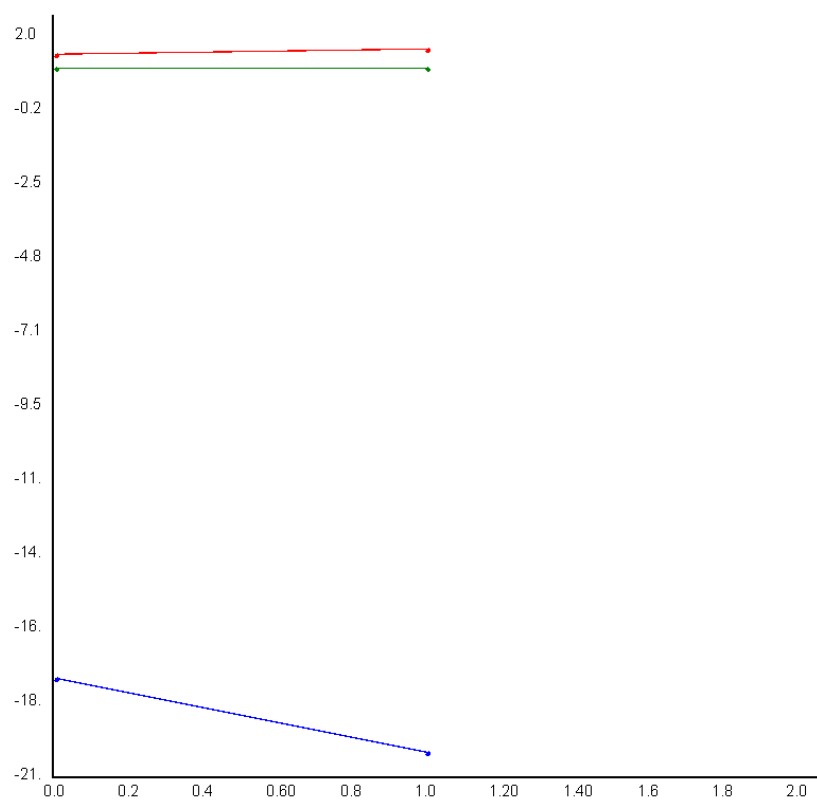
Quadratic Regression:

[info] REPORT
[info]   modelName mn = PolyRegression
[info]   hparameter hp = HyperParameter (HashMap(factorization ->
(Fac_Cholesky,Fac_Cholesky)))
[info]   features   fn = Array(x0, x1, x2)
[info]   parameter b   = VectorD(3.39145,   0.0459032,   0.025567)
[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.301324, rSqBar -> 0.298605, sst ->
780.529981, sse -> 545.337524, mse0 -> 1.054811, rmse -> 1.027040, mae -> 0.823228**, dfm
-> 2.000000, df -> 514.000000, fStat -> 110.838625, aic -> -741.385309, bic -> -728.641180,
mape -> 22.496013, smape -> 20.272131)

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var      Estimate    Std. Error      t value        Pr(>|t|)       VIF
[info] ----------------------------------------------------------------------------
[info]    x0    3.391446      0.187421       18.095352       0.000000      NA
[info]    x1    0.045903      0.090048        0.509766       0.610215      21.112255
[info]    x2    0.025567      0.009328        2.740913       0.006127      21.112255
[info]    Residual standard error: 1.030033 on 514.0 degrees of freedom
[info]    Multiple R-squared:  0.301324,     Adjusted R-squared:  0.298605
[info]    F-statistic: 110.83862513206601 on 2.0 and 514.0 DF,  p-value: 0.0
[info] ----------------------------------------------------------------------------
```

Cross-Validation:

```
[info] ------------------------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ------------------------------------------------------------------------------
[info] |    name |  num |       min |       max |      mean |     stdev |   interval |
[info] ------------------------------------------------------------------------------
[info] |     rSq |    5 |     0.160 |     0.370 |     0.283 |     0.079 |      0.098 |
[info] |  rSqBar |    5 |     0.157 |     0.367 |     0.280 |     0.079 |      0.099 |
[info] |     sst |    5 |   116.718 |   188.718 |   154.548 |    26.867 |     33.366 |
[info] |     sse |    5 |    94.843 |   127.205 |   109.723 |    15.255 |     18.945 |
[info] |    mse0 |    5 |     0.921 |     1.235 |     1.065 |     0.148 |      0.184 |
[info] |    rmse |    5 |     0.960 |     1.111 |     1.030 |     0.071 |      0.089 |
[info] |     mae |    5 |     0.757 |     0.916 |     0.826 |     0.069 |      0.086 |
[info] |     dfm |    5 |     2.000 |     2.000 |     2.000 |     0.000 |      0.000 |
[info] |      df |    5 |   514.000 |   514.000 |   514.000 |     0.000 |      0.000 |
[info] |   fStat |    5 |    49.037 |   150.724 |   104.832 |    38.029 |     47.229 |
[info] |     aic |    5 |  -151.696 |  -136.356 |  -143.409 |     7.231 |      8.980 |
[info] |     bic |    5 |  -143.792 |  -128.452 |  -135.505 |     7.231 |      8.980 |
[info] |    mape |    5 |    19.666 |    26.336 |    22.566 |     2.446 |      3.038 |
[info] |   smape |    5 |    18.127 |    23.160 |    20.341 |     1.978 |      2.457 |
[info] ------------------------------------------------------------------------------
```
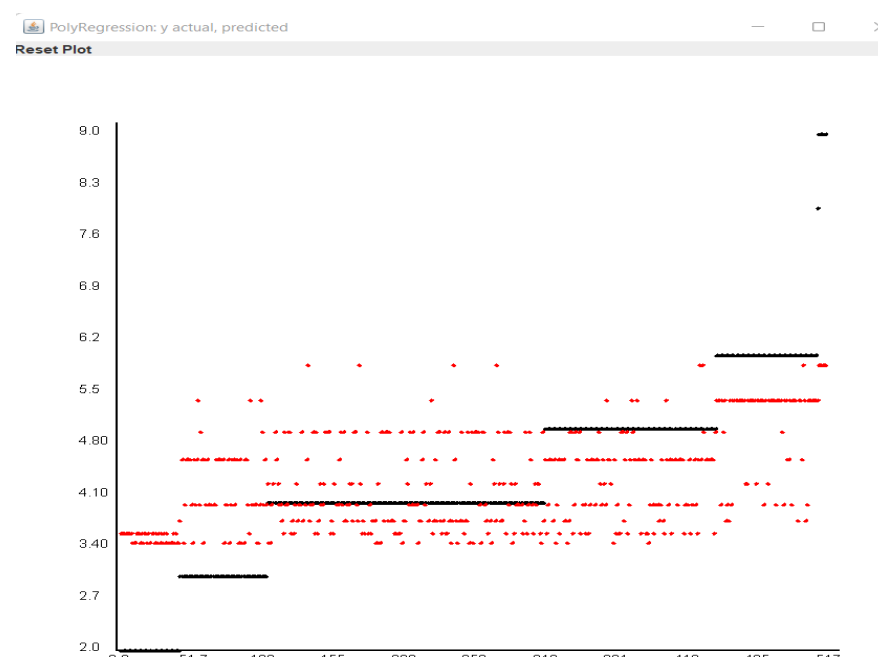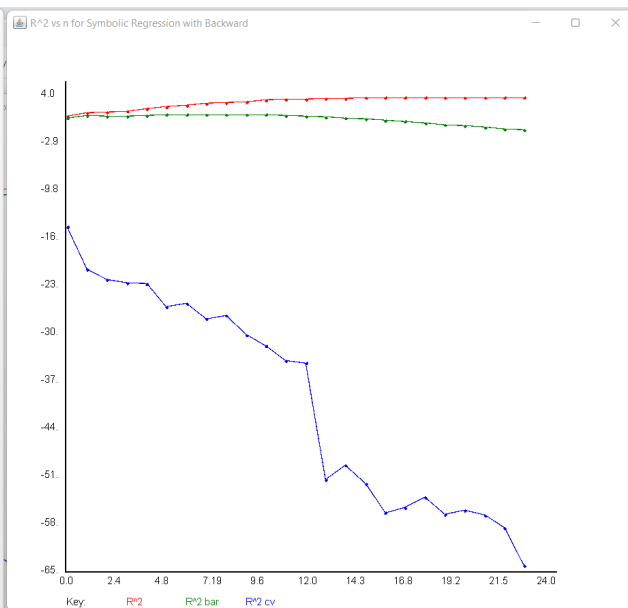

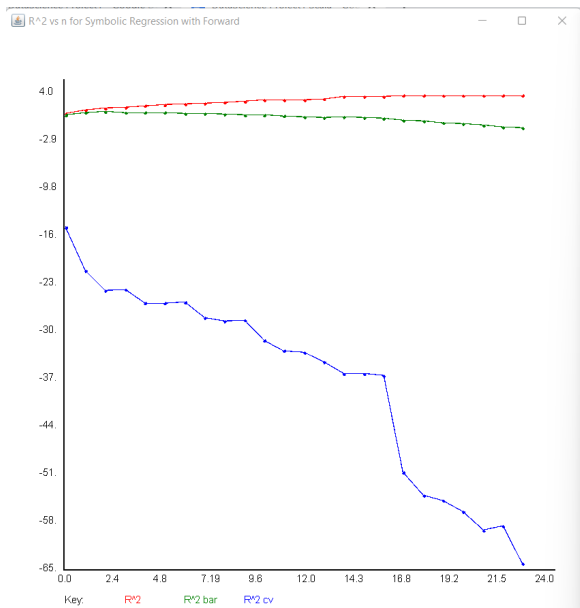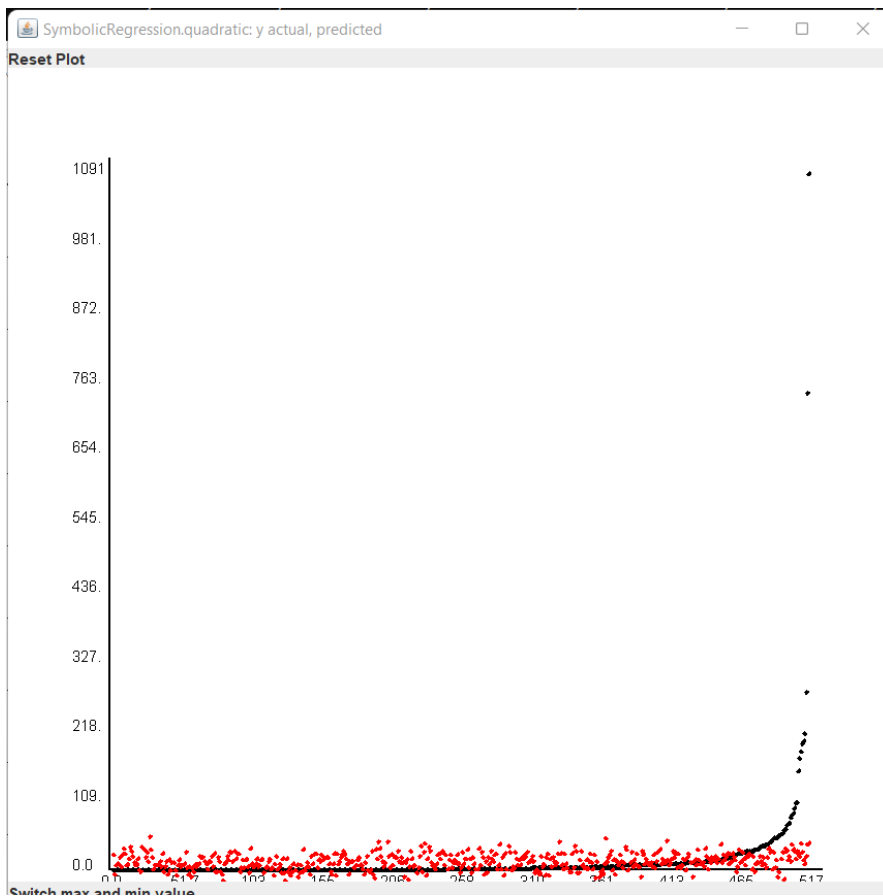PolyRegression: y actual, predicted
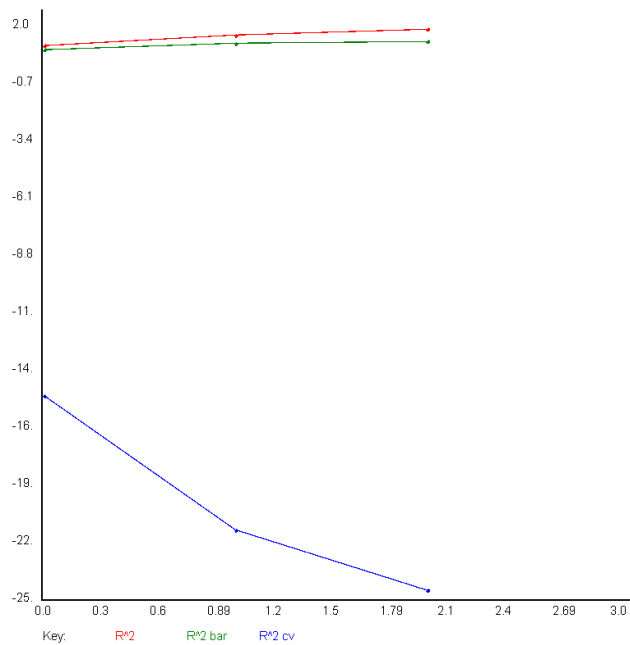
Symbolic Regression:

[info] REPORT

[info]   modelName mn = SymbolicRegression.quadratic

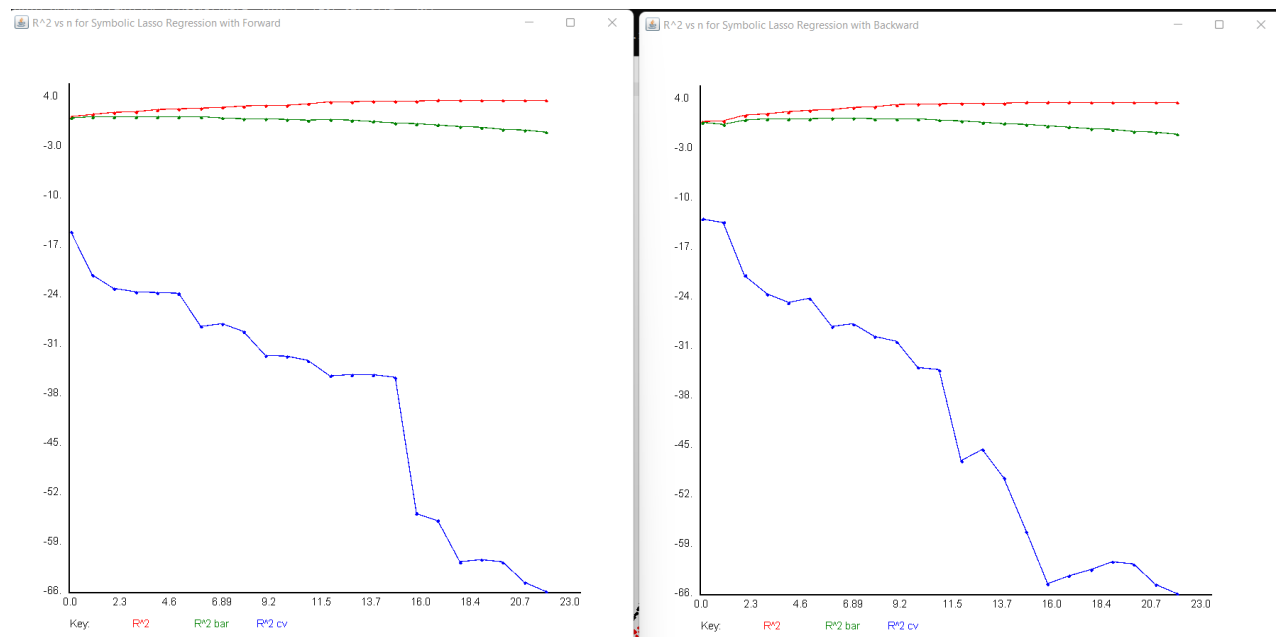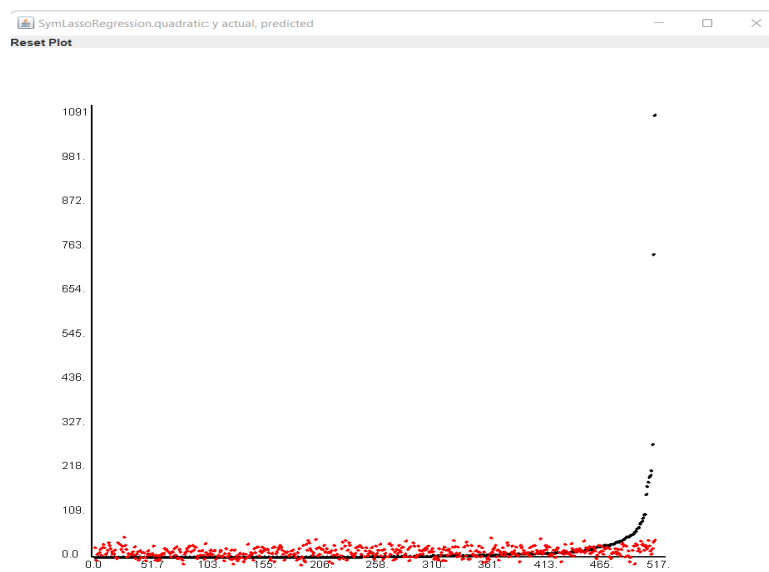[info]   hparameter hp = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))

[info]   features   fn = Array(intercept, X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain, X^2, Y^2, month^2, day^2, FFMC^2, DMC^2, DC^2, ISI^2, temp^2, RH^2, wind^2, rain^2)

[info]   parameter b   = VectorD(33.8738,   -0.252232,   -1.34350,   -13.8009,   -4.86745, -0.660461,   0.531026,   -0.0637533,   -1.45247,   0.612062,   -0.461598,   8.35282, -23.8968,   0.282141,   0.110506,   1.27133,   0.814943,   0.00509256, -0.00131266,   2.27165e-05,   0.0127240,   0.0142198,   0.00354986,   -0.757032, 3.3328)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.036853, rSqBar -> -0.010130, sst -> 2090864.624009, sse -> 2013810.569610, mse0 -> 3895.184854, rmse -> 62.411416, mae -> 19.712527**, dfm -> 24.000000, df -> 492.000000, fStat -> 0.784388, aic -> -2820.739045, bic -> -2714.537973, mape -> Infinity, smape -> 154.562640)

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var      Estimate      Std. Error      t value      Pr(>|t|)      VIF
[info] ----------------------------------------------------------------------------
[info]    x0    33.873753       121.338933       0.279166      0.780117      NA
[info]    x1    -0.252232         6.614921      -0.038131      0.969583      29.531729
[info]    x2    -1.343496        10.977560      -0.122386      0.902594      22.979879
[info]    x3   -13.800879        14.987043      -0.920854      0.357127     146.679318
[info]    x4    -4.867455         6.501757      -0.748637      0.454076      23.786627
[info]    x5    -0.660461         3.495172      -0.188964      0.850121      46.927668
[info]    x6     0.531026         0.310825       1.708442      0.087554      49.959470
[info]    x7    -0.063753         0.123401      -0.516634      0.605412     118.132937
[info]    x8    -1.452467         1.918072      -0.757253      0.448898       9.641757
[info]    x9     0.612062         3.236598       0.189107      0.850009      44.526772
[info]    x10   -0.461598         0.990324      -0.466108      0.641138      32.919718
[info]    x11    8.352821         6.514945       1.282102      0.199807      17.176129
[info]    x12  -23.896795        38.813508      -0.615682      0.538104      16.635101
[info]    x13    0.282141         0.697240       0.404654      0.685732      30.575021
[info]    x14    0.110506         1.187648       0.093046      0.925867      23.635443
[info]    x15    1.271333         1.054702       1.205395      0.228051     112.273831
[info]    x16    0.814943         0.806983       1.009863      0.312561      23.589064
[info]    x17    0.005093         0.025958       0.196184      0.844466      56.521938
[info]    x18   -0.001313         0.000931      -1.409855      0.158582      33.007449
[info]    x19    0.000023         0.000108       0.210328      0.833412      64.446159
[info]    x20    0.012724         0.041757       0.304714      0.760584       5.590794
[info]    x21    0.014220         0.075739       0.187748      0.851074      33.012194
[info]    x22    0.003550         0.009335       0.380286      0.703733      32.097929
[info]    x23   -0.757032         0.709920      -1.066362      0.286260      17.842384
[info]    x24    3.332823         6.268611       0.531668      0.594956      16.123072
[info]    Residual standard error: 63.977425 on 492.0 degrees of freedom
[info]    Multiple R-squared:  0.036853,      Adjusted R-squared:  -0.010130
[info]    F-statistic: 0.7843876375582628 on 24.0 and 492.0 DF,  p-value: 0.0
[info] ----------------------------------------------------------------------------
```

Cross-validation:

```
[info] --------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] --------------------------------------------------------
[info] |     name | num |        min |        max |        mean |        stdev |      interval |
[info] --------------------------------------------------------
[info] |      rSq |  5 |     -1.747 |     -0.024 |     -0.641 |      0.746 |       0.926 |
[info] |   rSqBar |  5 |     -1.881 |     -0.074 |     -0.721 |      0.782 |       0.971 |
[info] |      sst |  5 |  22086.277 | 1319497.171 | 412859.609 | 559096.891 | 694346.471 |
[info] |      sse |  5 |  60669.350 | 2722335.173 | 708117.178 | 1149864.998 | 1428025.653 |
[info] |     mse0 |  5 |    589.023 |  26430.439 |   6874.924 |  11163.738 |  13864.327 |
[info] |     rmse |  5 |     24.270 |    162.574 |     63.486 |     59.628 |      74.053 |
[info] |      mae |  5 |     17.137 |     39.399 |     23.692 |      9.434 |      11.716 |
[info] |      dfm |  5 |     24.000 |     24.000 |     24.000 |      0.000 |       0.000 |
[info] |       df |  5 |    492.000 |    492.000 |    492.000 |      0.000 |       0.000 |
[info] |    fStat |  5 |    -13.037 |     -0.490 |     -6.030 |      5.577 |       6.927 |
[info] |      aic |  5 |   -819.876 |   -478.214 |   -561.323 |    147.601 |     183.307 |
[info] |      bic |  5 |   -754.007 |   -412.346 |   -495.455 |    147.601 |     183.307 |
[info] |     mape |  5 | 17976931348623157000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000
[info] |    smape |  5 |    143.442 |    169.914 |    157.488 |     10.534 |      13.082 |
[info] --------------------------------------------------------
```

Reset Plot

1091
981.
872.
763.
654.
545.
436.
327.
218.
109.
0.0

Switch max and min value

R^2 vs n for Symbolic Regression with Forward   —   □   ✕

4.0
-2.9
-9.8
-16.
-23.
-30.
-37.
-44.
-51.
-58.
-65.

0.0   2.4   4.8   7.19   9.6   12.0   14.3   16.8   19.2   21.5   24.0

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

R^2 vs n for Symbolic Regression with Backward   —   □   ✕

4.0
-2.9
-9.8
-16.
-23.
-30.
-37.
-44.
-51.
-58.
-65.

0.0   2.4   4.8   7.19   9.6   12.0   14.3   16.8   19.2   21.5   24.0

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

R^2 vs n for Symbolic Regression with Stepwise

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

Symbolic Lasso Regression:

[info] REPORT

[info]   modelName mn = SymLassoRegression.quadratic

[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01)))

[info]   features   fn = Array(X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain, X^2, Y^2, month^2, day^2, FFMC^2, DMC^2, DC^2, ISI^2, temp^2, RH^2, wind^2, rain^2)
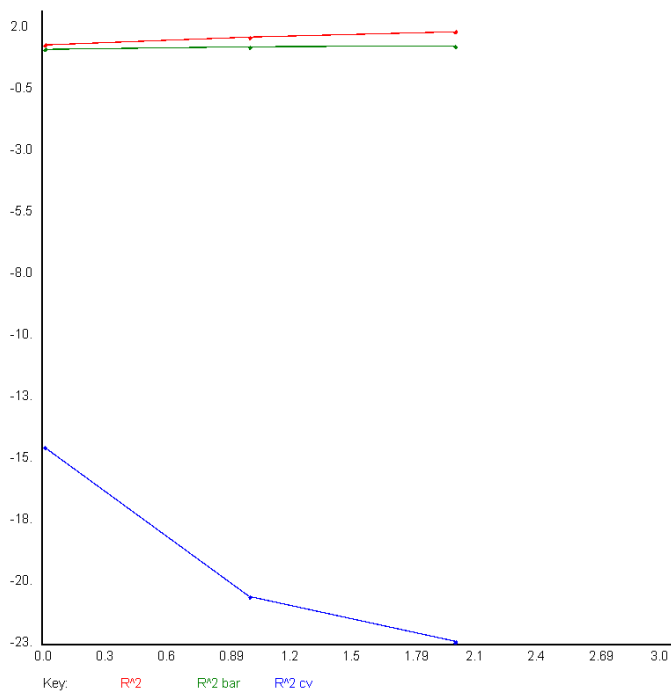
[info]   parameter b   = VectorD(-0.0338159, -0.800841,   -14.5594,   -4.67453,   0.199852,   0.527027,   -0.0573202,   -1.36578,   0.822753,   -0.487909,   8.34073, -24.5477, 0.255144,   0.0611386,   1.31842,   0.792161,   -0.000661795, -0.00129973,   1.69202e-05,   0.0115908,   0.0105412,   0.00400019,   -0.757373, 3.4301)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.036700, rSqBar -> -0.008241, sst -> 2090864.624009, sse -> 2014129.561691, mse0 -> 3895.801860, rmse -> 62.416359, mae -> 19.750481**, dfm -> 23.000000, df -> 493.000000, fStat -> 0.816630, aic -> -2822.779989, bic -> -2720.826960, mape -> Infinity, smape -> 154.122194)

```
[info] fname = Array(X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain, X^2, Y^2, month^2, day^2, FFMC^2, DMC^2, DC^2, ISI^2, temp^2, RH^2, wind^2, rain^2)
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]   Var     Estimate      Std. Error      t value      Pr(>|t|)      VIF
[info]  -----------------------------------------------------------------------------
[info]    x0    -0.033816     6.562345        -0.005153     0.995889      NA
[info]    x1    -0.800841     10.794149       -0.074192     0.940857      22.260038
[info]    x2    -14.559415    14.724352       -0.988798     0.322762      141.847722
[info]    x3    -4.674529     6.458901        -0.723735     0.469229      23.518071
[info]    x4    0.199852      1.647390        0.121315      0.903442      10.444738
[info]    x5    0.527027      0.310205        1.698964      0.089326      49.853750
[info]    x6    -0.057320     0.121116        -0.473265     0.636024      114.011829
[info]    x7    -1.365778     1.891004        -0.722250     0.470141      9.389107
[info]    x8    0.822753      3.144325        0.261663      0.793581      42.102879
[info]    x9    -0.487909     0.984904        -0.495387     0.620327      32.621435
[info]    x10   8.340726      6.508709        1.281472      0.200028      17.175386
[info]    x11   -24.547700    38.706231       -0.634205     0.525947      16.574271
[info]    x12   0.255144      0.689856        0.369851      0.711493      29.986954
[info]    x13   0.061139      1.173324        0.052107      0.958443      23.111970
[info]    x14   1.318423      1.040119        1.267569      0.204952      109.395071
[info]    x15   0.792161      0.802098        0.987611      0.323343      23.347995
[info]    x16   -0.000662     0.015763        -0.041984     0.966512      20.881669
[info]    x17   -0.001300     0.000929        -1.399001     0.161813      32.925981
[info]    x18   0.000017      0.000106        0.159790      0.873047      62.063511
[info]    x19   0.011591      0.041521        0.279157      0.780124      5.537981
[info]    x20   0.010541      0.074512        0.141469      0.887499      32.011619
[info]    x21   0.004000      0.009186        0.435487      0.663209      31.138735
[info]    x22   -0.757373     0.709255        -1.067844     0.285591      17.842328
[info]    x23   3.430108      6.252925        0.548561      0.583307      16.072544
[info]    Residual standard error: 63.917568 on 493.0 degrees of freedom
[info]    Multiple R-squared:  0.036700,     Adjusted R-squared:  -0.008241
[info]    F-statistic: 0.8166303749967986 on 23.0 and 493.0 DF,  p-value: 0.0
[info]  -----------------------------------------------------------------------------
```

Cross-validation:

```
[info] -------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -------------------------------------------------
[info] |     name | num |        min |         max |         mean |         stdev |       interval |
[info] -------------------------------------------------
[info] |      rSq |   5 |     -1.844 |       0.001 |       -0.658 |        0.791 |          0.983 |
[info] |   rSqBar |   5 |     -1.976 |      -0.045 |       -0.736 |        0.828 |          1.029 |
[info] |      sst |   5 |  22086.277 | 1319497.171 |   412859.609 |   559096.891 |     694346.471 |
[info] |      sse |   5 |  62803.043 | 2754270.936 |   711762.632 |  1163987.412 |    1445564.382 |
[info] |     mse0 |   5 |    609.738 |   26740.495 |     6910.317 |    11300.849 |      14034.606 |
[info] |     rmse |   5 |     24.693 |     163.525 |       63.549 |       59.915 |         74.409 |
[info] |      mae |   5 |     17.067 |      39.565 |       23.525 |        9.439 |         11.722 |
[info] |      dfm |   5 |     23.000 |      23.000 |       23.000 |        0.000 |          0.000 |
[info] |       df |   5 |    493.000 |     493.000 |      493.000 |        0.000 |          0.000 |
[info] |    fStat |   5 |    -13.897 |       0.032 |       -6.252 |        6.087 |          7.560 |
[info] |      aic |   5 |   -825.927 |    -480.495 |     -563.785 |      149.390 |        185.528 |
[info] |      bic |   5 |   -762.694 |    -417.262 |     -500.551 |      149.390 |        185.528 |
[info] |     mape |   5 | 179769313486231570000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
[info] |    smape |   5 |    143.517 |     169.135 |      157.638 |       10.391 |         12.905 |
[info] -------------------------------------------------
```


SymLassoRegression.quadratic: y actual, predicted

R^2 vs n for Symbolic Lasso Regression with Forward

R^2 vs n for Symbolic Lasso Regression with Backward

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

Symbolic Ridge Regression:

[info] REPORT

[info]   modelName mn = SymRidgeRegression.quadratic

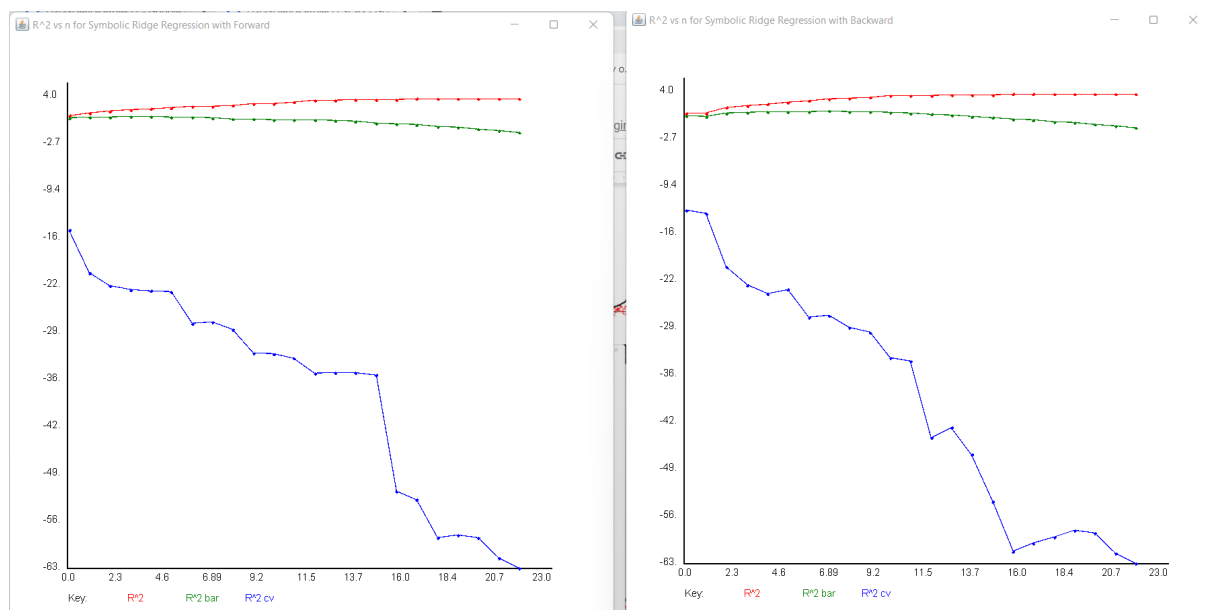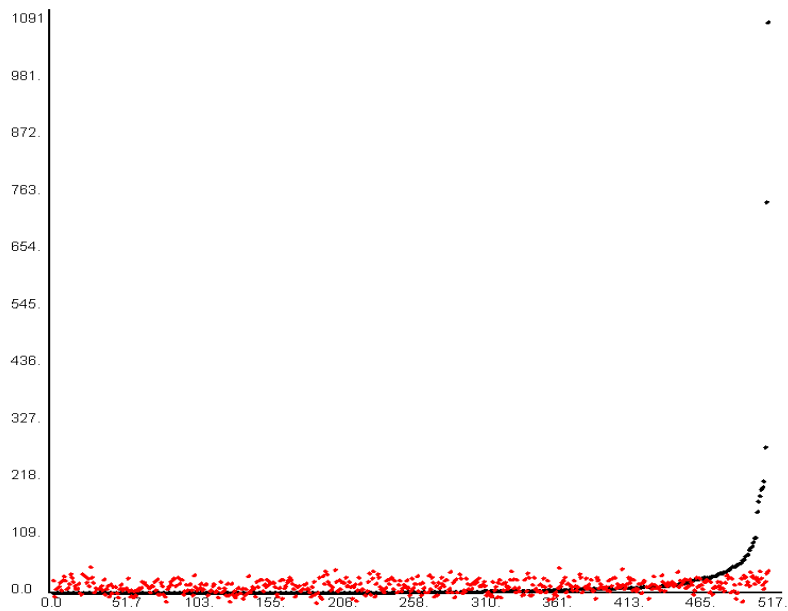[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky)))

[info]   features   fn = Array(X, Y, month, day, FFMC, DMC, DC, ISI, temp, RH, wind, rain, X^2, Y^2, month^2, day^2, FFMC^2, DMC^2, DC^2, ISI^2, temp^2, RH^2, wind^2, rain^2)

[info]   parameter b = VectorD(-0.0344328, -0.800701,   -14.5488,   -4.67538, 0.200168,   0.526929,   -0.0573334,   -1.36557,   0.821176,   -0.487953,   8.34000, -24.4586, 0.255175,   0.0611479,   1.31768,   0.792268,   -0.000664959, -0.00129945,   1.69325e-05,   0.0115876,   0.0105687,   0.00399885,   -0.757376, 3.4162)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.036700, rSqBar -> -0.010290, sst -> 2090864.624009, sse -> 2014129.586689, mse0 -> 3895.801908, rmse -> 62.416359, mae -> 19.749900**, dfm -> 24.000000, df -> 492.000000, fStat -> 0.781016, aic -> -2820.779992, bic -> -2714.578920, mape -> Infinity, smape -> 154.123007)

Cross-Validation:


```
[info] -----------------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -----------------------------------------------------------------------
[info] |   name | num |        min |         max |        mean |       stdev |     interval |
[info] -----------------------------------------------------------------------
[info] |    rSq |   5 |     -1.843 |       0.001 |      -0.629 |       0.774 |        0.962 |
[info] | rSqBar |   5 |     -1.982 |      -0.047 |      -0.709 |       0.812 |        1.008 |
[info] |    sst |   5 |  22086.277 | 1319497.171 |  412859.609 |  559096.891 |   694346.471 |
[info] |    sse |   5 |  62798.623 | 2562529.407 |  673412.035 | 1080003.030 |  1341263.571 |
[info] |   mse0 |   5 |    609.695 |   24878.926 |    6537.981 |   10485.466 |    13021.976 |
[info] |   rmse |   5 |     24.692 |     157.731 |      62.390 |      57.506 |       71.417 |
[info] |    mae |   5 |     17.066 |      38.749 |      23.361 |       9.093 |       11.293 |
[info] |    dfm |   5 |     24.000 |      24.000 |      24.000 |       0.000 |        0.000 |
[info] |     df |   5 |    492.000 |     492.000 |     492.000 |       0.000 |        0.000 |
[info] |  fStat |   5 |    -13.290 |       0.030 |      -5.832 |       5.681 |        7.055 |
[info] |    aic |   5 |   -799.318 |    -478.495 |    -556.863 |     138.611 |      172.142 |
[info] |    bic |   5 |   -733.450 |    -412.626 |    -490.995 |     138.611 |      172.142 |
[info] |   mape |   5 | 17976931348623157000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000
[info] |  smape |   5 |    143.494 |     169.131 |     157.634 |      10.397 |       12.912 |
[info] -----------------------------------------------------------------------
```

SymRidgeRegression.quadratic: y actual, predicted
Reset Plot

R^2 vs n for Symbolic Ridge Regression with Stepwise

Key:　　　R^2　　　R^2 bar　　　R^2 cv

Expedia Dataset:

Regression:
[info] REPORT
[info]　　modelName　mn　= Regression
[info]　　hparameter hp　= HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]　　features　fn　= Array(intercept, region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels)
[info]　　parameter　b　= VectorD(291.366,　-72.0484,　-2.09576,　116.166,　0.122837, -0.111580,　-22.5625,　-0.49375)
[info]　　fitMap　　qof = LinkedHashMap(**rSq -> 0.048879, rSqBar -> 0.046552, sst -> 473344895.742509, sse -> 450208477.787416, mse0 -> 156867.065431, rmse -> 396.064471, mae -> 138.951817**, dfm -> 7.000000, df -> 2862.000000, fStat -> 21.011354, aic -> -21223.479593, bic -> -21175.783054, mape -> 107.846900, smape -> 63.717290)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var      Estimate      Std. Error     t value        Pr(>|t|)      VIF
[info]     ------------------------------------------------------------------------
[info]     x0      291.365759    27.477024      10.603978      0.000000      NA
[info]     x1      -72.048363    9.269240       -7.772845      0.000000      1.058020
[info]     x2      -2.095756     10.805352      -0.193955      0.846211      1.187625
[info]     x3      116.166033    17.014840      6.827336       0.000000      1.320475
[info]     x4      0.122837      0.199793       0.614823       0.538672      1.049737
[info]     x5      -0.111580     0.281115       -0.396920      0.691427      1.520657
[info]     x6      -22.562545    7.349080       -3.070118      0.002140      1.620195
[info]     x7      -0.493754     0.285724       -1.728079      0.083974      1.093946
[info]     Residual standard error: 396.617634 on 2862.0 degrees of freedom
[info]     Multiple R-squared:  0.048879,      Adjusted R-squared:  0.046552
[info]     F-statistic: 21.011354089903588 on 7.0 and 2862.0 DF,  p-value: 0.0
[info]     ------------------------------------------------------------------------
```

```
[info] ----------------------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ----------------------------------------------------------------------------
[info] |        name | num |         min |         max |        mean |        stdev |      interval |
[info] ----------------------------------------------------------------------------
[info] |         rSq |  5  |       0.037 |       0.061 |       0.048 |        0.010 |         0.013 |
[info] |      rSqBar |  5  |       0.035 |       0.059 |       0.045 |        0.010 |         0.013 |
[info] |         sst |  5  | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |         sse |  5  | 61458971.265 | 150647766.310 | 90394318.546 | 38469064.711 | 47775009.558 |
[info] |        mse0 |  5  |  107071.379 |  262452.555 |  157481.391 |   67019.276 |    83231.724 |
[info] |        rmse |  5  |     327.218 |     512.301 |     390.295 |      80.246 |       99.658 |
[info] |         mae |  5  |     135.323 |     150.927 |     139.585 |       6.492 |        8.062 |
[info] |         dfm |  5  |       7.000 |       7.000 |       7.000 |       0.000 |        0.000 |
[info] |          df |  5  |    2862.000 |    2862.000 |    2862.000 |       0.000 |        0.000 |
[info] |       fStat |  5  |      15.819 |      26.699 |      20.504 |       4.732 |        5.877 |
[info] |         aic |  5  |   -4425.072 |   -4140.791 |   -4233.020 |     122.617 |      152.279 |
[info] |         bic |  5  |   -4390.251 |   -4105.970 |   -4198.199 |     122.617 |      152.279 |
[info] |        mape |  5  |     101.210 |     115.381 |     108.296 |       5.461 |        6.782 |
[info] |       smape |  5  |      61.940 |      66.447 |      64.042 |       1.729 |        2.148 |
[info] ----------------------------------------------------------------------------
```


Regression: y actual, predicted

Reset Plot
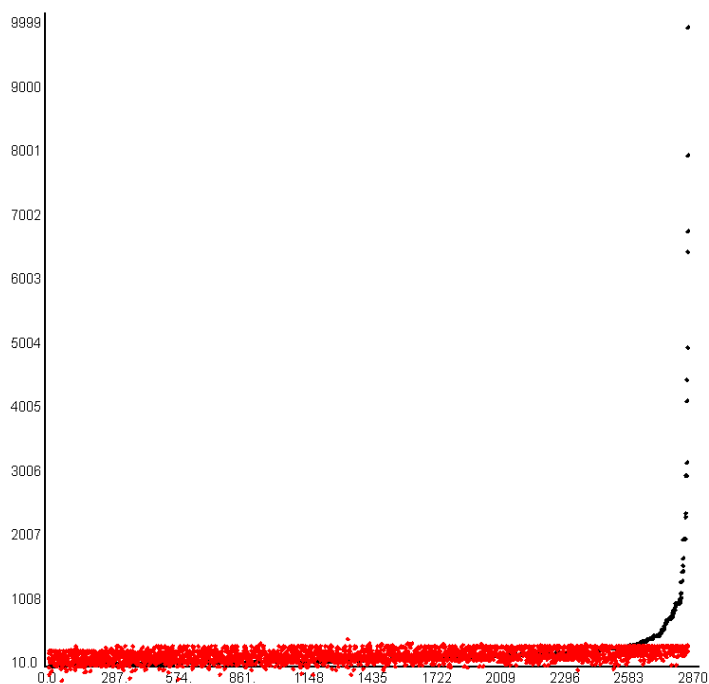
Linear Regression:
  Multiple R-squared: 0.048879
  Adjusted R-squared: 0.46552

Lasso Regression:
[info] REPORT
[info]   modelName mn = LassoRegression
[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]   features   fn = Array(x0, x1, x2, x3, x4, x5, x6)
[info]   parameter b   = VectorD(-16.6993, 75.5694,   116.332,   0.349717,   0.290970,
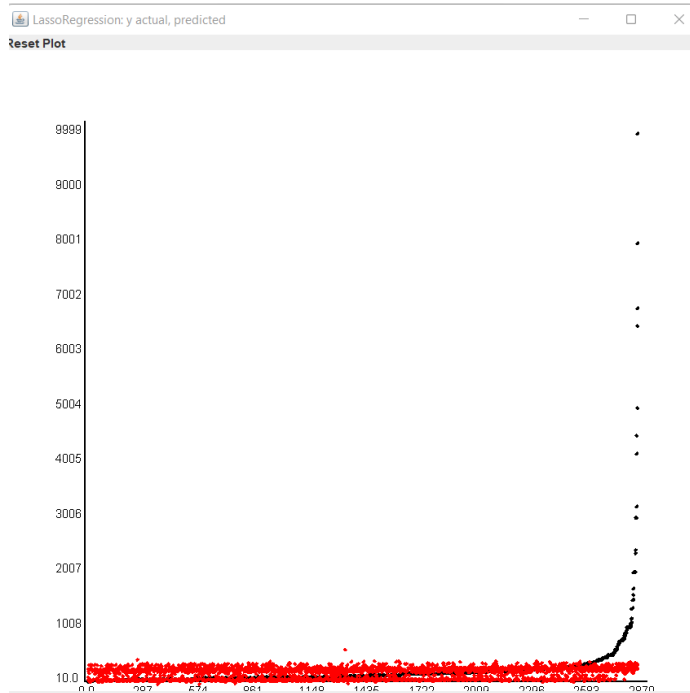-6.20799,   0.090560)
[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.011510, rSqBar -> 0.009439, sst ->
473344895.742509, sse -> 467896598.283164, mse0 -> 163030.173618, rmse -> 403.769951,**

mae -> 151.890328, dfm -> 6.000000, df -> 2863.000000, fStat -> 5.556240, aic -> -21280.784992, bic -> -21239.050521, mape -> 117.542858, smape -> 75.067564)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var       Estimate    Std. Error      t value       Pr(>|t|)        VIF
[info] ------------------------------------------------------------------------------
[info]     x0    -16.699272      7.807558     -2.138860       0.032447        NA
[info]     x1     75.569360      8.097478      9.332456       0.000000      0.641973
[info]     x2    116.332194     17.342829      6.707798       0.000000      1.320474
[info]     x3      0.349717      0.202473      1.727228       0.084127      1.037698
[info]     x4      0.290970      0.283909      1.024871       0.305424      1.492926
[info]     x5     -6.207991      7.323934     -0.847631       0.396644      1.548837
[info]     x6      0.090560      0.285765      0.316905       0.751316      1.053259
[info]     Residual standard error: 404.263256 on 2863.0 degrees of freedom
[info]     Multiple R-squared:  0.011510,      Adjusted R-squared:  0.009439
[info]     F-statistic: 5.55624030442443 on 6.0 and 2863.0 DF,  p-value: 0.0
[info] ------------------------------------------------------------------------------
```

Cross-Validation:

```
[info] -----------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -----------------------------------------------
[info] |    name | num |        min |        max |       mean |      stdev |    interval |
[info] ------------------------------------------------------------------------------------
[info] |     rSq |   5 |     -0.016 |      0.024 |      0.006 |      0.016 |       0.020 |
[info] |  rSqBar |   5 |     -0.018 |      0.022 |      0.004 |      0.016 |       0.020 |
[info] |     sst |   5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |     sse |   5 | 63887089.643 | 153988041.626 | 93885029.824 | 38345129.817 | 47621093.917 |
[info] |    mse0 |   5 | 111301.550 | 268271.850 | 163562.770 | 66803.362 | 82963.578 |
[info] |    rmse |   5 |    333.619 |    517.950 |    398.269 |     78.615 |      97.633 |
[info] |     mae |   5 |    146.119 |    156.987 |    152.344 |      4.411 |       5.479 |
[info] |     dfm |   5 |      6.000 |      6.000 |      6.000 |      0.000 |       0.000 |
[info] |      df |   5 |   2863.000 |   2863.000 |   2863.000 |      0.000 |       0.000 |
[info] |   fStat |   5 |     -7.317 |     11.840 |      2.938 |      7.812 |       9.702 |
[info] |     aic |   5 |  -4430.735 |  -4153.643 |  -4245.897 |    117.925 |     146.452 |
[info] |     bic |   5 |  -4400.267 |  -4123.174 |  -4215.429 |    117.925 |     146.452 |
[info] |    mape |   5 |    110.991 |    127.404 |    117.919 |      7.633 |       9.480 |
[info] |   smape |   5 |     70.770 |     79.015 |     75.273 |      3.337 |       4.145 |
[info] -----------------------------------------------
```


LassoRegression: y actual, predicted

Reset Plot

Lasso Regression:

    Multiple R-squared: 0.011510

    Adjusted R-squared: 0.009439

Ridge Regression:

[info] REPORT
[info]   modelName mn = RidgeRegression
[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky)))
[info]   features   fn = Array(region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels)
[info]   parameter b   = VectorD(-72.0478,   -2.09529,   116.164,   0.122841, -0.111580,   -22.5627,   -0.49374)
[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.048879, rSqBar -> 0.046552, sst -> 473344895.742509, sse -> 450208477.790368, mse0 -> 156867.065432, rmse -> 396.064471, mae -> 138.951617**, dfm -> 7.000000, df -> 2862.000000, fStat -> 21.011354, aic -> -21223.479593, bic -> -21175.783054, mape -> 331.066601, smape -> 128.639245)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var       Estimate      Std. Error       t value        Pr(>|t|)        VIF
[info] -----------------------------------------------------------------------------------
[info]     x0     -72.047806      9.269240       -7.772785       0.000000        NA
[info]     x1      -2.095288     10.805352       -0.193912       0.846245        1.187625
[info]     x2     116.163820     17.014840        6.827206       0.000000        1.320475
[info]     x3       0.122841      0.199793        0.614843       0.538658        1.049737
[info]     x4      -0.111580      0.281115       -0.396920       0.691427        1.520657
[info]     x5     -22.562680      7.349080       -3.070137       0.002140        1.620195
[info]     x6      -0.493745      0.285724       -1.728049       0.083980        1.093946
[info]     Residual standard error: 396.617634 on 2862.0 degrees of freedom
[info]     Multiple R-squared:  0.048879,       Adjusted R-squared:  0.046552
[info]     F-statistic: 21.01135408708497 on 7.0 and 2862.0 DF,  p-value: 0.0
[info] -----------------------------------------------------------------------------------
```

Cross-Validation:

```
[info] -----------------------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -----------------------------------------------------------------
[info] |      name | num |         min |         max |        mean |        stdev |     interval |
[info] -----------------------------------------------------------------
[info] |       rSq |   5 |       0.037 |       0.061 |       0.048 |        0.010 |        0.013 |
[info] |    rSqBar |   5 |       0.035 |       0.059 |       0.046 |        0.010 |        0.013 |
[info] |       sst |   5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |       sse |   5 | 61458775.771 | 150606394.364 | 90371609.659 | 38459384.951 | 47762988.194 |
[info] |      mse0 |   5 | 107071.038 | 262380.478 | 157441.829 | 67002.413 | 83210.781 |
[info] |      rmse |   5 |     327.217 |     512.231 |     390.246 |       80.236 |       99.646 |
[info] |       mae |   5 |     135.220 |     151.965 |     139.533 |        7.044 |        8.748 |
[info] |       dfm |   5 |       7.000 |       7.000 |       7.000 |        0.000 |        0.000 |
[info] |        df |   5 |    2862.000 |    2862.000 |    2862.000 |        0.000 |        0.000 |
[info] |     fStat |   5 |      15.891 |      26.701 |      20.613 |        4.717 |        5.858 |
[info] |       aic |   5 |   -4424.941 |   -4140.790 |   -4232.947 |      122.586 |      152.240 |
[info] |       bic |   5 |   -4390.120 |   -4105.969 |   -4198.126 |      122.586 |      152.240 |
[info] |      mape |   5 |     233.764 |     420.446 |     331.829 |       71.912 |       89.308 |
[info] |     smape |   5 |     122.942 |     132.594 |     128.794 |        4.127 |        5.126 |
[info] -----------------------------------------------------------------
```

Reset Plot
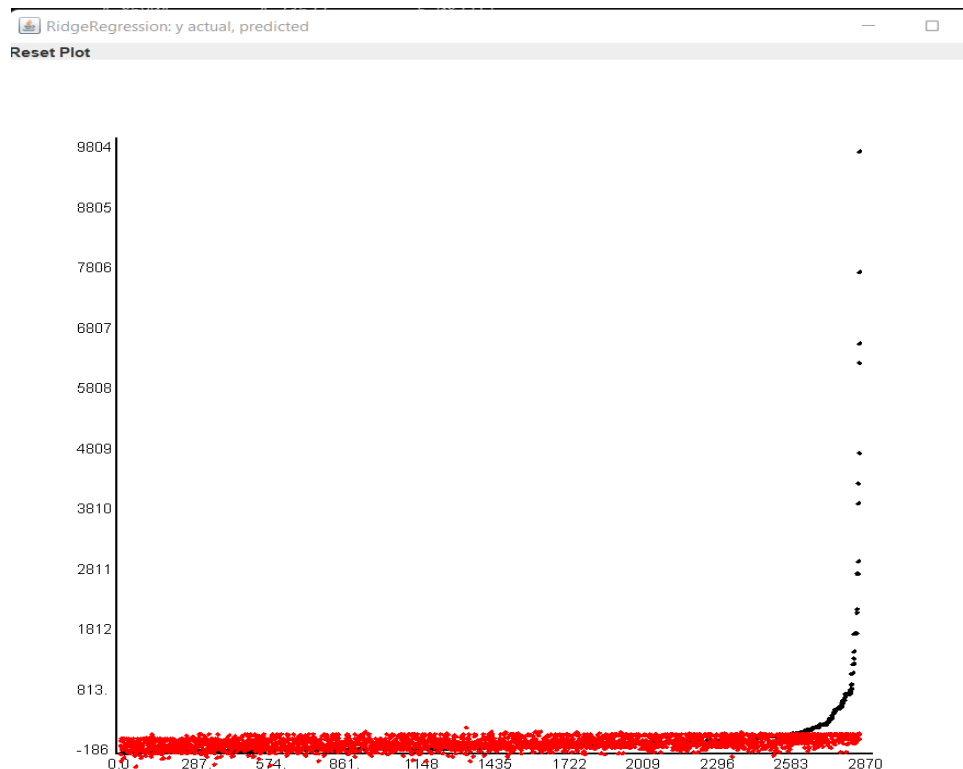
Ridge Regression:

         Multiple R-squared: 0.048879

         Adjusted R-squared: 0.046552

Symbolic Regression:

[info] REPORT

[info]   modelName mn = SymbolicRegression.quadratic

[info]   hparameter hp = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))

[info]   features   fn = Array(intercept, region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels, region^2, accommodation_type^2, yearly_availability^2, minimum_nights^2, number_of_reviews^2, reviews_per_month^2, owned_hotels^2)

[info]   parameter b = VectorD(380.487,   -170.932,   28.3262,   -5.56516e+13, 0.350423,   -0.250481,   -60.1795,   -2.05479,   22.5090,   -7.89439, 5.56516e+13,   -0.000540263,   0.00153794,   7.18678,   0.0065093)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.056665, rSqBar -> 0.052039, sst -> 473344895.742509, sse -> 446522666.146301, mse0 -> 155582.810504, rmse -> 394.439869, mae -> 139.477534**, dfm -> 14.000000, df -> 2855.000000, fStat -> 12.249813, aic -> -21197.683037, bic -> -21108.252027, mape -> 106.793957, smape -> 65.310552)

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var      Estimate      Std. Error      t value          Pr(>|t|)          VIF
[info] ------------------------------------------------------------------------------------------
[info]    x0    380.486725        NaN       NaN    0.000000    NA
[info]    x1    -170.932022       NaN       NaN    0.000000    17.309176
[info]    x2    28.326164         NaN       NaN    0.000000    79.760563
[info]    x3    -55651632671836.130000   NaN    NaN    0.000000         Infinity
[info]    x4    0.350423          NaN       NaN    0.000000    3.543752
[info]    x5    -0.250481         NaN       NaN    0.000000    6.196127
[info]    x6    -60.179486        NaN       NaN    0.000000    8.128142
[info]    x7    -2.054790         NaN       NaN    0.000000    5.334912
[info]    x8    22.508976         NaN       NaN    0.000000    17.134495
[info]    x9    -7.894391         NaN       NaN    0.000000    73.409289
[info]    x10   55651632671945.820000    NaN    NaN    0.000000         Infinity
[info]    x11   -0.000540         NaN       NaN    0.000000    3.379866
[info]    x12   0.001538          NaN       NaN    0.000000    4.664131
[info]    x13   7.186780          NaN       NaN    0.000000    6.059455
[info]    x14   0.006509          NaN       NaN    0.000000    4.898707
[info]    Residual standard error: 395.474694 on 2855.0 degrees of freedom
[info]    Multiple R-squared:  0.056665,      Adjusted R-squared:  0.052039
[info]    F-statistic: 12.249812560000395 on 14.0 and 2855.0 DF,  p-value: 0.0
[info] ------------------------------------------------------------------------------------------
```

Cross-validation:

```
[info] -----------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -----------------------------------------------
[info] |    name | num |        min |        max |        mean |    stdev |    interval |
[info] -----------------------------------------------
[info] |     rSq |  5 |      0.033 |      0.075 |       0.051 |    0.018 |      0.023 |
[info] |  rSqBar |  5 |      0.028 |      0.070 |       0.047 |    0.018 |      0.023 |
[info] |     sst |  5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |     sse |  5 | 60562171.612 | 149582876.103 | 90067557.337 | 38327769.315 | 47599533.784 |
[info] |    mse0 |  5 | 105509.010 | 260597.345 | 156912.121 | 66773.117 | 82926.017 |
[info] |    rmse |  5 |    324.822 |    510.487 |     389.555 |   80.307 |     99.734 |
[info] |     mae |  5 |    135.179 |    150.843 |     141.401 |    5.940 |      7.377 |
[info] |     dfm |  5 |     14.000 |     14.000 |      14.000 |    0.000 |      0.000 |
[info] |      df |  5 |   2855.000 |   2855.000 |    2855.000 |    0.000 |      0.000 |
[info] |   fStat |  5 |      6.874 |     16.534 |      11.124 |    4.160 |      5.167 |
[info] |     aic |  5 |  -4409.254 |  -4123.167 |   -4217.989 |  123.175 |    152.972 |
[info] |     bic |  5 |  -4343.965 |  -4057.877 |   -4152.699 |  123.175 |    152.972 |
[info] |    mape |  5 |    104.255 |    114.963 |     108.579 |    4.151 |      5.156 |
[info] |   smape |  5 |     64.422 |     67.400 |      66.332 |    1.392 |      1.728 |
[info] -----------------------------------------------
```


SymbolicRegression.quadratic: y actual, predicted

Reset Plot
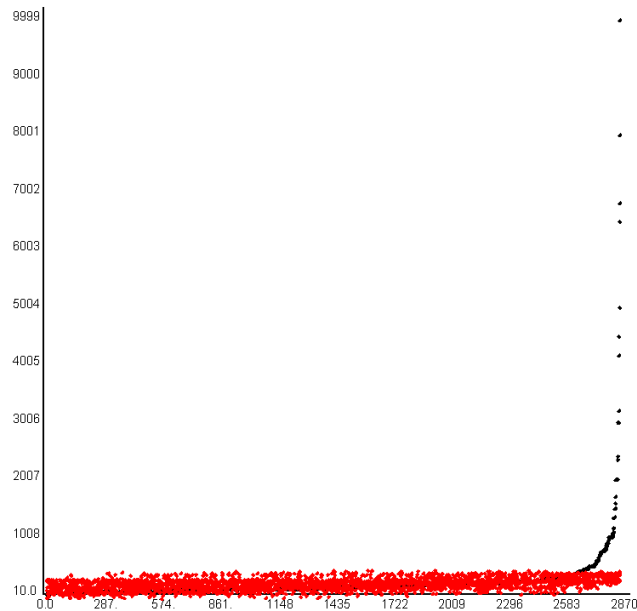
Symbolic Regression:

Multiple R-squared: 0.056665

Adjusted R-squared: 0.052039

Quadratic Regression:

[info] REPORT

[info]   modelName mn = PolyRegression

[info]   hparameter hp = HyperParameter (HashMap(factorization -> (Fac_Cholesky,Fac_Cholesky)))

[info]   features   fn = Array(x0, x1, x2)

[info]   parameter b   = VectorD(1.87193,   -0.0197617,   0.010120)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.000839, rSqBar -> 0.000142, sst -> 1600.093728, sse -> 1598.751508, mse0 -> 0.557056, rmse -> 0.746362, mae -> 0.609821**, dfm -> 2.000000, df -> 2867.000000, fStat -> 1.203485, aic -> -3226.750857, bic -> -3208.864655, mape -> 41.527495, smape -> 34.293227)

```
[info] SUMMARY
[info]     Parameters/Coefficients:
[info]     Var      Estimate     Std. Error        t value         Pr(>|t|)          VIF
[info]     ----------------------------------------------------------------------------
[info]     x0      1.871927      0.067571        27.703297        0.000000        NA
[info]     x1     -0.019762      0.069351        -0.284951        0.775682        16.707235
[info]     x2      0.010121      0.015585         0.649397        0.516082        16.707235
[info]     Residual standard error: 0.746752 on 2867.0 degrees of freedom
[info]     Multiple R-squared:  0.000839,       Adjusted R-squared:  0.000142
[info]     F-statistic: 1.2034849518511246 on 2.0 and 2867.0 DF,  p-value: 0.0
[info]     ----------------------------------------------------------------------------
[info]
```

Cross-Validation:

```
[info] ------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ------------------------------------------------
[info] |     name | num |      min |      max |     mean |    stdev | interval |
[info] ------------------------------------------------------------------------
[info] |      rSq |   5 |   -0.003 |    0.000 |   -0.002 |    0.001 |    0.002 |
[info] |   rSqBar |   5 |   -0.004 |   -0.000 |   -0.003 |    0.001 |    0.002 |
[info] |      sst |   5 |  313.401 |  325.814 |  319.778 |    4.820 |    5.985 |
[info] |      sse |   5 |  314.133 |  326.737 |  320.371 |    4.755 |    5.905 |
[info] |     mse0 |   5 |    0.547 |    0.569 |    0.558 |    0.008 |    0.010 |
[info] |     rmse |   5 |    0.740 |    0.754 |    0.747 |    0.006 |    0.007 |
[info] |      mae |   5 |    0.601 |    0.620 |    0.610 |    0.007 |    0.009 |
[info] |      dfm |   5 |    2.000 |    2.000 |    2.000 |    0.000 |    0.000 |
[info] |       df |   5 | 2867.000 | 2867.000 | 2867.000 |    0.000 |    0.000 |
[info] |    fStat |   5 |   -4.052 |    0.632 |   -2.655 |    1.868 |    2.319 |
[info] |      aic |   5 | -646.821 | -635.508 | -641.107 |    4.268 |    5.300 |
[info] |      bic |   5 | -633.763 | -622.450 | -628.049 |    4.268 |    5.300 |
[info] |     mape |   5 |   40.450 |   43.270 |   41.570 |    1.068 |    1.326 |
[info] |    smape |   5 |   33.906 |   35.196 |   34.316 |    0.515 |    0.639 |
[info] ------------------------------------------------------------------------
```
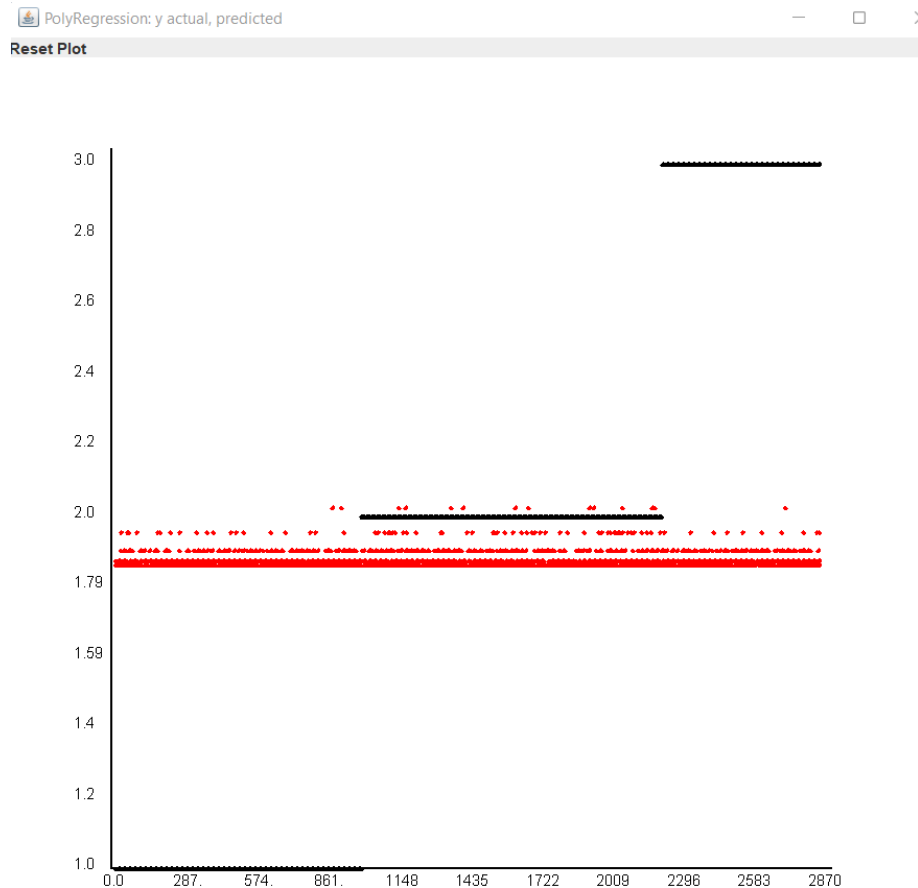
Ridge Regression:
                Multiple R-squared: 0.000839
                Adjusted R-squared: 0.000142

Symbolic Lasso Regression:

[info] REPORT

[info]   modelName mn = SymLassoRegression.quadratic

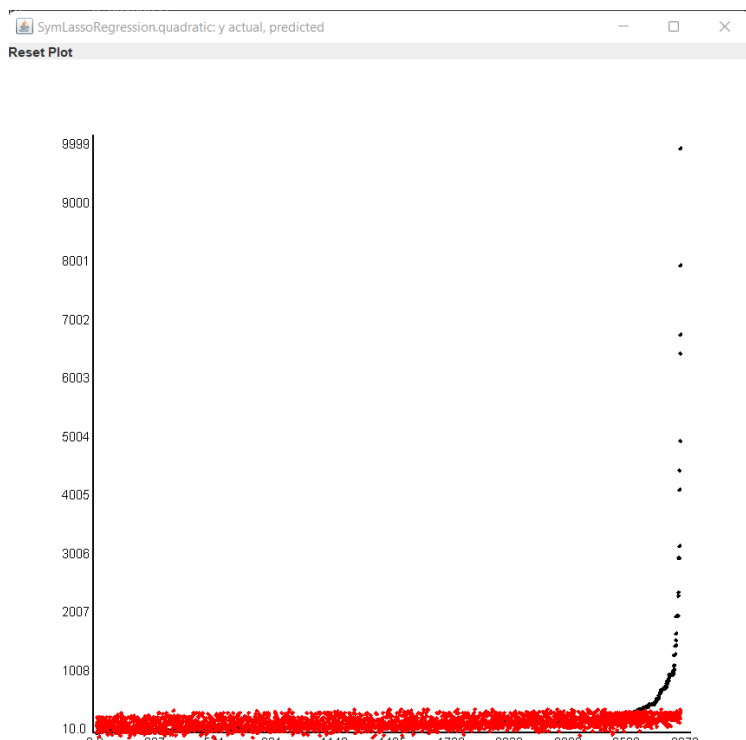[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01)))

[info]   features   fn = Array(region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels, region^2, accommodation_type^2, yearly_availability^2, minimum_nights^2, number_of_reviews^2, reviews_per_month^2, owned_hotels^2)
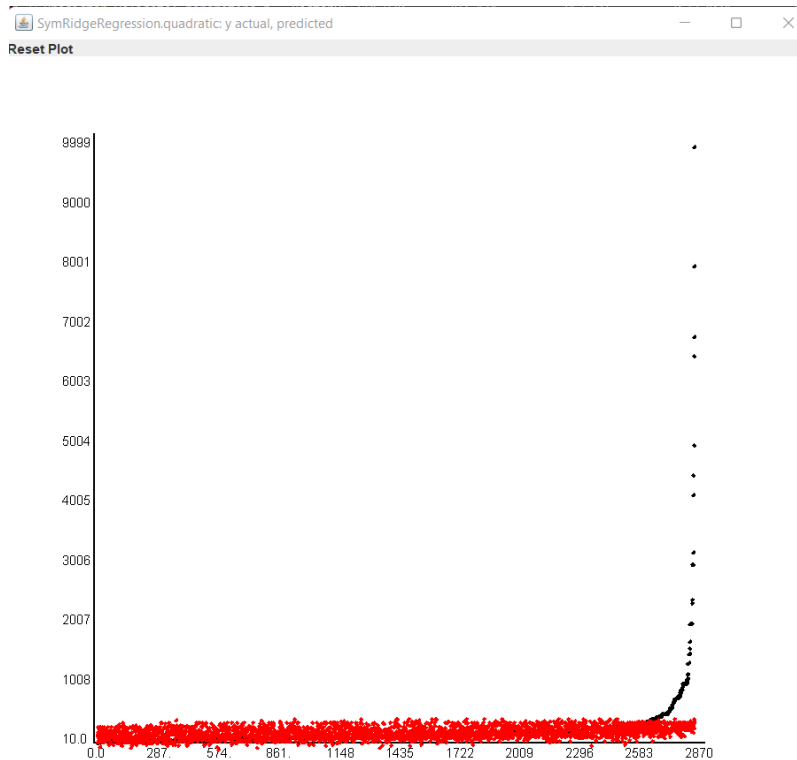
[info]   parameter b = VectorD(-93.5917,   376.349,   26.1912,   0.373307, -0.130269,   -44.6193,   -1.65233,   6.67263,   -90.2034,   26.1912, -0.000582570, 0.000987059,   5.30805,   0.0054025)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.050100, rSqBar -> 0.045776, sst -> 473344895.742509, sse -> 449630542.789972, mse0 -> 156665.694352, rmse -> 395.810175, mae -> 139.209646,** dfm -> 13.000000, df -> 2856.000000, fStat -> 11.586982, aic -> -21209.636294, bic -> -21126.167352, mape -> 106.118428, smape -> 64.262340)

```
[info] -------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] -------------------------------------------------
[info] |     name | num |         min |         max |        mean |      stdev |   interval |
[info] -------------------------------------------------
[info] |      rSq |   5 |       0.031 |       0.068 |       0.045 |      0.015 |      0.019 |
[info] |   rSqBar |   5 |       0.026 |       0.064 |       0.040 |      0.015 |      0.019 |
[info] |      sst |   5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |      sse |   5 | 61008561.271 | 150517814.952 | 90641793.955 | 38346216.804 | 47622443.854 |
[info] |     mse0 |   5 |  106286.692 |  262226.158 |  157912.533 |  66805.256 |  82965.930 |
[info] |     rmse |   5 |     326.016 |     512.080 |     390.887 |     79.998 |     99.350 |
[info] |      mae |   5 |     134.522 |     150.629 |     140.623 |      6.005 |      7.458 |
[info] |      dfm |   5 |      13.000 |      13.000 |      13.000 |      0.000 |      0.000 |
[info] |       df |   5 |    2856.000 |    2856.000 |    2856.000 |      0.000 |      0.000 |
[info] |    fStat |   5 |       6.972 |      16.074 |      10.280 |      3.732 |      4.635 |
[info] |      aic |   5 |   -4412.916 |   -4127.232 |   -4221.811 |    122.389 |    151.995 |
[info] |      bic |   5 |   -4351.979 |   -4066.295 |   -4160.875 |    122.389 |    151.995 |
[info] |     mape |   5 |     103.255 |     114.642 |     107.370 |      4.494 |      5.581 |
[info] |    smape |   5 |      62.595 |      66.148 |      64.796 |      1.397 |      1.735 |
[info] -------------------------------------------------
```


Symbolic Lasso Regression:

Multiple R-squared: 0.050100

Adjusted R-squared: 0.045776

Symbolic Ridge Regression:

[info] REPORT

[info]   modelName mn = SymRidgeRegression.quadratic

[info]   hparameter hp = HyperParameter (HashMap(lambda -> (0.01,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky)))

[info]   features   fn = Array(region, accommodation_type, yearly_availability, minimum_nights, number_of_reviews, reviews_per_month, owned_hotels, region^2, accommodation_type^2, yearly_availability^2, minimum_nights^2, number_of_reviews^2, reviews_per_month^2, owned_hotels^2)

[info]   parameter b   = VectorD(-93.5564,   376.299,   26.1966,   0.373350, -0.130273,   -44.6175,   -1.65224,   6.66527,   -90.1907,   26.1966, -0.000582615, 0.000987056,   5.30785,   0.0054022)

[info]   fitMap   qof = LinkedHashMap(**rSq -> 0.050100, rSqBar -> 0.045442, sst -> 473344895.742509, sse -> 449630543.023248, mse0 -> 156665.694433, rmse -> 395.810175, mae -> 139.208477**, dfm -> 14.000000, df -> 2855.000000, fStat -> 10.755573, aic -> -21207.636295, bic -> -21118.205285, mape -> 106.117717, smape -> 64.261292)

```
[info] --------------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] --------------------------------------------------
[info] |    name | num |        min |         max |         mean |       stdev |    interval |
[info] --------------------------------------------------
[info] |     rSq |   5 |      0.031 |       0.068 |        0.045 |       0.015 |       0.019 |
[info] |  rSqBar |   5 |      0.026 |       0.064 |        0.040 |       0.015 |       0.019 |
[info] |     sst |   5 | 65472382.329 | 156684138.578 | 94645824.164 | 39425546.436 | 48962871.126 |
[info] |     sse |   5 | 61008752.764 | 150517860.398 | 90641807.607 | 38346144.710 | 47622354.320 |
[info] |    mse0 |   5 | 106287.026 | 262226.238 | 157912.557 | 66805.130 | 82965.774 |
[info] |    rmse |   5 |    326.017 |     512.080 |      390.887 |      79.997 |      99.349 |
[info] |     mae |   5 |    134.521 |     150.628 |      140.622 |       6.005 |       7.458 |
[info] |     dfm |   5 |     14.000 |      14.000 |       14.000 |       0.000 |       0.000 |
[info] |      df |   5 |   2855.000 |    2855.000 |     2855.000 |       0.000 |       0.000 |
[info] |   fStat |   5 |      6.471 |      14.920 |        9.543 |       3.464 |       4.302 |
[info] |     aic |   5 |  -4410.916 |   -4125.232 |    -4219.812 |     122.388 |     151.995 |
[info] |     bic |   5 |  -4345.627 |   -4059.943 |    -4154.522 |     122.388 |     151.995 |
[info] |    mape |   5 |    103.255 |     114.641 |      107.369 |       4.494 |       5.581 |
[info] |   smape |   5 |     62.594 |      66.147 |       64.795 |       1.397 |       1.735 |
[info] --------------------------------------------------
```


SymRidgeRegression.quadratic: y actual, predicted

Reset Plot

Symbolic Lasso Regression:

        Multiple R-squared: 0.050100

        Adjusted R-squared: 0.045442

Auto MPG:

Data Set Information:

This dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg", 8 of the original instances were removed because they had unknown values for the "mpg" attribute. The original dataset is available in the file "auto-mpg.data-original".

Attribute Information:

1. mpg: continuous

2. cylinders: multi-valued discrete

3. displacement: continuous

4. horsepower: continuous

5. weight: continuous

6. acceleration: continuous

7. model year: multi-valued discrete

8. origin: multi-valued discrete

9. car name: string (unique for each instance)
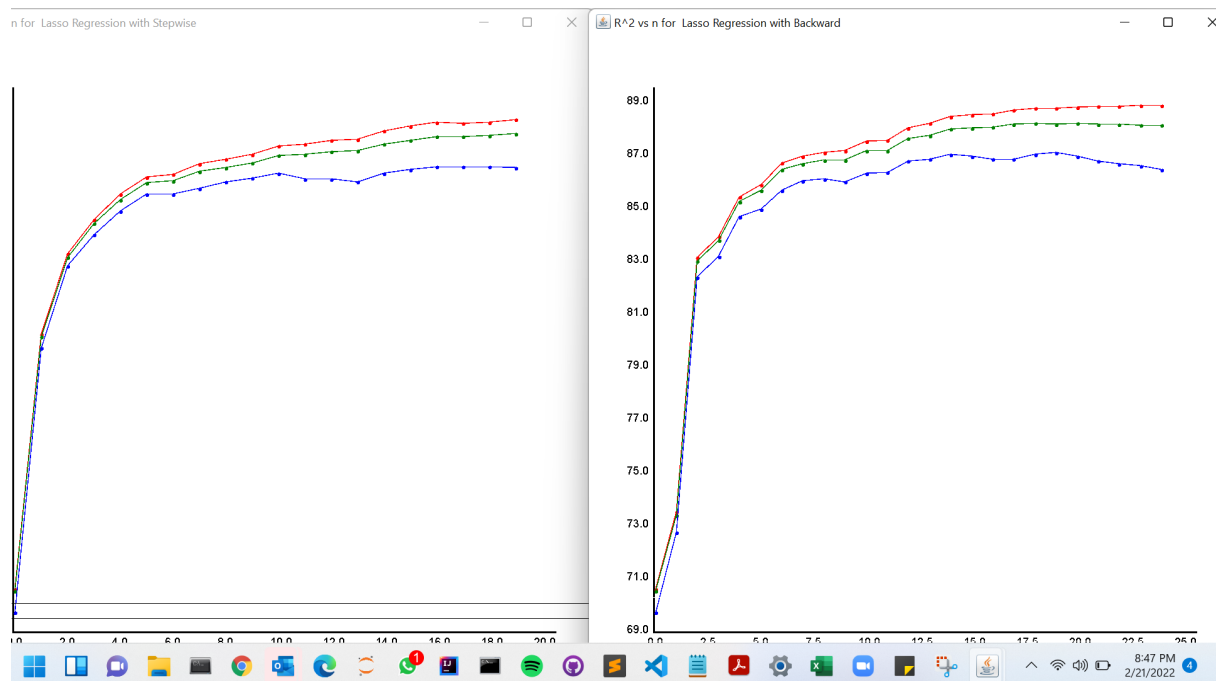
Scala Methods to Execute:

      LasoTest9
      symLassoRegressionTest11
      regressionTest9
      QuadraticTest9
      RidgeRegressionTest9
      symbolicRegressionTest9
      symRidgeRegressionTest11

Models:

Linear Regression:
      Multiple R-squared: 0.792574
      Adjusted R-squared: 0.789341

Lasso Regression
      Multiple R-squared: 0.804643,
      Adjusted R-squared: 0.802112

Ridge Regression

      Multiple R-squared: 0.804643,
      Adjusted R-squared: 0.801598

Quadratic Regression

```
[info] ----------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ----------------------------------------------
[info] |    name | num |       min |       max |      mean |    stdev |  interval |
[info] ----------------------------------------------
[info] |     rSq |   5 |     0.888 |     0.923 |     0.903 |    0.013 |     0.016 |
[info] |  rSqBar |   5 |     0.888 |     0.923 |     0.903 |    0.013 |     0.016 |
[info] |     sst |   5 | 723406.718| 981947.295| 847932.772| 94732.611| 117649.115|
[info] |     sse |   5 | 64848.888 | 94144.564 | 81567.946 | 11218.803| 13932.713 |
[info] |    mse0 |   5 |   831.396 |  1206.982 |  1045.743 |  143.831 |   178.625 |
[info] |    rmse |   5 |    28.834 |    34.742 |    32.274 |    2.265 |     2.813 |
[info] |     mae |   5 |    23.021 |    27.148 |    25.222 |    1.835 |     2.279 |
[info] |     dfm |   5 |     2.000 |     2.000 |     2.000 |    0.000 |     0.000 |
[info] |      df |   5 |   389.000 |   389.000 |   389.000 |    0.000 |     0.000 |
[info] |   fStat |   5 |  1547.303 |  2333.118 |  1849.146 |  292.714 |   363.524 |
[info] |     aic |   5 |  -381.941 |  -367.705 |  -375.829 |    5.452 |     6.770 |
[info] |     bic |   5 |  -374.870 |  -360.635 |  -368.759 |    5.452 |     6.770 |
[info] |    mape |   5 |    13.382 |    16.794 |    14.759 |    1.413 |     1.755 |
[info] |   smape |   5 |    12.897 |    15.686 |    14.317 |    1.201 |     1.491 |
[info] ----------------------------------------------
```

```
fname = Array(x0, x1, x2, x3)
SUMMARY
    Parameters/Coefficients:
    Var      Estimate    Std. Error      t value       Pr(>|t|)          VIF
    ------------------------------------------------------------------------
    x0    -1.678041      0.015371    -109.167715      0.000000          NA
    x1    -0.272654      0.008015     -34.019350      0.000000          3.899141
    x2     0.502796      0.000416    1209.082517      0.000000          0.009250
    x3    -0.099925      0.004406     -22.677139      0.000000          1.556891
    Residual standard error: 5.049825 on 9564.0 degrees of freedom
    Multiple R-squared:  0.918517,       Adjusted R-squared:  0.918483
    F-statistic: 26952.511083359277 on 4.0 and 9564.0 DF,  p-value: 0.0
    ------------------------------------------------------------------------


REPORT
    modelName  mn  = RidgeRegression
    hparameter hp  = HyperParameter (HashMap(lambda -> (1.0,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky))
    features   fn  = Array(x0, x1, x2, x3)
    parameter  b   = VectorD(-1.67804,   -0.272654,      0.502796,      -0.099924)
    fitMap     qof = LinkedHashMap(rSq -> 0.918517, rSqBar -> 0.918483, sst -> 2993123.926550, sse -> 243888.9909
e0 -> 25.487406, rmse -> 5.048505, mae -> 4.019090, dfm -> 4.000000, df -> 9564.000000, fStat -> 26952.511083, aic
```

Multiple R-squared: 0.898704,
Adjusted R-squared: 0.898183

Symbolic Regression
        Multiple R-squared: 0.881253
        Adjusted R-squared: 0.874851

Ridge Symbolic Regression
        Multiple R-squared: 0.889053
        Adjusted R-squared: 0.881150

Lasso Symbolic Regression
        Multiple R-squared: 0.889053,
        Adjusted R-squared: 0.881474

Best Model Quadratic Regression **Based on R2**

FoldsCpp

The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (EP) of the plant. A combined cycle power plant (CCPP) is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the Vacuum is colected from and has effect on the Steam Turbine, he other three of the ambient variables effect the GT performance.

For comparability with our baseline studies, and to allow 5x2 fold statistical tests be carried out, we provide the data shuffled five times. For each shuffling 2-fold CV is carried out and the resulting 10 measurements are used for statistical testing.

```
info] stepRegressionAll: features in/out   = ArrayBuffer(5, 3, 7, 1, 2, 9, 24, 25, 8, 16, 4, 23, 22, 12, 6, 15, 21, 7,
0, 19, 9, 18, 10, 13)
info] k = 21, n = 6
info] Run + title
info] x-axis: minX = 0.0, maxX = 20.0
info] y-axis: minY = 69.0, maxY = 89.0
info] Stepwise: rSq =
info] MatrixD(70.6260, 70.4750,        69.7413,
info]         80.2612,        80.1085,        79.7154,
info]         83.3067,        83.1341,        82.8338,
info]         84.6036,        84.4042,        84.0090,
info]         85.5426,        85.3173,        84.9240,
info]         86.1867,        85.9349,        85.5414,
info]         86.2924,        86.0061,        85.5695,
info]         86.6843,        86.3706,        85.7534,
info]         86.8779,        86.5335,        86.0013,
info]         87.0722,        86.6980,        86.1777,
info]         87.3721,        86.9723,        86.3406,
info]         87.4492,        87.0176,        86.1168,
info]         87.5822,        87.1211,        86.1419,
info]         87.6392,        87.1461,        86.0202,
info]         87.9441,        87.4297,        86.3310,
info]         88.1138,        87.5735,        86.4840,
info]         88.2580,        87.6914,        86.5743,
info]         88.2479,        87.7138,        86.5770,
info]         88.2861,        87.7208,        86.5971,
info]         88.3907,        87.7978,        86.5435)
```

Scala Methods to Execute:

       regressionTest8
       LasoTest8
       QuadraticTest8
       RidgeRegressionTest8
       symbolicRegressionTest11
       symRidgeRegressionTest10
       symLassoRegressionTest10

Models:

Linear Regression:
       Multiple R-squared: 0.918517
       Adjusted R-squared: 0.918491

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var     Estimate    Std. Error      t value       Pr(>|t|)       VIF
[info]    -----------------------------------------------------------------------
[info]    x0     -1.678056    0.015370     -109.174372     0.000000        NA
[info]    x1     -0.272647    0.008014      -34.020349     0.000000      3.899141
[info]    x2      0.502796    0.000416     1209.146016     0.000000      0.009250
[info]    x3     -0.099927    0.004406      -22.678864     0.000000      1.556891
[info]    Residual standard error: 5.049561 on 9565.0 degrees of freedom
[info]    Multiple R-squared:  0.918517,      Adjusted R-squared:  0.918491
[info]    F-statistic: 35940.43894310781 on 3.0 and 9565.0 DF,  p-value: 0.0
[info]    -----------------------------------------------------------------------
[info]
[info] REPORT
[info]    modelName  mn  = Regression
[info]    hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]    features   fn  = Array(x0, x1, x2, x3)
[info]    parameter  b   = VectorD(-1.67806,  -0.272647,    0.502796,   -0.099927)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.918517, rSqBar -> 0.918491, sst -> 2993123.926550, sse -> 243888.990904, mse0 -> 25.487406, rmse -> 5.048505, mae -> 4.01
9089, dfm -> 3.000000, df -> 9565.000000, fStat -> 35940.438943, aic -> -29062.916372, bic -> -29034.251236, mape -> NaN, smape -> NaN)
```

```
[info]    -------------------------------------------------------
[info]    | showQofStatTable: Statistical Table for QoF |
[info]    -------------------------------------------------------
[info]    |     name |  num |        min |        max |       mean |     stdev |   interval |
[info]    ----------------------------------------------------------------------------------
[info]    |      rSq |    5 |      0.910 |      0.934 |      0.917 |     0.009 |      0.012 |
[info]    |   rSqBar |    5 |      0.910 |      0.934 |      0.917 |     0.009 |      0.012 |
[info]    |      sst |    5 | 549043.244 | 755436.033 | 597905.395 | 88255.289 | 109604.881 |
[info]    |      sse |    5 |  47882.308 |  50201.425 |  48834.736 |   974.055 |   1209.686 |
[info]    |     mse0 |    5 |     25.030 |     26.242 |     25.528 |     0.509 |      0.632 |
[info]    |     rmse |    5 |      5.003 |      5.123 |      5.052 |     0.050 |      0.062 |
[info]    |      mae |    5 |      3.996 |      4.052 |      4.022 |     0.020 |      0.025 |
[info]    |      dfm |    5 |      3.000 |      3.000 |      3.000 |     0.000 |      0.000 |
[info]    |       df |    5 |   9565.000 |   9565.000 |   9565.000 |     0.000 |      0.000 |
[info]    |    fStat |    5 |  32240.928 |  44790.023 |  35792.136 |  5089.330 |   6320.476 |
[info]    |      aic |    5 |  -5832.081 |  -5786.585 |  -5805.270 |    19.109 |     23.731 |
[info]    |      bic |    5 |  -5809.855 |  -5764.360 |  -5783.044 |    19.109 |     23.731 |
[info]    |     mape |    5 |      0.879 |      0.889 |        NaN |       NaN |        NaN |
[info]    |    smape |    5 |      0.878 |      0.889 |        NaN |       NaN |        NaN |
[info]    ----------------------------------------------------------------------------------
[success] Total time: 11 s, completed Feb 21, 2022, 10:36:34 PM
```

Lasso Regression
Multiple R-squared: 0.918517
Adjusted R-squared: 0.918491

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var     Estimate    Std. Error      t value       Pr(>|t|)       VIF
[info]    -----------------------------------------------------------------------
[info]    x0     -1.678056    0.015370     -109.174368     0.000000        NA
[info]    x1     -0.272647    0.008014      -34.020352     0.000000      3.899141
[info]    x2      0.502796    0.000416     1209.146013     0.000000      0.009250
[info]    x3     -0.099927    0.004406      -22.678860     0.000000      1.556891
[info]    Residual standard error: 5.049561 on 9565.0 degrees of freedom
[info]    Multiple R-squared:  0.918517,      Adjusted R-squared:  0.918491
[info]    F-statistic: 35940.438943107714 on 3.0 and 9565.0 DF,  p-value: 0.0
[info]    -----------------------------------------------------------------------
[info]
[info] REPORT
[info]    modelName  mn  = LassoRegression
[info]    hparameter hp  = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]    features   fn  = Array(x0, x1, x2, x3)
[info]    parameter  b   = VectorD(-1.67806,  -0.272647,    0.502796,   -0.099927)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.918517, rSqBar -> 0.918491, sst -> 2993123.926550, sse -> 243888.990904, mse0 -> 25.487406, rmse -> 5.048505, mae -> 4.01
9089, dfm -> 3.000000, df -> 9565.000000, fStat -> 35940.438943, aic -> -29062.916372, bic -> -29034.251236, mape -> NaN, smape -> NaN)
[info]
```

```
[info]    -------------------------------------------------------
[info]    | showQofStatTable: Statistical Table for QoF |
[info]    -------------------------------------------------------
[info]    |     name |  num |        min |        max |       mean |     stdev |   interval |
[info]    ----------------------------------------------------------------------------------
[info]    |      rSq |    5 |      0.910 |      0.934 |      0.917 |     0.009 |      0.012 |
[info]    |   rSqBar |    5 |      0.910 |      0.934 |      0.917 |     0.009 |      0.012 |
[info]    |      sst |    5 | 549043.244 | 755436.033 | 597905.395 | 88255.289 | 109604.881 |
[info]    |      sse |    5 |  47882.308 |  50201.425 |  48834.736 |   974.055 |   1209.686 |
[info]    |     mse0 |    5 |     25.030 |     26.242 |     25.528 |     0.509 |      0.632 |
[info]    |     rmse |    5 |      5.003 |      5.123 |      5.052 |     0.050 |      0.062 |
[info]    |      mae |    5 |      3.996 |      4.052 |      4.022 |     0.020 |      0.025 |
[info]    |      dfm |    5 |      3.000 |      3.000 |      3.000 |     0.000 |      0.000 |
[info]    |       df |    5 |   9565.000 |   9565.000 |   9565.000 |     0.000 |      0.000 |
[info]    |    fStat |    5 |  32240.928 |  44790.024 |  35792.136 |  5089.330 |   6320.476 |
[info]    |      aic |    5 |  -5832.081 |  -5786.585 |  -5805.270 |    19.109 |     23.731 |
[info]    |      bic |    5 |  -5809.855 |  -5764.360 |  -5783.044 |    19.109 |     23.731 |
[info]    |     mape |    5 |      0.879 |      0.889 |        NaN |       NaN |        NaN |
[info]    |    smape |    5 |      0.878 |      0.889 |        NaN |       NaN |        NaN |
[info]    ----------------------------------------------------------------------------------
```

Ridge Regression

```
[info] SUMMARY
[info]    Parameters/Coefficients:
[info]    Var      Estimate    Std. Error      t value        Pr(>|t|)        VIF
[info] ----------------------------------------------------------------------------
[info]    x0      -1.678041    0.015371      -109.167715     0.000000        NA
[info]    x1      -0.272654    0.008015       -34.019350     0.000000        3.899141
[info]    x2       0.502796    0.000416      1209.082517     0.000000        0.009250
[info]    x3      -0.099925    0.004406       -22.677139     0.000000        1.556891
[info]    Residual standard error: 5.049825 on 9564.0 degrees of freedom
[info]    Multiple R-squared:  0.918517,      Adjusted R-squared:  0.918483
[info]    F-statistic: 26952.511083359277 on 4.0 and 9564.0 DF,  p-value: 0.0
[info] ----------------------------------------------------------------------------
[info]
[info] REPORT
[info]    modelName  mn  = RidgeRegression
[info]    hparameter hp  = HyperParameter (HashMap(lambda -> (1.0,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky))
[info]    features   fn  = Array(x0, x1, x2, x3)
[info]    parameter  b   = VectorD(-1.67804,   -0.272654,      0.502796,      -0.099924)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.918517, rSqBar -> 0.918483, sst -> 2993123.926550, sse -> 243888.9909
27, mse0 -> 25.487406, rmse -> 5.048505, mae -> 4.019090, dfm -> 4.000000, df -> 9564.000000, fStat -> 26952.511083, aic
 -> -29060.916372, bic -> -29025.084952, mape -> NaN, smape -> NaN)
[info]
```

Multiple R-squared: 0.918517
Adjusted R-squared: 0.918483

Quadratic Regression

PolyRegression: y actual, predicted

Reset Plot
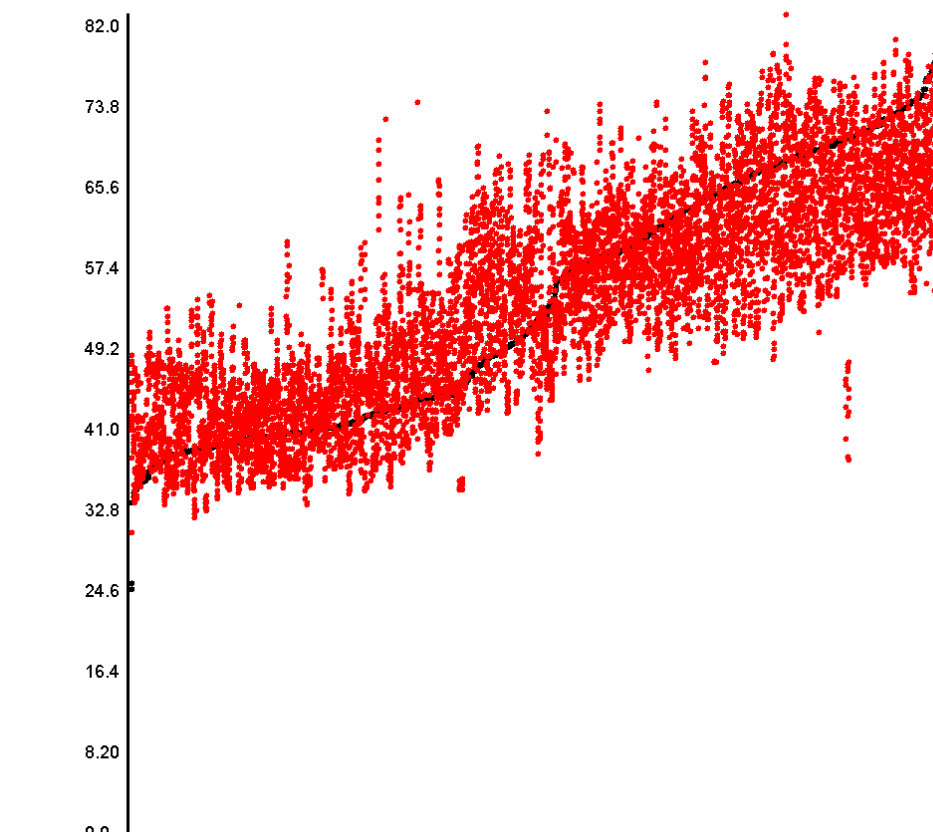
Switch max and min value

Multiple R-squared: 0.722751
Adjusted R-squared: 0.722693

Symbolic Regression

Multiple R-squared: 0.940227
Adjusted R-squared: 0.940170

Ridge Symbolic Regression

```
[info] ---------------------------------------------
[info] | showQofStatTable: Statistical Table for QoF |
[info] ---------------------------------------------
[info] |      name | num |       min |       max |      mean |     stdev |  interval |
[info] ---------------------------------------------
[info] |       rSq |   5 |     0.926 |     0.946 |     0.933 |     0.007 |     0.009 |
[info] |    rSqBar |   5 |     0.926 |     0.946 |     0.933 |     0.007 |     0.009 |
[info] |       sst |   5 | 549043.244 | 755436.033 | 597905.395 | 88255.289 | 109604.881 |
[info] |       sse |   5 | 38125.965 | 41044.776 | 39329.414 |  1476.752 |  1833.989 |
[info] |      mse0 |   5 |    19.930 |    21.456 |    20.559 |     0.772 |     0.959 |
[info] |      rmse |   5 |     4.464 |     4.632 |     4.534 |     0.085 |     0.105 |
[info] |       mae |   5 |     3.568 |     3.637 |     3.597 |     0.030 |     0.037 |
[info] |       dfm |   5 |     8.000 |     8.000 |     8.000 |     0.000 |     0.000 |
[info] |        df |   5 |  9560.000 |  9560.000 |  9560.000 |     0.000 |     0.000 |
[info] |     fStat |   5 | 14870.169 | 20799.177 | 16944.541 |  2248.522 |  2792.456 |
[info] |       aic |   5 | -5630.054 | -5558.870 | -5588.220 |    36.015 |    44.727 |
[info] |       bic |   5 | -5580.046 | -5508.862 | -5538.212 |    36.015 |    44.727 |
[info] |      mape |   5 |     0.785 |     0.801 |       NaN |       NaN |       NaN |
```

Multiple R-squared: 0.934456,
Adjusted R-squared: 0.934401

Lasso Symbolic Regression

```
rname   Array(x0, x1, x2, x3, x4_v_x1, x1_v_x1, x1_x2_x1, x1_x2_v)
SUMMARY
    Parameters/Coefficients:
    Var     Estimate      Std. Error      t value        Pr(>|t|)       VIF
    ---------------------------------------------------------------------------------
    x0      1.124533      0.114373        9.832133       0.000000       NA
    x1      1.181933      0.062586        18.884876      0.000000       295.495848
    x2      0.429808      0.002535        169.545283     0.000000       0.427206
    x3      1.092398      0.031154        35.064556      0.000000       96.718200
    x4     -0.000052      0.000002        -25.804227     0.000000       760.425537
    x5      0.000883      0.000026        34.240191      0.000000       529.005085
    x6     -0.000048      0.000001        -31.857766     0.000000       236.374465
    x7     -0.000023      0.000001        -30.586820     0.000000       291.846128
    Residual standard error: 4.529796 on 9561.0 degrees of freedom
    Multiple R-squared:  0.934456,       Adjusted R-squared:  0.934408
    F-statistic: 19472.787284755657 on 7.0 and 9561.0 DF,  p-value: 0.0
    ---------------------------------------------------------------------------------
```

Multiple R-squared: 0.934456
Adjusted R-squared: 0.934408

Best Model Symbolic Regression **Based on R2**

USA Housing Dataset:

We have taken this dataset from Kaggle. The main motive is to predict the price of the house. This dataset contains 6 attributes which are - Avg. Area Income, Avg. Area House Age,Avg. Area Number of Rooms,Avg. Area Number of Bedrooms,Area Population,Price.

Linear Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var      Estimate     Std. Error       t value         Pr(>|t|)          VIF
------------------------------------------------------------------------------
    x0     10.237202      0.268622       38.110015        0.000000          NA
    x1     51296.068671   2960.620169    17.326123        0.000000          0.740165
    x2     -9614.433350   3262.670492    -2.946799        0.003211          0.927345
    x3     4082.312867    3132.019425    1.303412         0.192434          1.276879
    x4     8.091825       0.326978       24.747299        0.000000          0.900695
    Residual standard error: 242106.334210 on 4996.0 degrees of freedom
    Multiple R-squared:  0.531341,       Adjusted R-squared:  0.530966
    F-statistic: 1416.0525157356897 on 4.0 and 4996.0 DF,   p-value: 0.0
------------------------------------------------------------------------------
```

```
[info] REPORT
[info]      modelName  mn  = Regression
[info]      hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]      features   fn  = Array(x0, x1, x2, x3, x4)
[info]      parameter  b   = VectorD(10.2372,    51296.1,      -9614.43,     4082.31,      8.0918)
[info]      fitMap     qof = LinkedHashMap(rSq -> 0.531341, rSqBar -> 0.530966, sst -> 624853298447502.600000, sse -> 292
842923414396.300000, mse0 -> 58556873308.217620, rmse -> 241985.274982, mae -> 193470.071203, dfm -> 4.000000, df -> 499
6.000000, fStat -> 1416.052516, aic -> -69081.710487, bic -> -69049.123521, mape -> NaN, smape -> NaN)
```
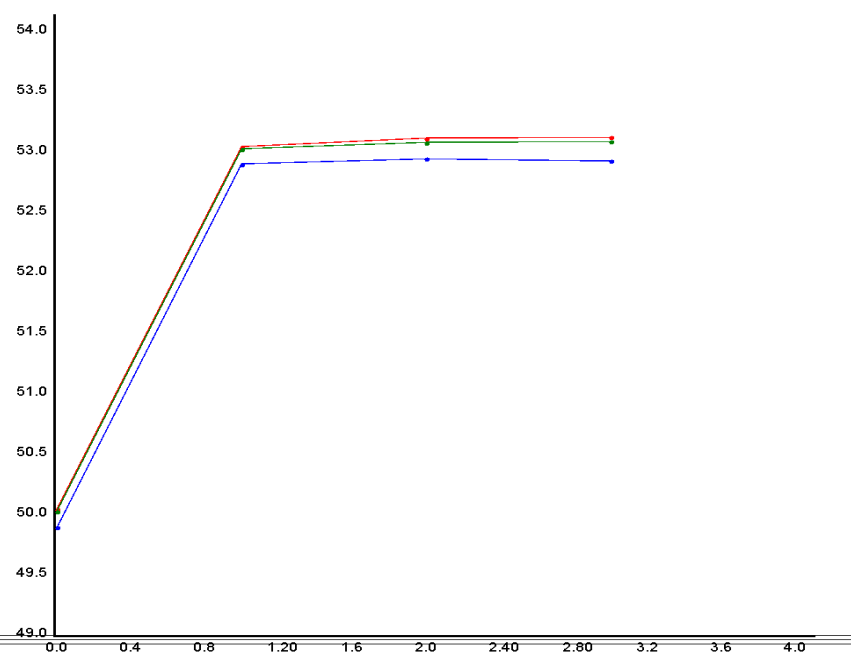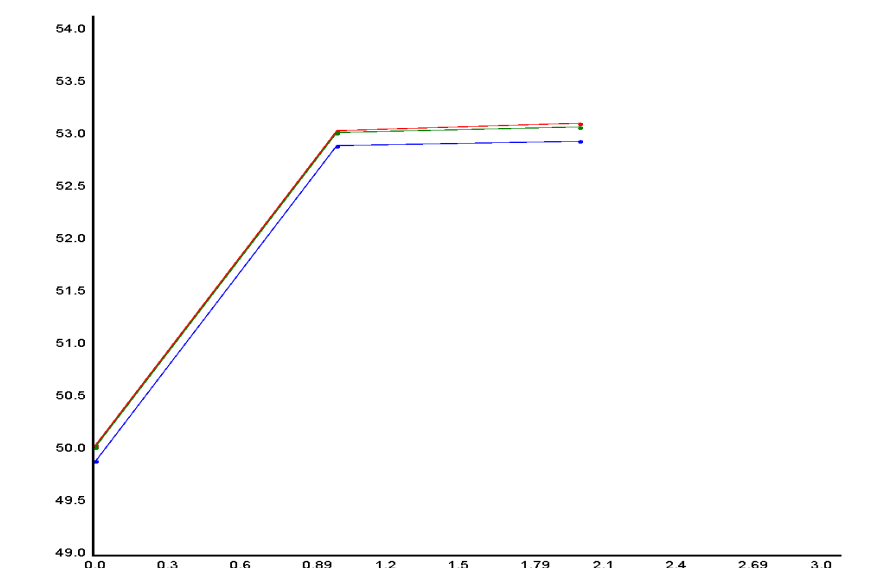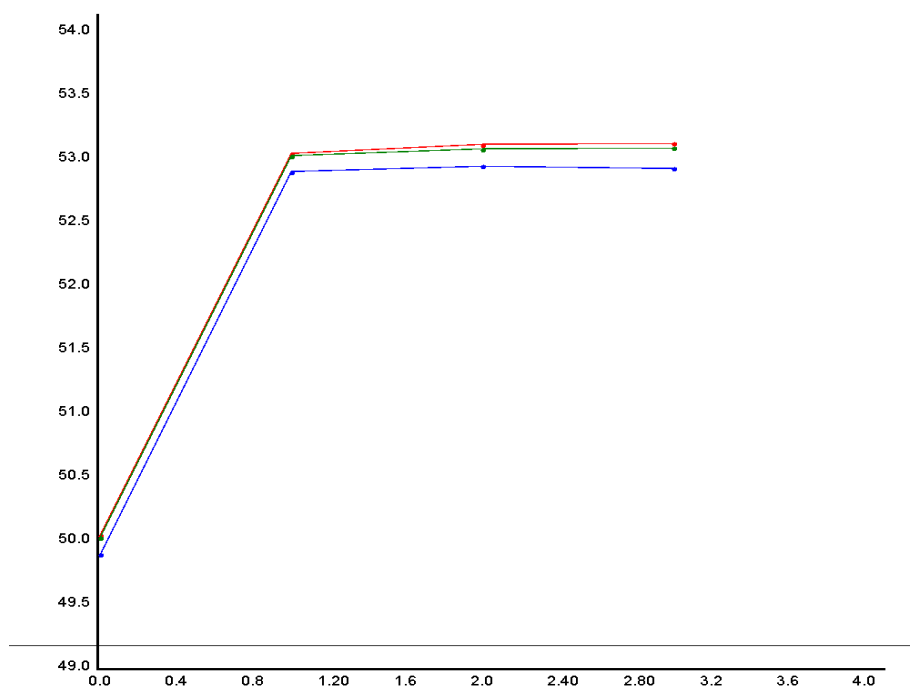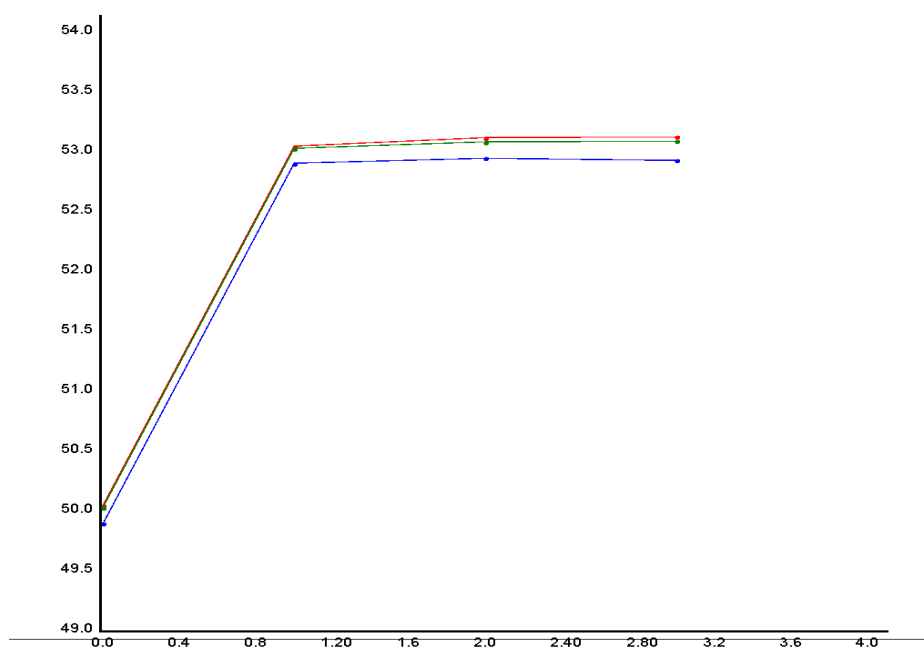

R^2 vs n for Regression with Forward
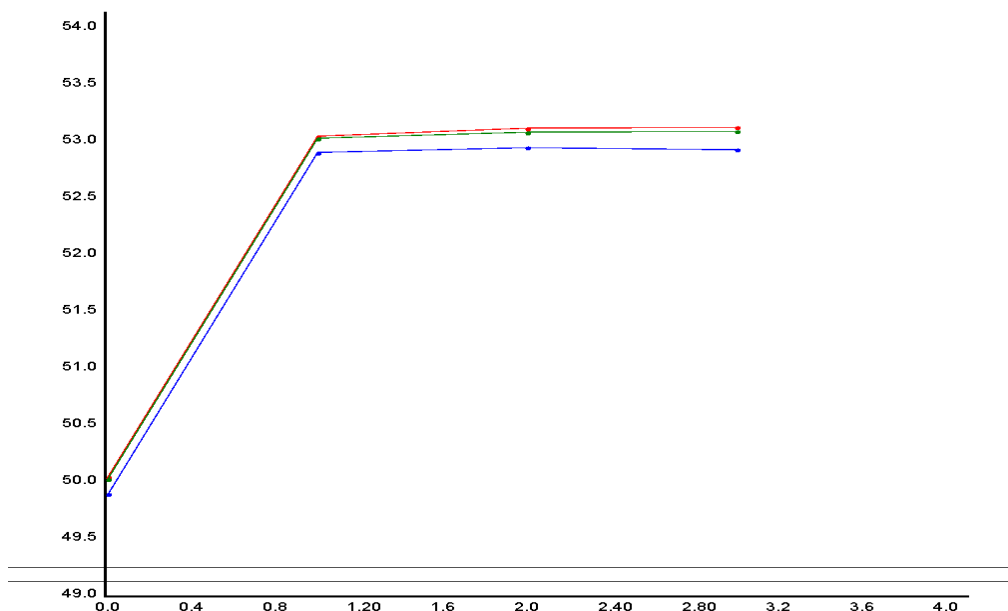
R^2 vs n for Regression with Backward
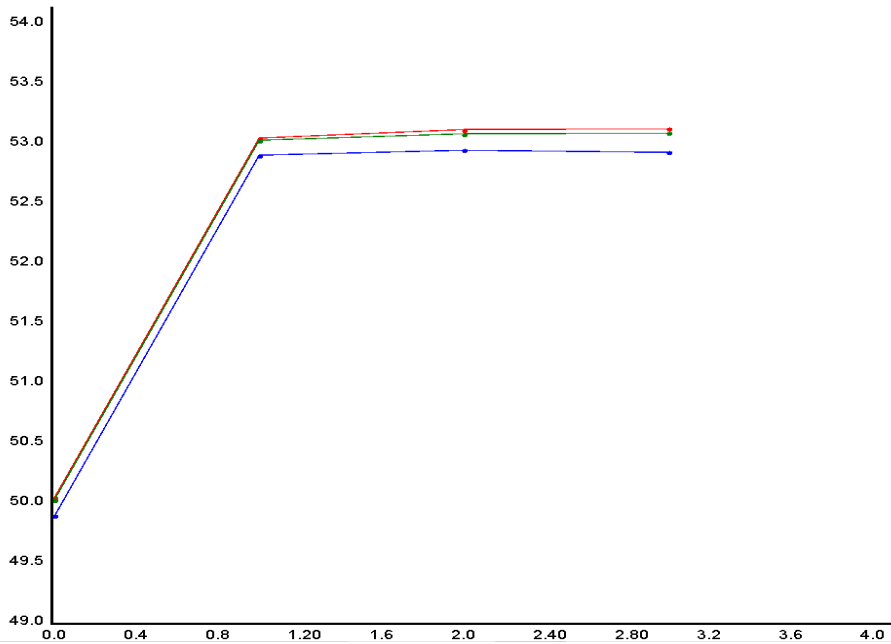
R^2 vs n for Regression with Stepwise
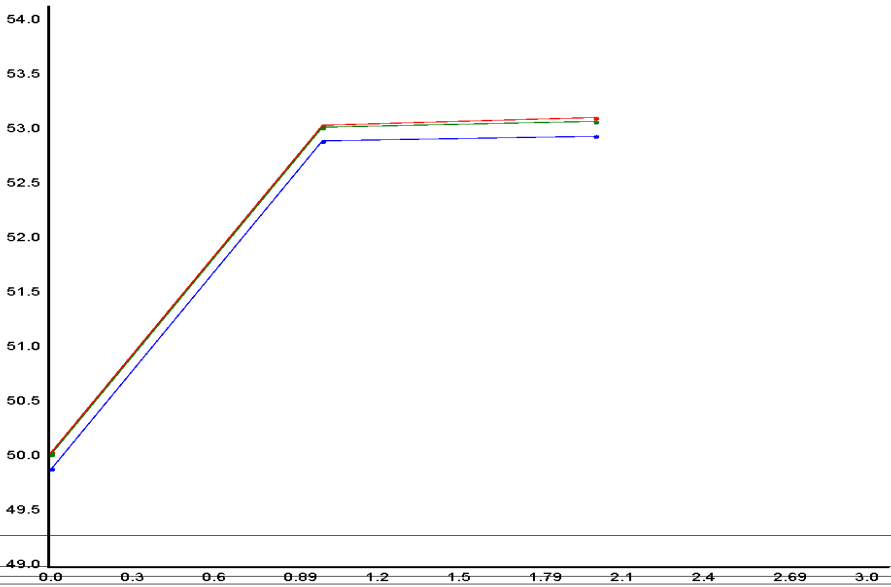
Lasso Regression:
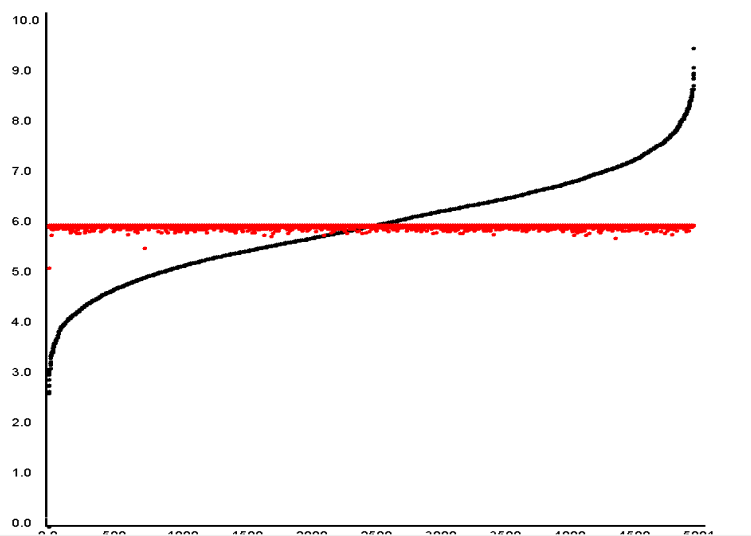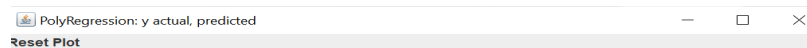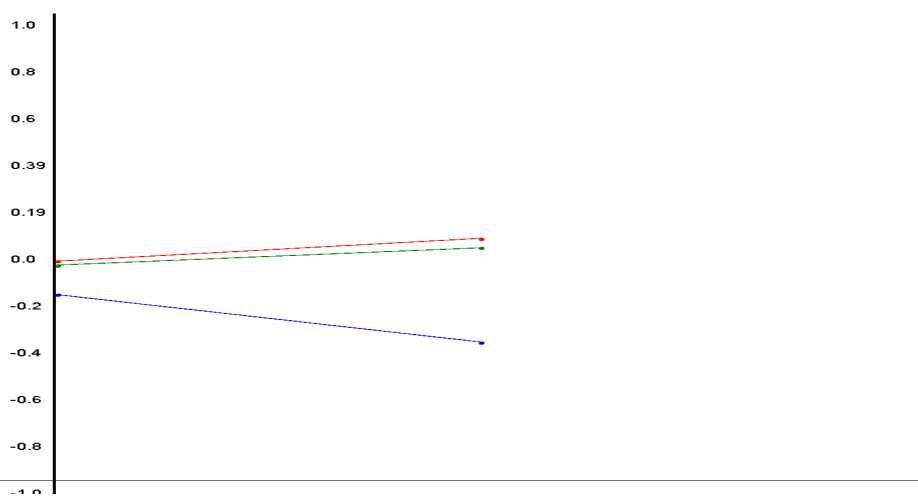
```
SUMMARY
    Parameters/Coefficients:
    Var      Estimate      Std. Error      t value       Pr(>|t|)       VIF
-----------------------------------------------------------------------------
    x0    10.237202       0.268622        38.110015     0.000000       NA
    x1    51296.068669    2960.620169     17.326123     0.000000       0.740165
    x2    -9614.433347    3262.670492     -2.946799     0.003211       0.927345
    x3    4082.312865     3132.019425     1.303412      0.192434       1.276879
    x4    8.091825        0.326978        24.747299     0.000000       0.900695
    Residual standard error: 242106.334210 on 4996.0 degrees of freedom
    Multiple R-squared:  0.531341,        Adjusted R-squared:  0.530966
    F-statistic: 1416.0525157356935 on 4.0 and 4996.0 DF,  p-value: 0.0
-----------------------------------------------------------------------------
```

```
[info] REPORT
[info]    modelName  mn  = LassoRegression
[info]    hparameter hp  = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]    features   fn  = Array(x0, x1, x2, x3, x4)
[info]    parameter  b   = VectorD(10.2372,    51296.1,       -9614.43,       4082.31,        8.0918)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.531341, rSqBar -> 0.530966, sst -> 624853298447502.600000, sse -> 292
842923414395.900000, mse0 -> 58556873308.217530, rmse -> 241985.274982, mae -> 193470.071204, dfm -> 4.000000, df -> 499
6.000000, fStat -> 1416.052516, aic -> -69081.710487, bic -> -69049.123521, mape -> NaN, smape -> NaN)
```
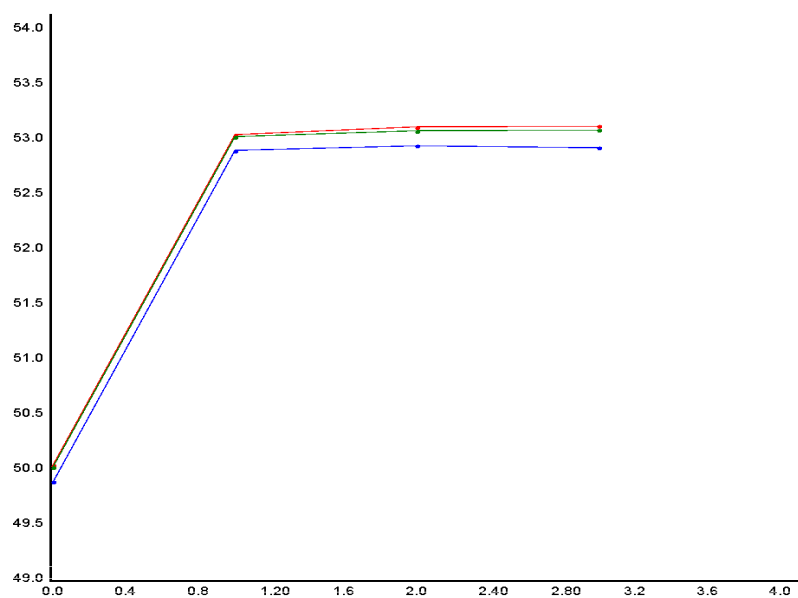
R^2 vs n for Regression with Stepwise

Ridge Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var        Estimate     Std. Error       t value        Pr(>|t|)       VIF
    ----------------------------------------------------------------------------
    x0     10.237202        0.268622         38.110015      0.000000       NA
    x1     51296.068671     2960.620169      17.326123      0.000000       0.740165
    x2     -9614.433350     3262.670492      -2.946799      0.003211       0.927345
    x3     4082.312867      3132.019425      1.303412       0.192434       1.276879
    x4     8.091825         0.326978         24.747299      0.000000       0.900695
    Residual standard error: 242106.334210 on 4996.0 degrees of freedom
    Multiple R-squared:  0.531341,        Adjusted R-squared:  0.530966
    F-statistic: 1416.0525157356897 on 4.0 and 4996.0 DF,  p-value: 0.0
    ----------------------------------------------------------------------------
```

```
[info] REPORT
[info]      modelName  mn  = Regression
[info]      hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]      features   fn  = Array(x0, x1, x2, x3, x4)
[info]      parameter  b   = VectorD(10.2372,    51296.1,        -9614.43,       4082.31,        8.0918)
[info]      fitMap     qof = LinkedHashMap(rSq -> 0.531341, rSqBar -> 0.530966, sst -> 624853298447502.600000, sse -
842923414396.300000, mse0 -> 58556873308.217620, rmse -> 241985.274982, mae -> 193470.071203, dfm -> 4.000000, df -
6.000000, fStat -> 1416.052516, aic -> -69081.710487, bic -> -69049.123521, mape -> NaN, smape -> NaN)
[info]
```

R^2 vs n for RidgeRegression with Forward

R^2 vs n for RidgeRegression with Backward
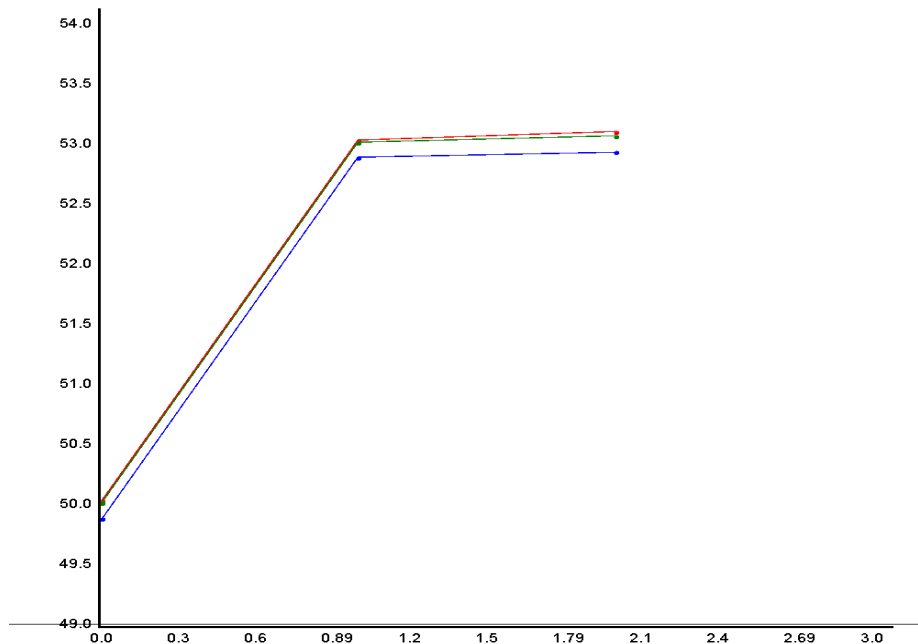
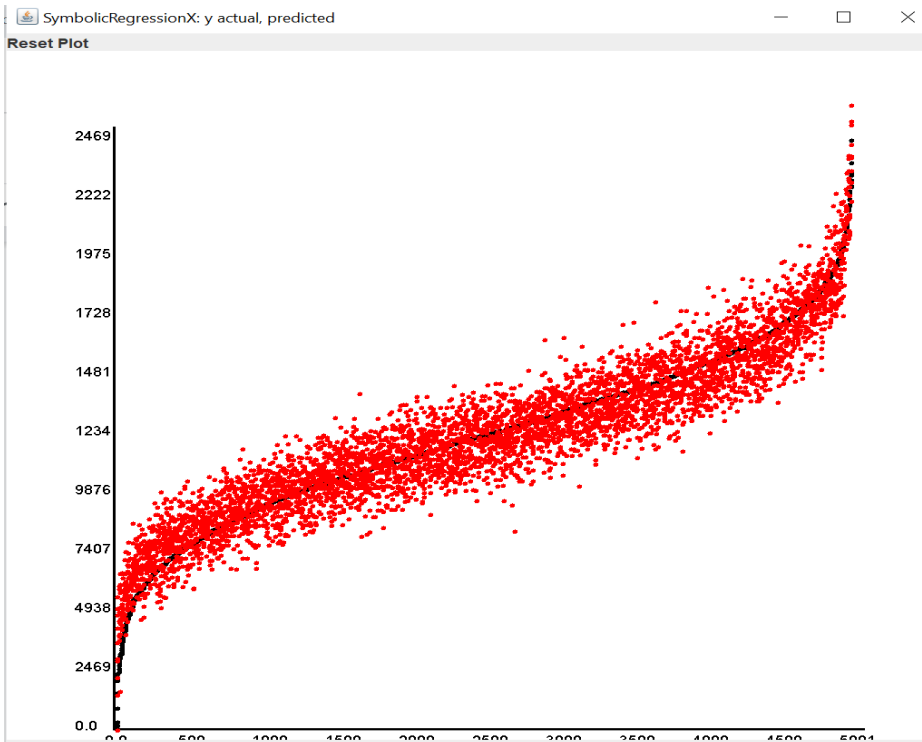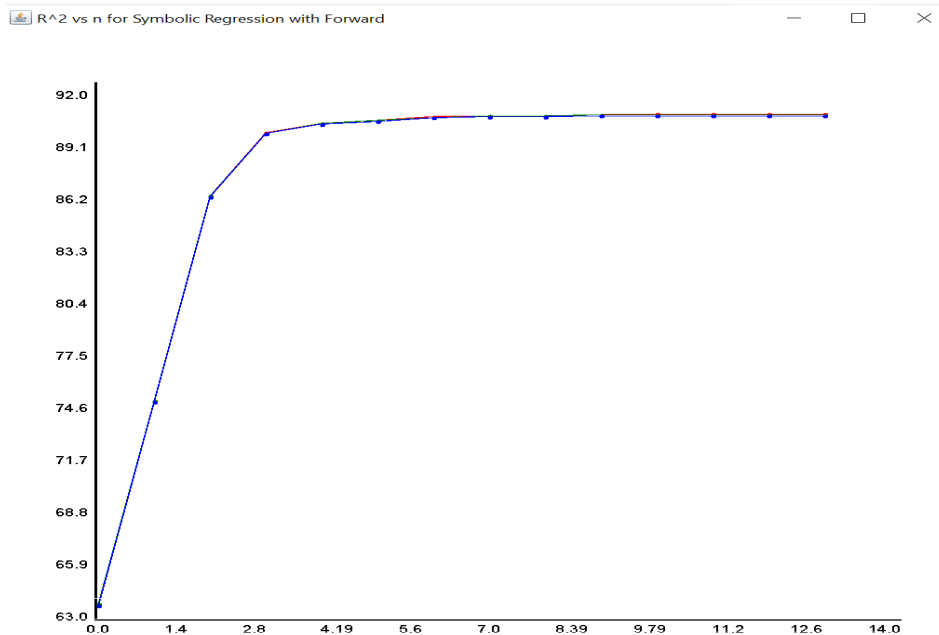## Quadratic Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var      Estimate     Std. Error       t value        Pr(>|t|)       VIF
    ---------------------------------------------------------------------------
    x0     5.136085      0.375942        13.661891      0.000000       NA
    x1     0.000025      0.000011        2.235661       0.025374       70.737970
    x2    -0.000000      0.000000        -2.203290      0.027574       70.737970
    Residual standard error: 0.994654 on 4998.0 degrees of freedom
    Multiple R-squared:  0.001003,        Adjusted R-squared:  0.000603
    F-statistic: 2.508735121336463 on 2.0 and 4998.0 DF,  p-value: 0.0
    ---------------------------------------------------------------------------
```

```
[info] REPORT
[info]     modelName  mn  = PolyRegression
[info]     hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_Cholesky,Fac_Cholesky)))
[info]     features   fn  = Array(x0, x1, x2)
[info]     parameter  b   = VectorD(5.13608,    2.47170e-05,    -1.77501e-1)
[info]     fitMap     qof = LinkedHashMap(rSq -> 0.001003, rSqBar -> 0.000603, sst -> 4949.663837, sse -> 4944.699874, m
se0 -> 0.988742, rmse -> 0.994355, mae -> 0.795422, dfm -> 2.000000, df -> 4998.000000, fStat -> 2.508735, aic -> -7061.
801889, bic -> -7042.249709, mape -> Infinity, smape -> 13.516651)
```
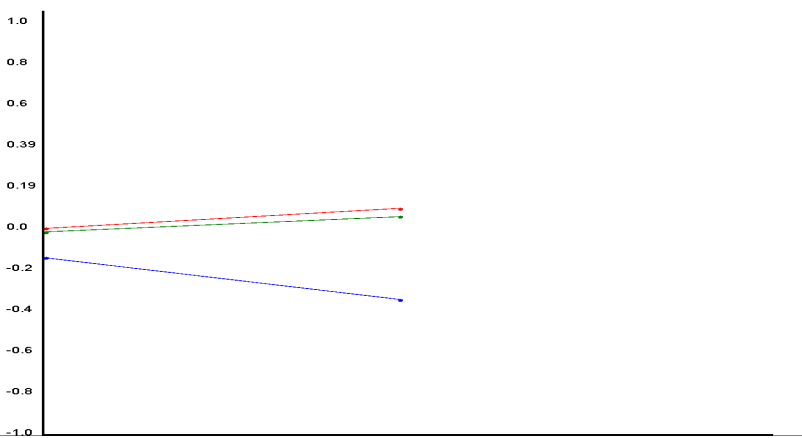
Symbolic Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var      Estimate       Std. Error        t value         Pr(>|t|)         VIF
----------------------------------------------------------------------------------
    x0       0.748740       0.968482          0.773107        0.439459         NA
    x1       -59820.197647  11031.304218      -5.422767       0.000000         53.835975
    x2       -116041.009552 11497.329285      -10.092866      0.000000         60.331523
    x3       -50652.522718  16740.784098      -3.025696       0.002481         191.120743
    x4       -1.247255      1.482206          -0.841485       0.400076         96.964524
    x5       1.443778       0.125642          11.491168       0.000000         61.922070
    x6       1.169153       0.138727          8.427694        0.000000         89.127736
    x7       15293.731548   1534.481878       9.966707        0.000000         89.547897
    x8       0.131378       0.127026          1.034256        0.301016         67.160827
    x9       -1695.790808   1379.002077       -1.229723       0.218801         61.331687
    x10      7681.708699    1356.387824       5.663357        0.000000         104.384962
    x11      0.000090       0.000014          6.657232        0.000000         50.560730
    x12      0.681448       0.141904          4.802189        0.000002         43.444824
    x13      0.901487       0.160334          5.622567        0.000000         72.323231
    x14      -0.043388      0.137809          -0.314839       0.752884         30.348317
    Residual standard error: 105773.876897 on 4986.0 degrees of freedom
    Multiple R-squared:  0.910725,         Adjusted R-squared:  0.910474
    F-statistic: 3633.1248885464265 on 14.0 and 4986.0 DF,  p-value: 0.0
----------------------------------------------------------------------------------
```

```
[info] REPORT
[info]    modelName  mn  = SymbolicRegressionX
[info]    hparameter hp  = HyperParameter (HashMap(factorization -> (Fac_QR,Fac_QR)))
[info]    features   fn  = Array(Avg. Area Income, Avg. Area House Age, Avg. Area Number of Rooms, Avg. Area Number of
Bedrooms, Area Population, Avg. Area House Age_Avg. Area Income, Avg. Area Number of Rooms_Avg. Area Income, Avg. Area N
umber of Rooms_Avg. Area House Age, Avg. Area Number of Bedrooms_Avg. Area Income, Avg. Area Number of Bedrooms_Avg. Are
a House Age, Avg. Area Number of Bedrooms_Avg. Area Number of Rooms, Area Population_Avg. Area Income, Area Population_A
vg. Area House Age, Area Population_Avg. Area Number of Rooms, Area Population_Avg. Area Number of Bedrooms)
[info]    parameter  b   = VectorD(0.748740,   -59820.2,      -116041,      -50652.5,      -1.24725,      1.44378,
1.16915,      15293.7,      0.131378,     -1695.79,      7681.71,      9.00561e-05,   0.681448,      0.901487
,     -0.043387)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.910725, rSqBar -> 0.910474, sst -> 624853298447502.600000, sse -> 557
83931586116.445000, mse0 -> 11154555406.142061, rmse -> 105615.128680, mae -> 84679.343701, dfm -> 14.000000, df -> 4986
.000000, fStat -> 3633.124889, aic -> -64915.463682, bic -> -64817.702784, mape -> NaN, smape -> NaN)
[info]
```
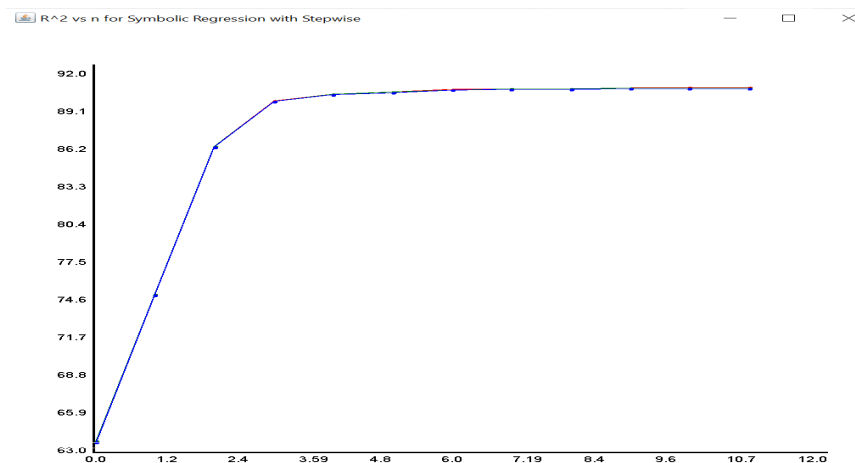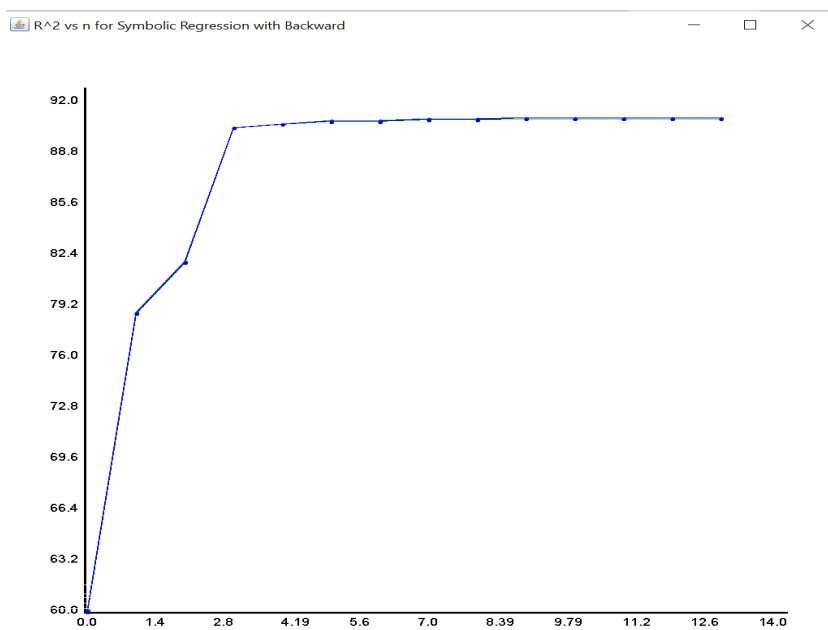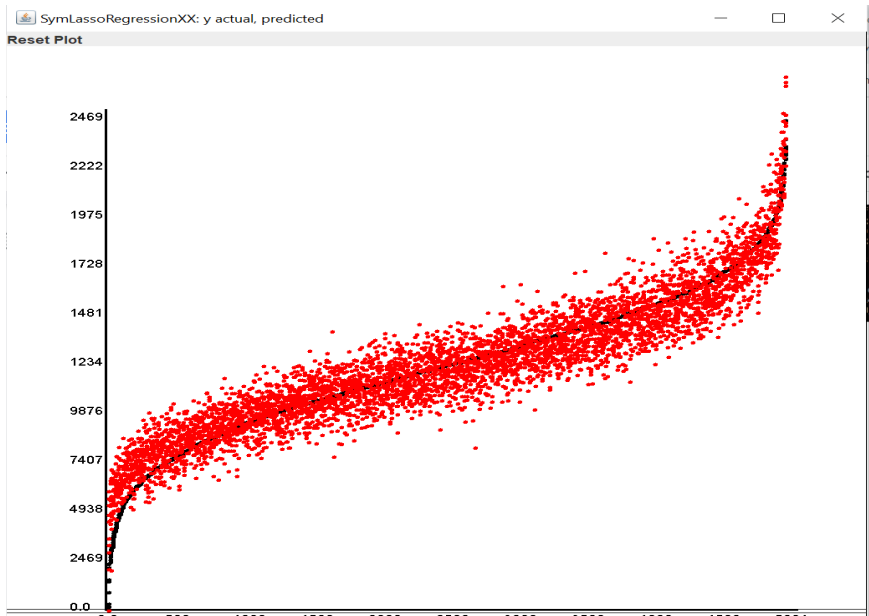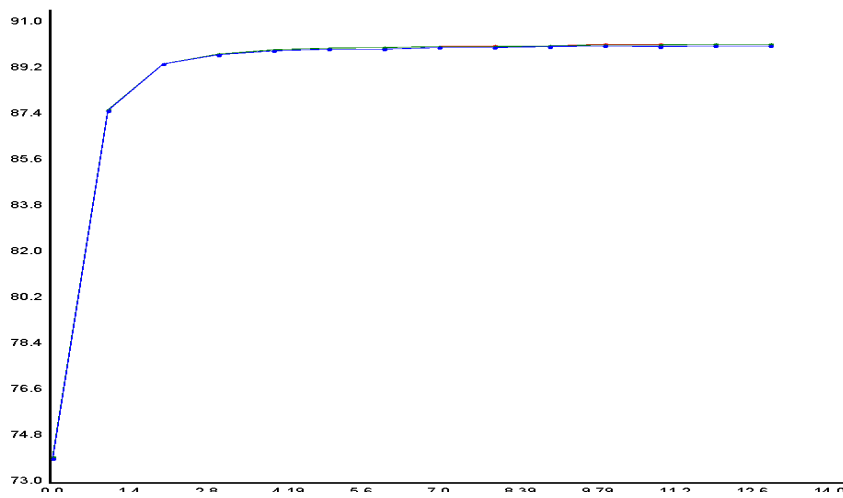
R^2 vs n for Symbolic Regression with Forward

R^2 vs n for Symbolic Regression with Backward

R^2 vs n for Symbolic Regression with Stepwise

Symbolic Lasso Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var       Estimate      Std. Error      t value        Pr(>|t|)       VIF
----------------------------------------------------------------------------------
    x0     6.049819        0.511927      11.817730       0.000000       NA
    x1     -652.522107     5866.278979   -0.111233       0.911432       13.861542
    x2     -56546.883536   6073.730779   -9.310074       0.000000       15.329525
    x3     -40405.242861   8796.889354   -4.593128       0.000004       48.048714
    x4     2.957676        0.787922      3.753766        0.000174       24.947663
    x5     0.232008        0.014462      16.042840       0.000000       51.193915
    x6     -0.030833       0.023823      -1.294232       0.195585       96.059965
    x7     0.051243        0.016452      3.114658        0.001842       78.264382
    x8     719.782992      196.628326    3.660627        0.000252       86.040597
    x9     0.000015        0.000002      6.831178        0.000000       54.746866
    x10    0.000005        0.000002      2.253994        0.024197       86.659911
    x11    0.083322        0.026699      3.120784        0.001804       83.595858
    x12    0.000000        0.000004      0.022806        0.981805       121.812318
    x13    -0.132079       0.045437      -2.906840       0.003651       126.901650
    x14    0.104313        0.030581      3.411038        0.000647       95.514297
    Residual standard error: 110852.319815 on 4986.0 degrees of freedom
    Multiple R-squared:  0.901946,       Adjusted R-squared:  0.901671
    F-statistic: 3275.9794038948385 on 14.0 and 4986.0 DF,  p-value: 0.0
----------------------------------------------------------------------------------
```

```
[info] REPORT
[info]      modelName  mn  = SymLassoRegressionXX
[info]      hparameter hp  = HyperParameter (HashMap(lambda -> (0.01,0.01)))
[info]      features   fn  = Array(Avg. Area Income, Avg. Area House Age, Avg. Area Number of Rooms, Avg. Area Number of
Bedrooms, Area Population, Avg. Area Number of Rooms_Avg. Area House Age_Avg. Area Income, Avg. Area Number of Bedrooms_
Avg. Area House Age_Avg. Area Income, Avg. Area Number of Bedrooms_Avg. Area Number of Rooms_Avg. Area Income, Avg. Area
 Number of Bedrooms_Avg. Area Number of Rooms_Avg. Area House Age, Area Population_Avg. Area House Age_Avg. Area Income,
 Area Population_Avg. Area Number of Rooms_Avg. Area Income, Area Population_Avg. Area Number of Rooms_Avg. Area House A
ge, Area Population_Avg. Area Number of Bedrooms_Avg. Area Income, Area Population_Avg. Area Number of Bedrooms_Avg. Are
a House Age, Area Population_Avg. Area Number of Bedrooms_Avg. Area Number of Rooms)
[info]      parameter  b   = VectorD(6.04982,     -652.522,       -56546.9,       -40405.2,       2.95768,        0.232008
,       -0.0308328,     0.0512426,      719.783,        1.49384e-05,    5.41563e-06,    0.0833218,      8.89805e-08,    -
0.132079,       0.10431)
[info]      fitMap     qof = LinkedHashMap(rSq -> 0.901946, rSqBar -> 0.901671, sst -> 624853298447502.600000, sse -> 612
69148726467.500000, mse0 -> 12251379469.399620, rmse -> 110685.949738, mae -> 88342.124664, dfm -> 14.000000, df -> 4986
.000000, fStat -> 3275.979404, aic -> -65149.987003, bic -> -65052.226105, mape -> NaN, smape -> NaN)
[info]
```
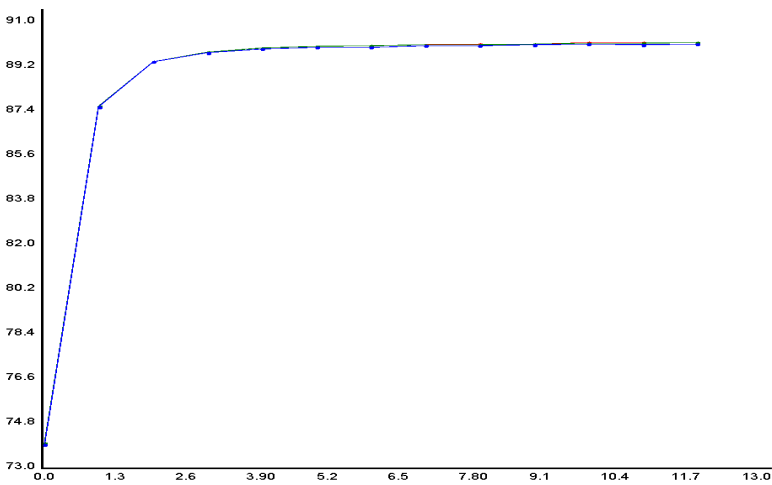
R^2 vs n for Symbolic Lasso Regression with Forward

R^2 vs n for Symbolic Lasso Regression with Backward
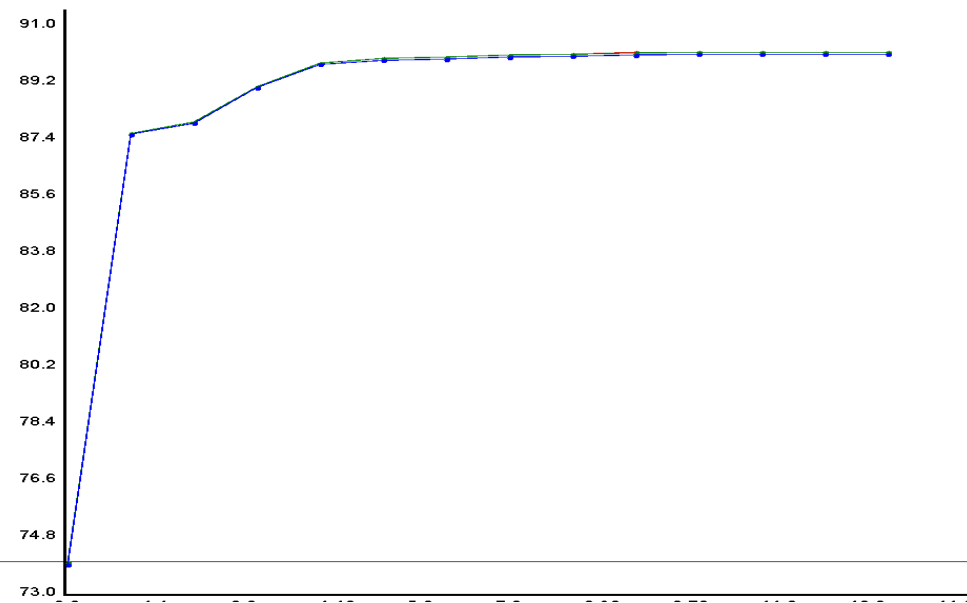
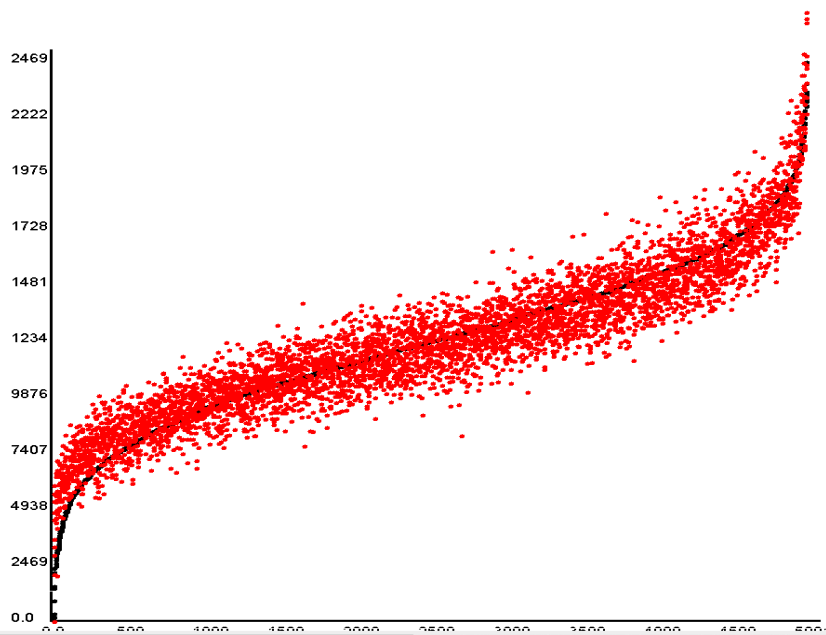R^2 vs n for Symbolic Lasso Regression with Stepwise

Symbolic Ridge Regression:

```
SUMMARY
    Parameters/Coefficients:
    Var      Estimate     Std. Error    t value      Pr(>|t|)      VIF
    -----------------------------------------------------------------------
    x0      6.049747      0.511979      11.816404    0.000000      NA
    x1      -653.277645   5866.867342   -0.111350    0.911339      13.861542
    x2      -56545.921993 6074.339949   -9.308982    0.000000      15.329525
    x3      -40403.726429 8797.771646   -4.592495    0.000004      48.048714
    x4      2.957598      0.788001      3.753290     0.000175      24.947663
    x5      0.232008      0.014463      16.041286    0.000000      51.193915
    x6      -0.030831     0.023826      -1.294025    0.195657      96.059965
    x7      0.051240      0.016454      3.114209     0.001844      78.264382
    x8      719.759222    196.648047    3.660139     0.000252      86.040597
    x9      0.000015      0.000002      6.830608     0.000000      54.746866
    x10     0.000005      0.000002      2.253767     0.024211      86.659911
    x11     0.083321      0.026702      3.120459     0.001806      83.595858
    x12     0.000000      0.000004      0.022802     0.981808      121.812318
    x13     -0.132079     0.045442      -2.906555    0.003654      126.901650
    x14     0.104313      0.030584      3.410690     0.000648      95.514297
    Residual standard error: 110863.437846 on 4985.0 degrees of freedom
    Multiple R-squared:  0.901946,      Adjusted R-squared:  0.901651
    F-statistic: 3056.967543696419 on 15.0 and 4985.0 DF,  p-value: 0.0
    -----------------------------------------------------------------------
```

```
[info] REPORT
[info]    modelName  mn  = SymRidgeRegressionXX
[info]    hparameter hp  = HyperParameter (HashMap(lambda -> (0.01,0.01), factorization -> (Fac_Cholesky,Fac_Cholesky))
)
[info]    features   fn  = Array(Avg. Area Income, Avg. Area House Age, Avg. Area Number of Rooms, Avg. Area Number of
Bedrooms, Area Population, Avg. Area Number of Rooms_Avg. Area House Age_Avg. Area Income, Avg. Area Number of Bedrooms_
Avg. Area House Age_Avg. Area Income, Avg. Area Number of Bedrooms_Avg. Area Number of Rooms_Avg. Area Income, Avg. Area
 Number of Bedrooms_Avg. Area Number of Rooms_Avg. Area House Age, Area Population_Avg. Area House Age_Avg. Area Income,
 Area Population_Avg. Area Number of Rooms_Avg. Area Income, Area Population_Avg. Area Number of Rooms_Avg. Area House A
ge, Area Population_Avg. Area Number of Bedrooms_Avg. Area Income, Area Population_Avg. Area Number of Bedrooms_Avg. Are
a House Age, Area Population_Avg. Area Number of Bedrooms_Avg. Area Number of Rooms)
[info]    parameter  b   = VectorD(6.04975,     -653.278,       -56545.9,        -40403.7,       2.95760,        0.232008
,      -0.0308309,     0.0512403,      719.759,       1.49387e-05,    5.41563e-06,    0.0833215,      8.89750e-08,    -
0.132079,       0.10431)
[info]    fitMap     qof = LinkedHashMap(rSq -> 0.901946, rSqBar -> 0.901651, sst -> 624853298447502.600000, sse -> 612
69148727619.380000, mse0 -> 12251379469.629951, rmse -> 110685.949739, mae -> 88342.129691, dfm -> 15.000000, df -> 4985
.000000, fStat -> 3056.967544, aic -> -65147.987003, bic -> -65043.708712, mape -> NaN, smape -> NaN)
```
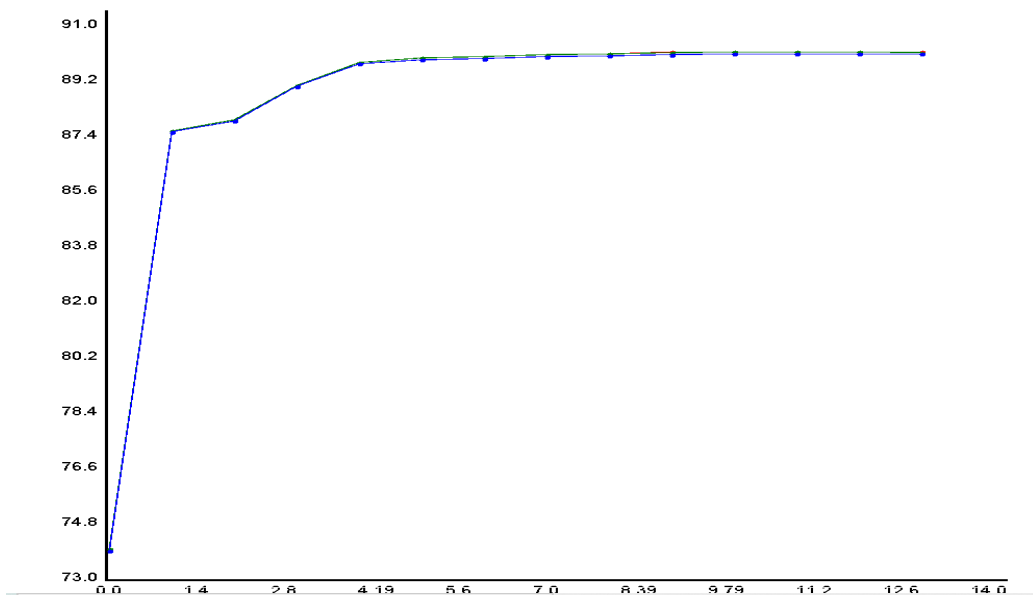
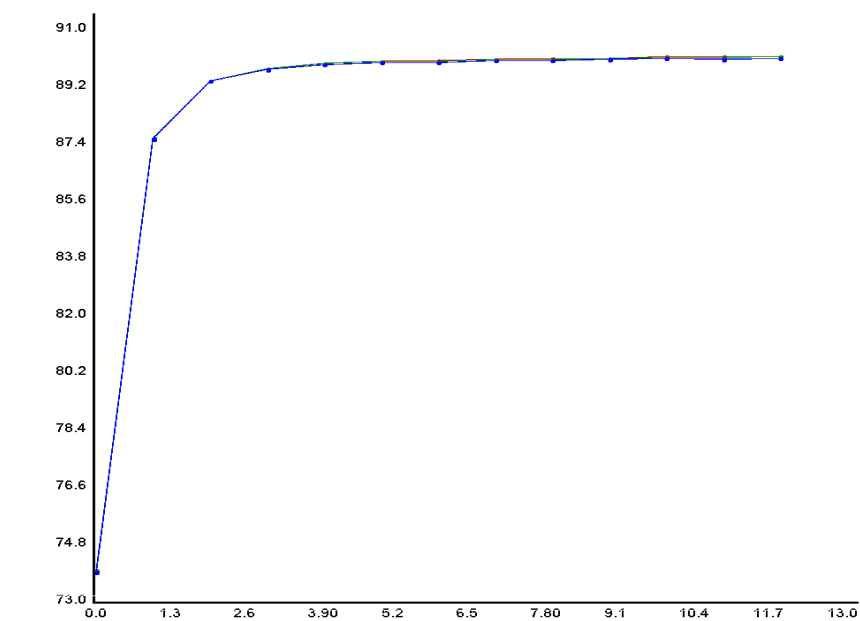SymRidgeRegressionXX: y actual, predicted
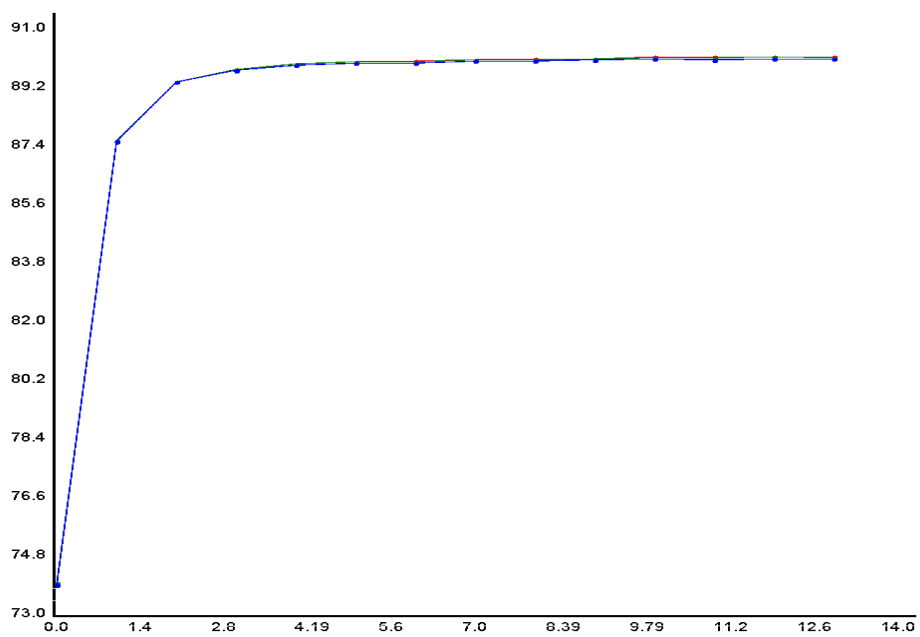
Reset Plot

R^2 vs n for Symbolic Ridge Regression with Forward

R^2 vs n for Symbolic Ridge Regression with Backward

R^2 vs n for Symbolic Ridge Regression with Stepwise

Scala Methods to Execute:

LassoRegressionTest100
symLassoRegressionTest10
Linearregression100
QuadraticTest100
RidgeRegressionTest9
symbolicRegressionTest100
symRidgeRegressionTest10

Models:

Linear Regression:
     Multiple R-squared: 0.531341
     Adjusted R-squared: 0.530966

Lasso Regression
     Multiple R-squared: 0.530966
     Adjusted R-squared: 0.531341

Ridge Regression

     Multiple R-squared: 0.531341
     Adjusted R-squared: 0.530966

Quadratic Regression

     Multiple R-squared: 0.001003
     Adjusted R-squared: 0.0.000603

Symbolic Regression
     Multiple R-squared: 0.910725
     Adjusted R-squared: 0.910474

Ridge Symbolic Regression
     Multiple R-squared: 0.901946
     Adjusted R-squared: 0.901651

Lasso Symbolic Regression
     Multiple R-squared: 0.901946
     Adjusted R-squared: 0.901671

Best Model Symbolic Lasso Regression Regression & Symbolic Ridge Regression