Franklin D. Worrell
24 February 2017
6990-603 S17

Report on Programming Assignment 1: Linear Regression for Forest Fire Prediction

This experiment attempts to develop a model for forest fire prediction using a set of meteorological data collected for forest fires that occurred in Montesinho natural park in Portugal.[1] Two of the input features of the data set were qualitative values and were replaced by scalar values (month by 1-12 and day by 1-7). Another first input feature consisting of scalar value 1 was added to all instances for calculation of a zeroth beta value—the y-intercept. The input features were initially ranked based on magnitude of Pearson Correlation Coefficient (PCC) with the dependent variable, area burned (Table 1). The top five features were selected based on PCC ranking: temperature, relative humidity, Duff Moisture Code, the x-axis spatial coordinate of a park map, and month (Graphs 1-5 plot each of these features vs. area burned). Models of order 1, 2, 4, 6, and 9 were developed for both the 12 feature set and the top five feature set using the non-iterative equation and 10-fold cross validation. For each model for both the 12 feature and top five feature sets, the average of and standard-deviation of both mean-absolute-error (MAE) and root-mean-squared-error (RMSE) were calculated (Tables 2 and 3). Plots of MAE and RMSE vs Model reveal the expected trend for the 12 feature set—both values peak are roughly model of order 4 and then drop off substantially (Graphs 6 and 7). Conversely, similar plots for the top five feature set show a drop after model order 4 for MAE, whereas RMSE does not peak until model order 6 (Graphs 8 and 9). Regularization was applied for both sets for Models 6 and 9, with 6 different values of λ (Tables 4-7). Graphs 10-17 are log-log plots of this data.

[1] The data set used here is that collected and analyzed by P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data" in J. Neves, M. F. Santos and J. Machado Eds., *New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence*, December, Guimaraes, Portugal, pp. 512-523, 2007.

Table 1: Feature SL by PCC Magnitude

ID	PCC
9	0.097844107
10	-0.075518563
6	0.072994296
1	0.063385299
3	0.056495774
7	0.049383225
2	0.044873225
5	0.040122004
4	0.023226372
11	0.012317277
8	0.008257688
12	-0.007365729

Table 2: All 12 Input Features: Model, MAE, and RMSE

Model	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
1	20.24718217	5.882529115	49.15959682	43.59655179
2	20.37399357	6.069860713	49.33431673	43.61040537
4	20.47197906	6.075768455	49.36329082	43.54954282
6	20.29894136	5.827871942	49.30558971	43.436139
9	20.01526519	6.289133875	48.964272	43.74695297

Table 3: Top 5 Input Features: Model, MAE, and RMSE

Model	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
1	19.79633101	6.027032588	48.5597626	43.80273736
2	19.35742088	6.186835598	48.6773433	43.71206309
4	19.64086239	6.452025943	48.88378257	43.86654177
6	13.54302301	7.186543292	48.9488658	45.06565288
9	13.30135531	7.225033814	48.91734323	45.13994377

Table 4: MAE and RMSE by λ-Value for Model 6 of All 12 Input Features

λ	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
0	20.29894136	5.827871942	49.30558971	43.436139
5.00E-09	20.29894136	5.827871942	49.30558971	43.436139
1.50E-06	20.29894136	5.827871942	49.30558971	43.436139
0.0002	20.29894108	5.827871937	49.30558954	43.4361391
1	20.29718704	5.828021391	49.30431888	43.43688813
2	20.29543919	5.828171417	49.30305457	43.43763414

Table 5: MAE and RMSE by λ-Value for Model 9 of All 12 Input Features

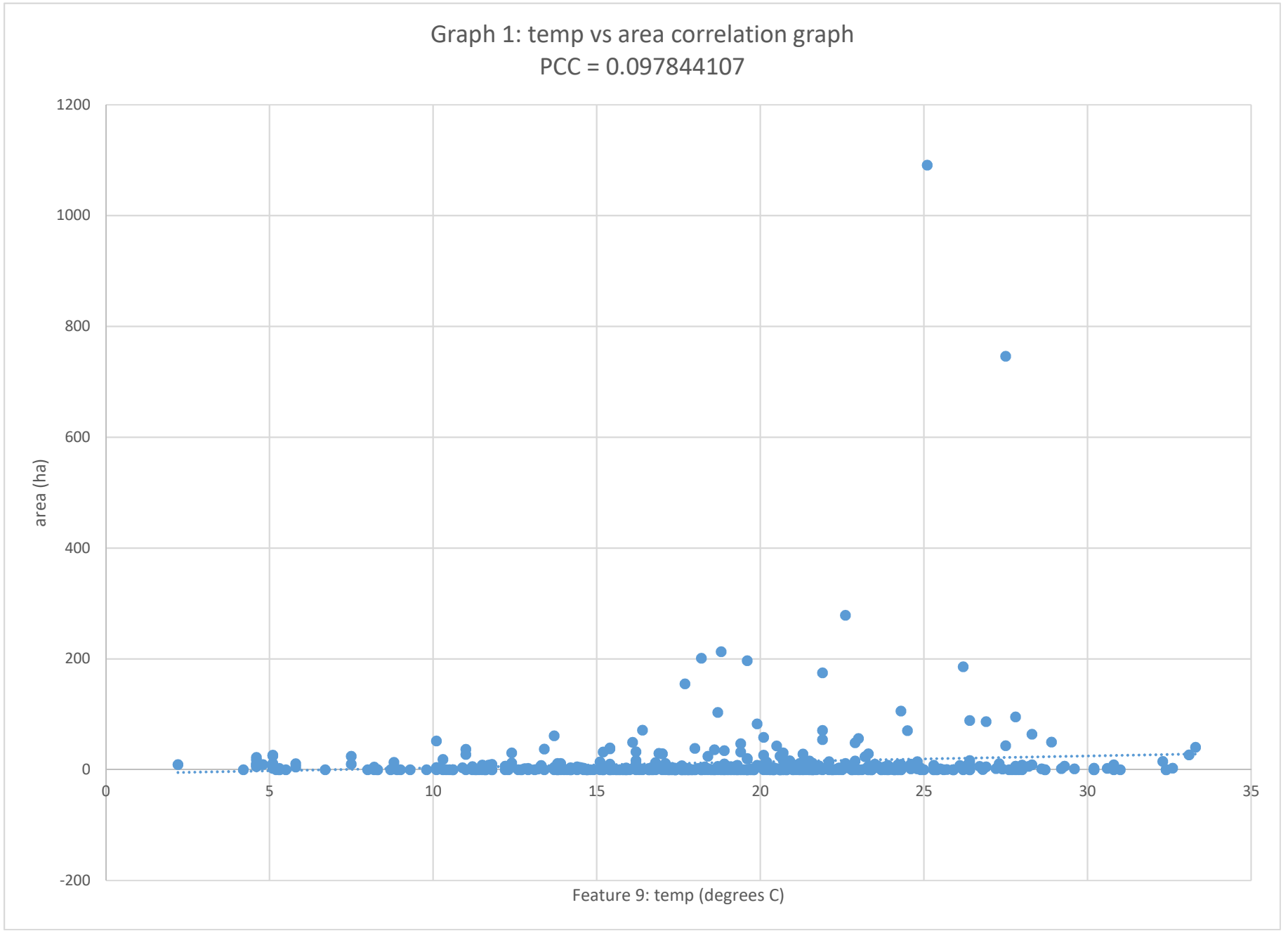
λ	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
0	20.01526519	6.289133875	48.964272	43.74695297
5.00E-09	20.01526519	6.289133875	48.964272	43.74695297
1.50E-06	20.01526519	6.289133875	48.964272	43.74695297
0.0002	20.01526519	6.289133877	48.96427199	43.74695298
1	20.0152327	6.289144995	48.96422642	43.7469651
2	20.01520023	6.289156113	48.96418092	43.74697721

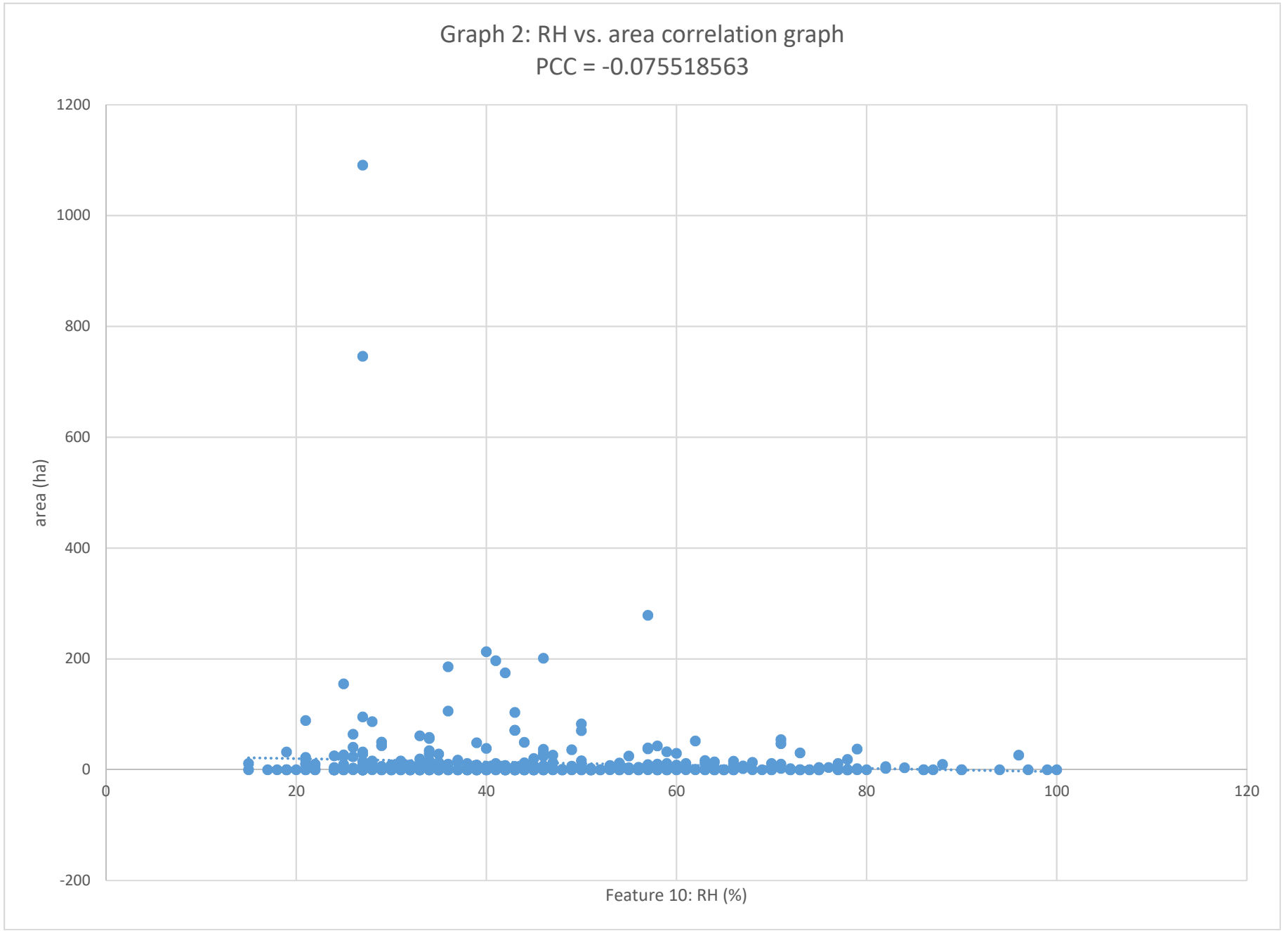
Table 6: MAE and RMSE by λ-Value for Model 6 of Top 5 Input Features

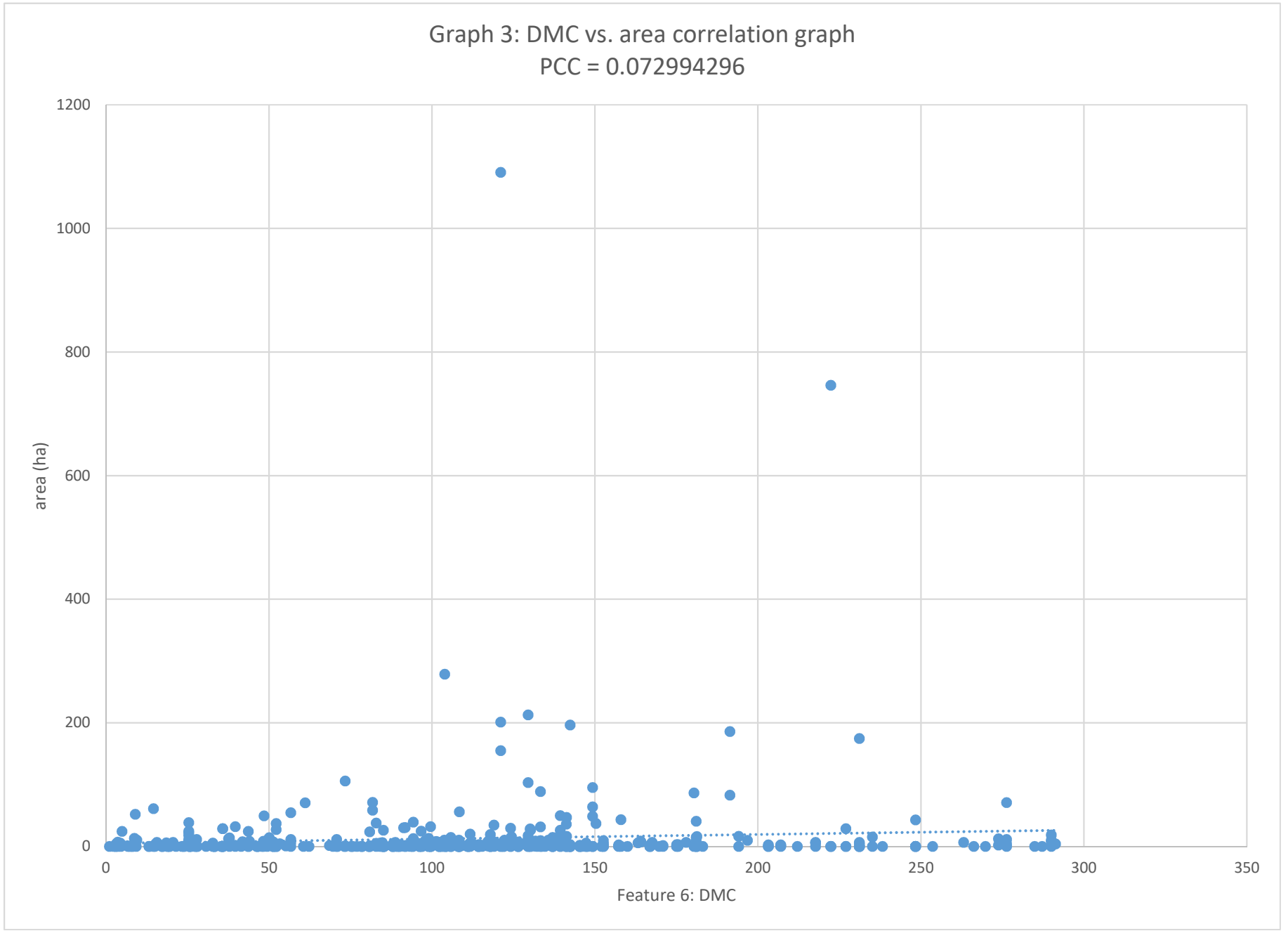
λ	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
0	13.54302301	7.186543292	48.9488658	45.06565288
5.00E-09	13.54302301	7.186543292	48.9488658	45.06565288
1.50E-06	13.54302301	7.186543292	48.9488658	45.06565288
0.0002	13.54302301	7.186543292	48.9488658	45.06565288
1	13.54302301	7.186543292	48.9488658	45.06565288
2	13.54302301	7.186543292	48.9488658	45.06565288

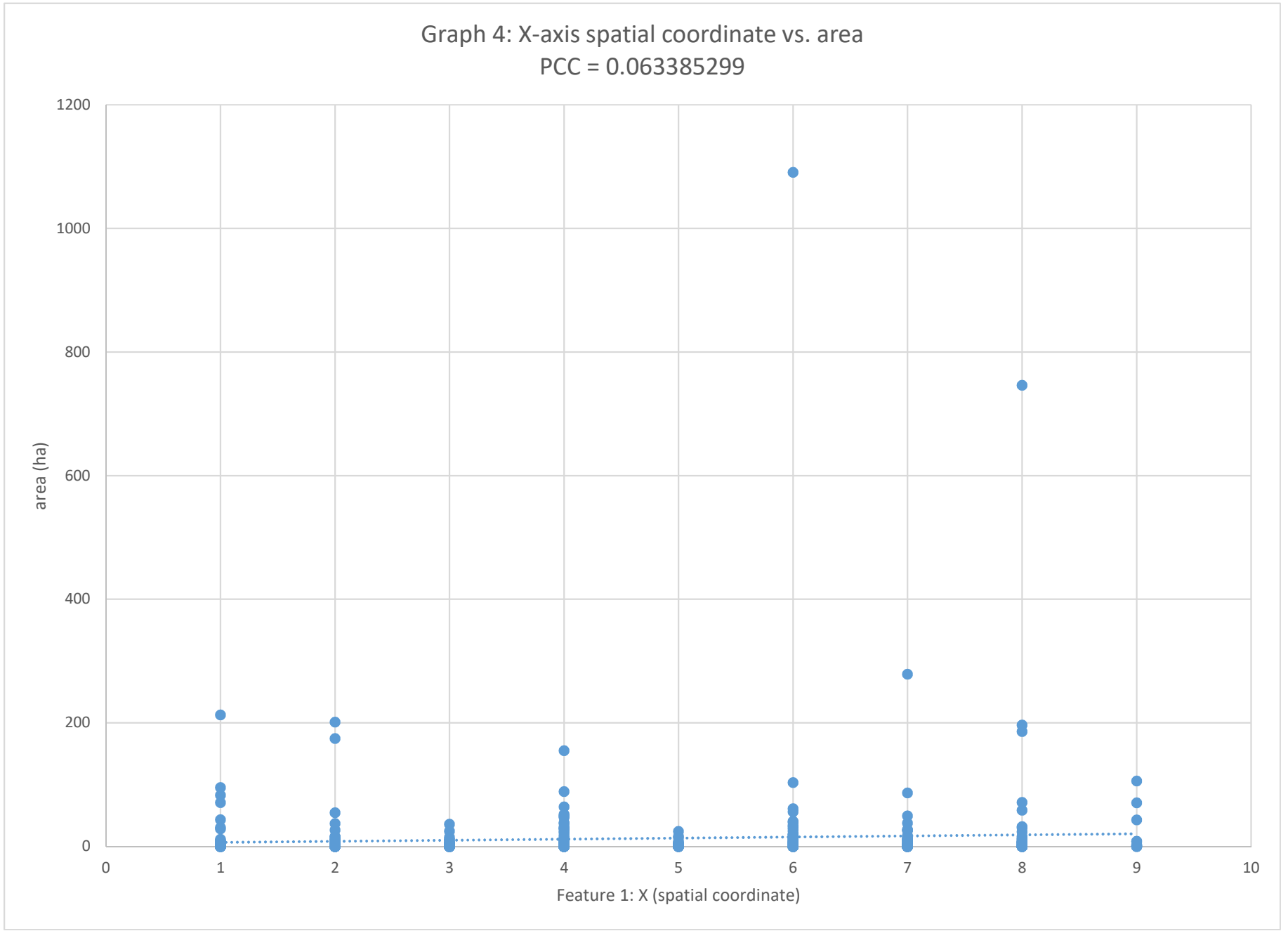
Table 7: MAE and RMSE by λ-Value for Model 9 of Top 5 Input Features

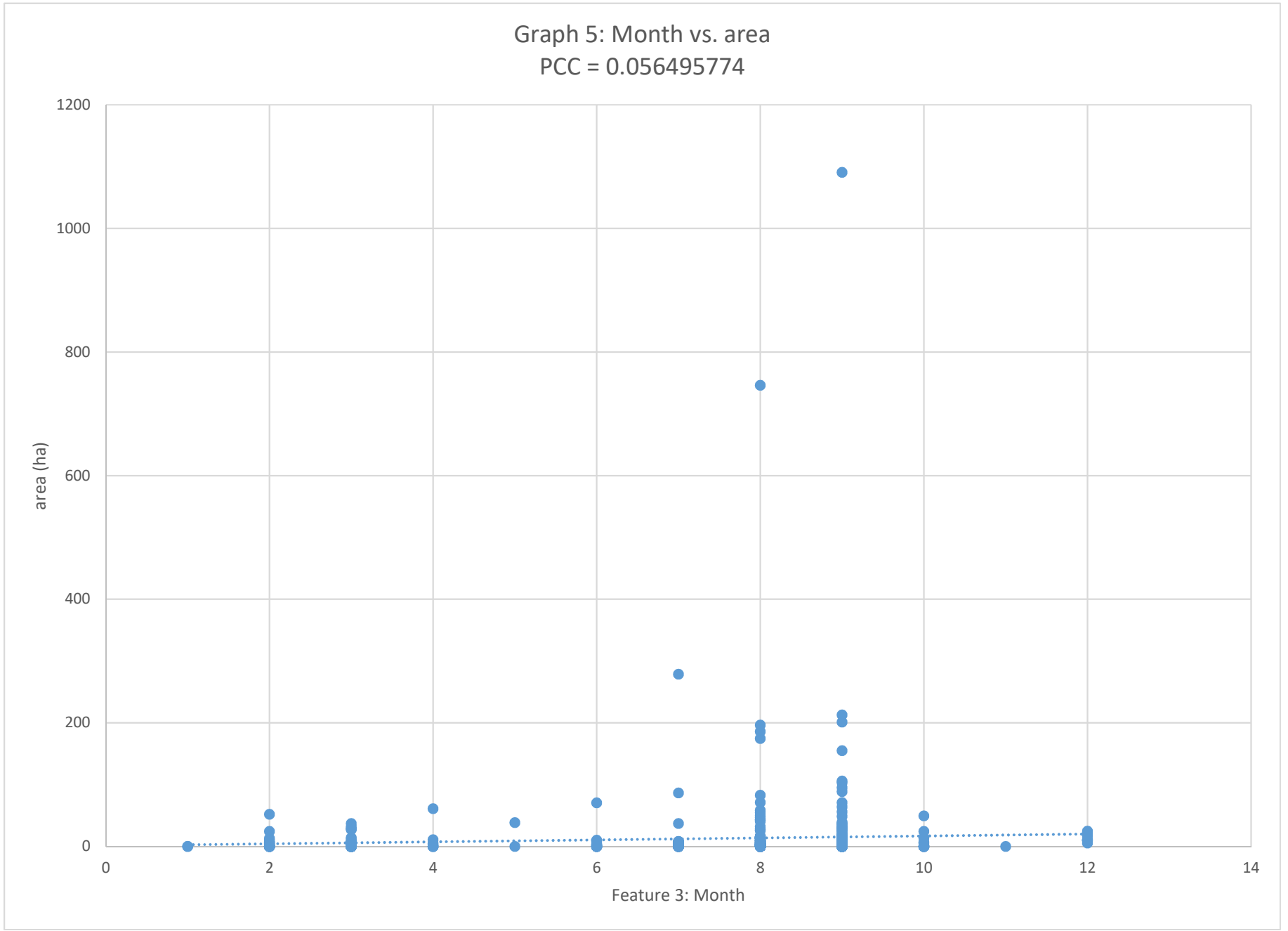
λ	Avg of MAE	SD of MAE	Avg of RMSE	SD of RMSE
0	13.30135531	7.225033814	48.91734323	45.13994377
5.00E-09	13.30135531	7.225033814	48.91734323	45.13994377
1.50E-06	13.30135531	7.225033814	48.91734323	45.13994377
0.0002	13.30135531	7.225033814	48.91734323	45.13994377
1	13.30135531	7.225033814	48.91734323	45.13994377
2	13.30135531	7.225033814	48.91734323	45.13994377


Graph 1: temp vs area correlation graph
PCC = 0.097844107
1200
1000
800
600
400
200
0
-200
area (ha)
0
5
10
15
20
25
30
35
Feature 9: temp (degrees C)

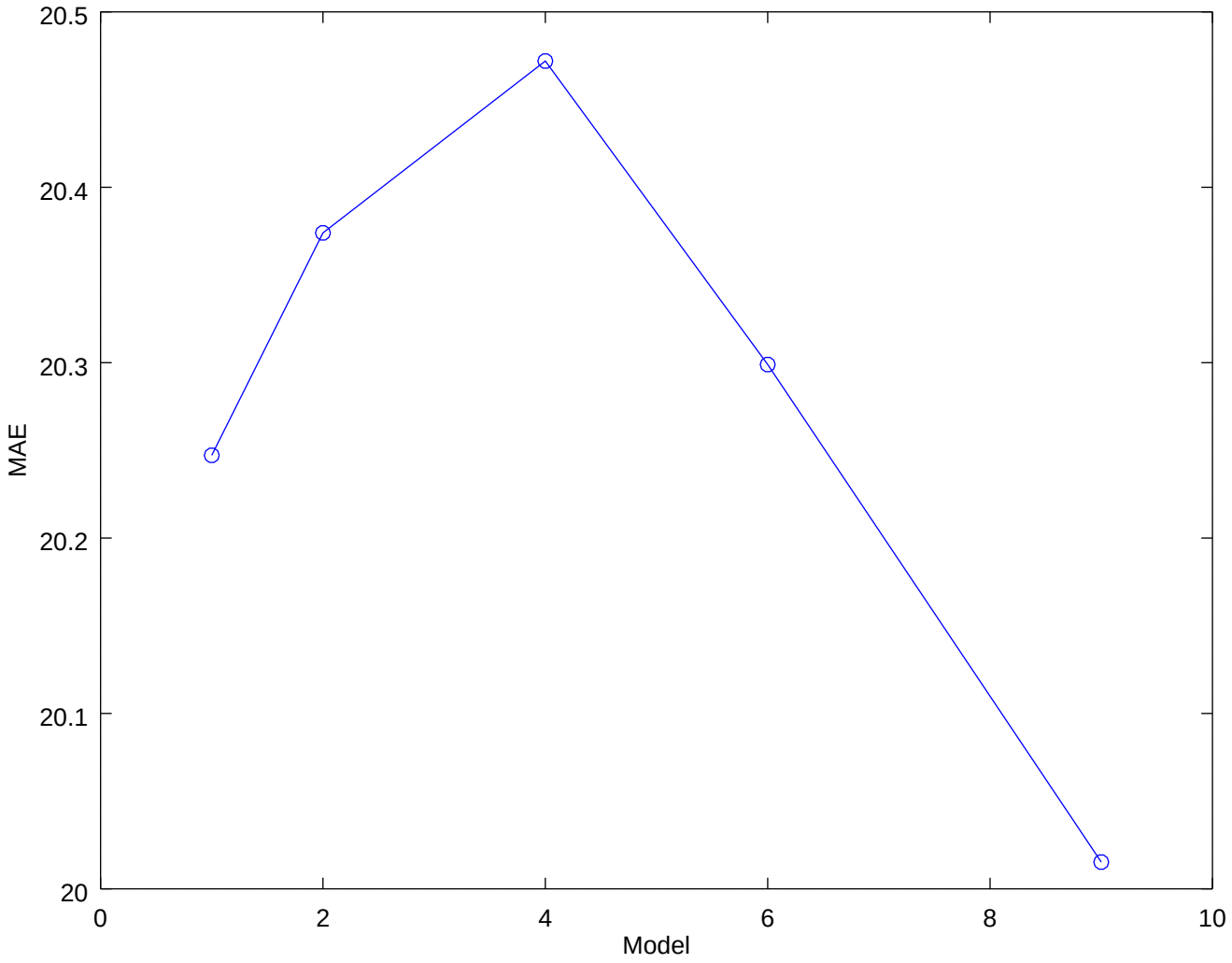

Graph 2: RH vs. area correlation graph
PCC = -0.075518563
1200
1000
800
600
400
200
0
-200
area (ha)
0
20
40
60
80
100
120
Feature 10: RH (%)


Graph 3: DMC vs. area correlation graph
PCC = 0.072994296
1200
1000
800
600
400
200
0
area (ha)
0
50
100
150
200
250
300
350
Feature 6: DMC

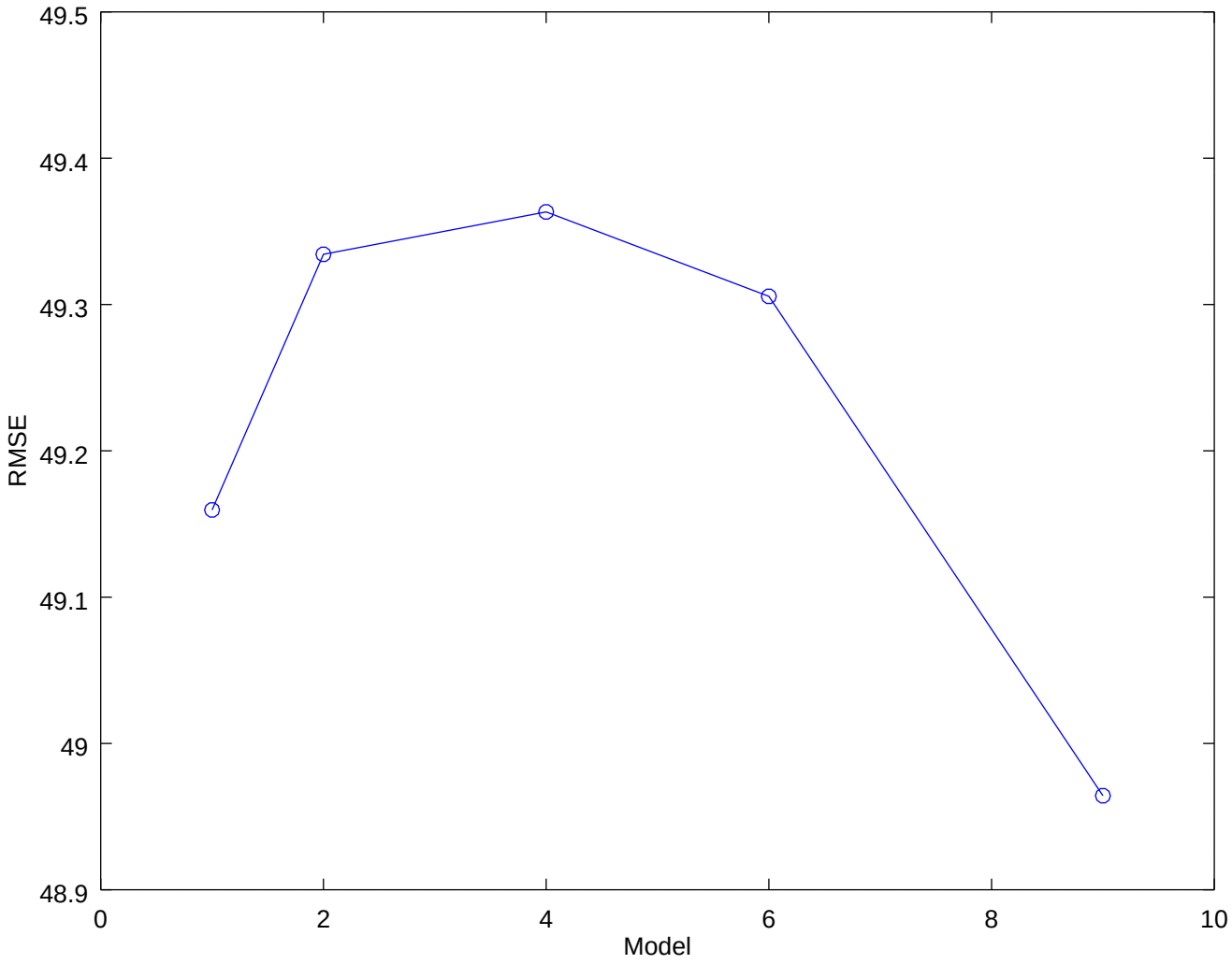

Graph 4: X-axis spatial coordinate vs. area
PCC = 0.063385299
area (ha)
1200
1000
800
600
400
200
0
0
1
2
3
4
5
6
7
8
9
10
Feature 1: X (spatial coordinate)


Graph 5: Month vs. area
PCC = 0.056495774
1200
1000
800
600
400
200
0
area (ha)
0
2
4
6
8
10
12
14
Feature 3: Month

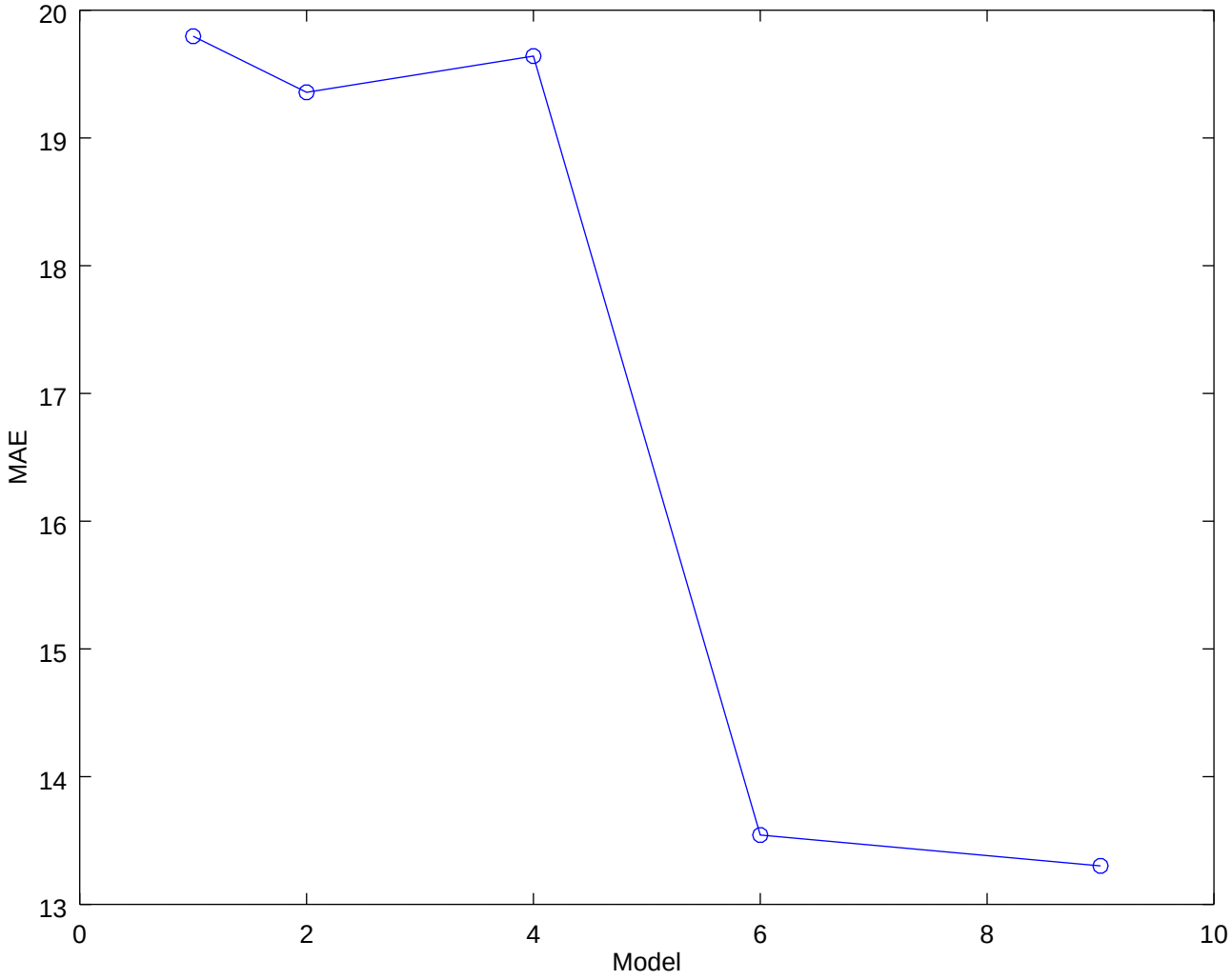
Graph 6: Model vs. MAE (12 Features)



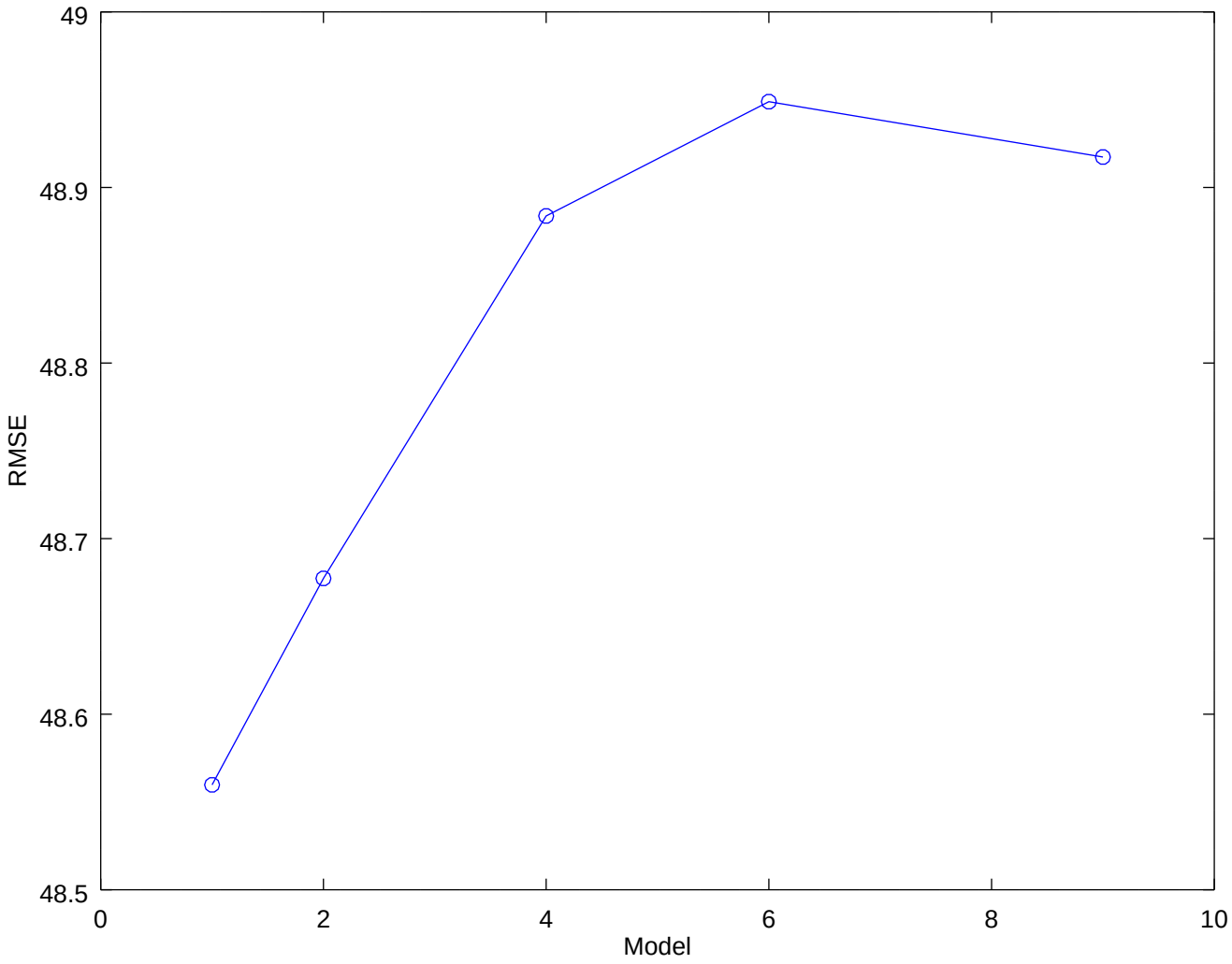
Graph 7: Model v. RMSE (12 Features)



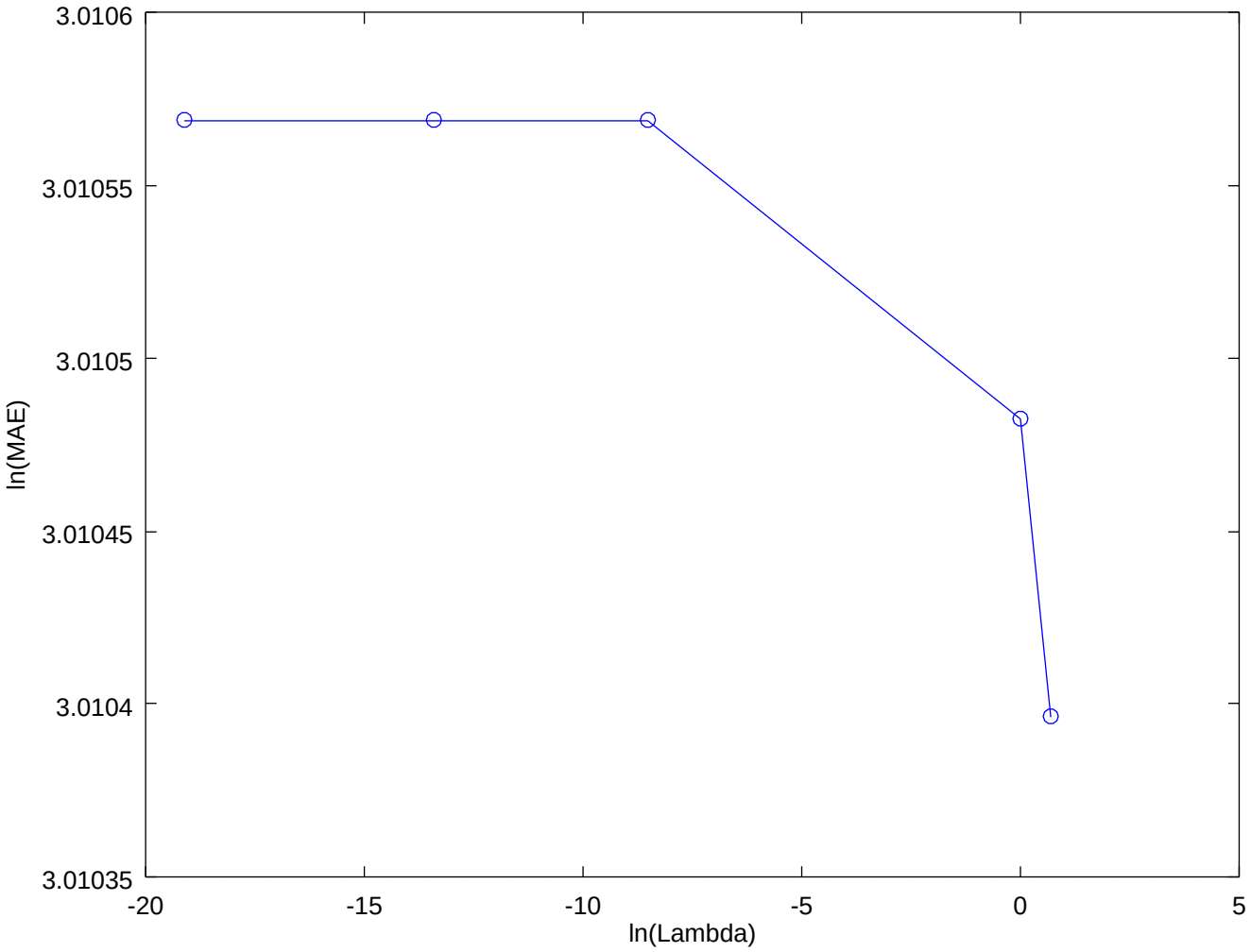
Graph 8: Model vs. MAE (Top 5 Features)



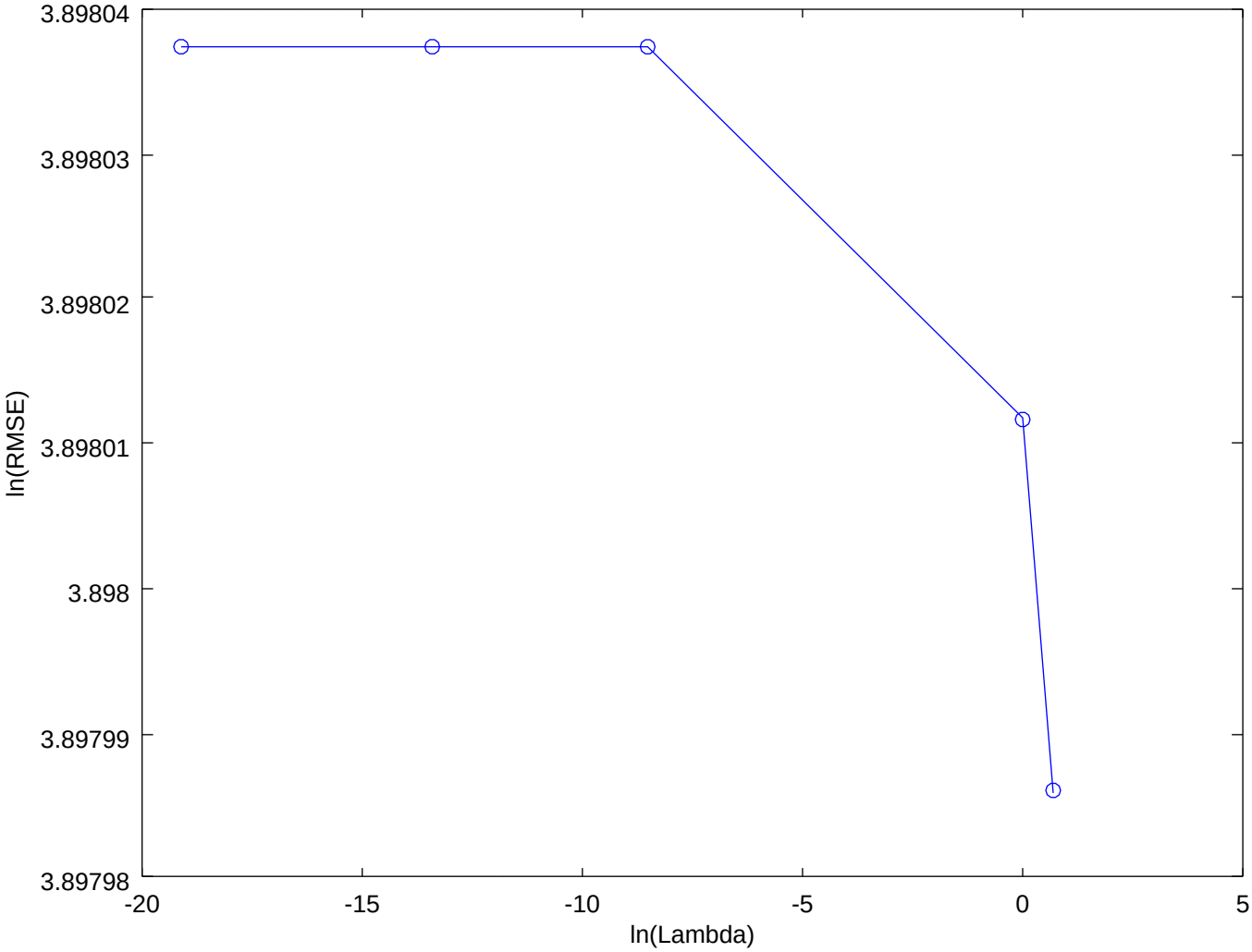
Graph 9: Model v. RMSE (Top 5 Features)



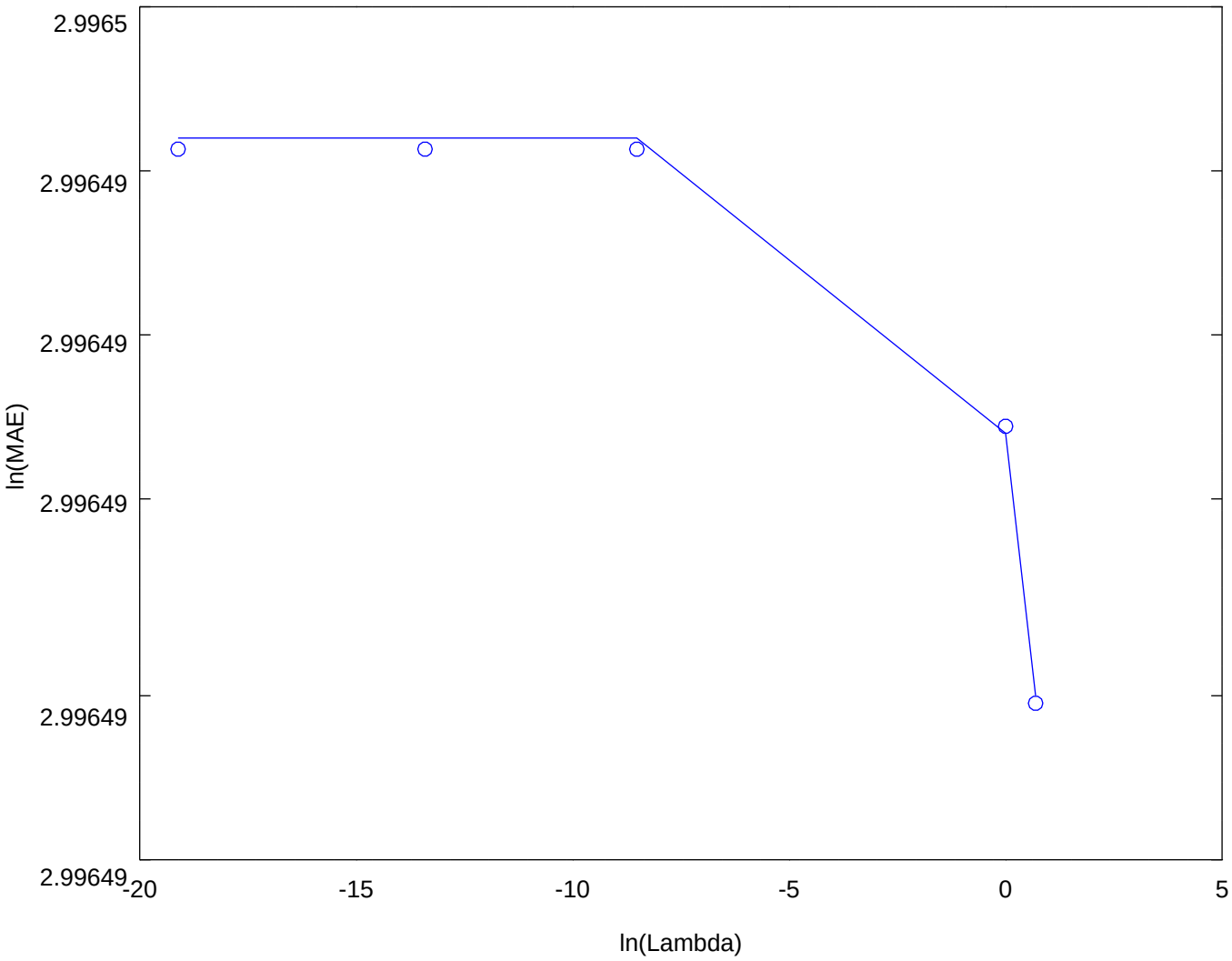
Graph 10: Lambda vs. MAE for Model 6 (12 Features)

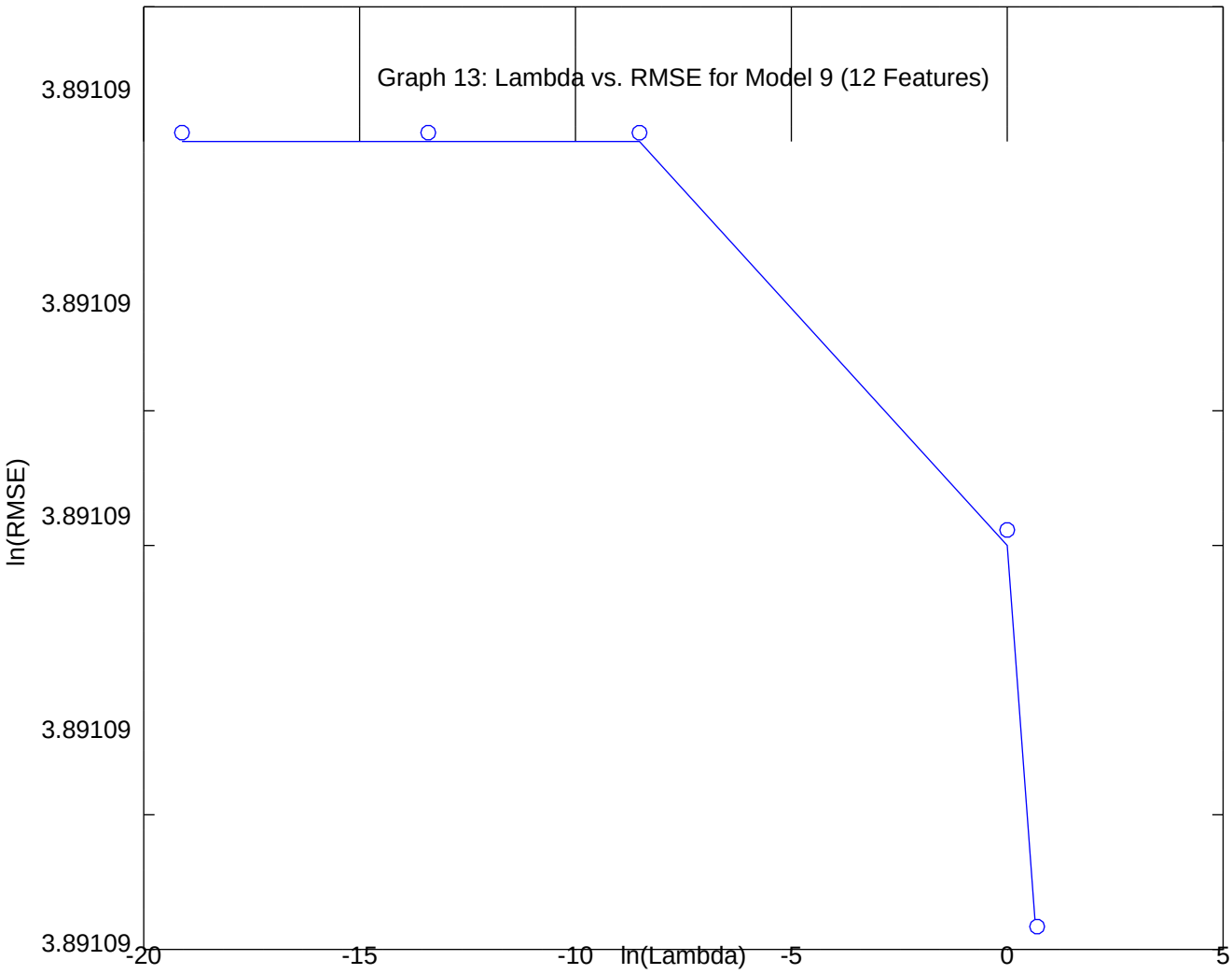


Graph 11: Lambda vs. RMSE for Model 6 (12 Features)

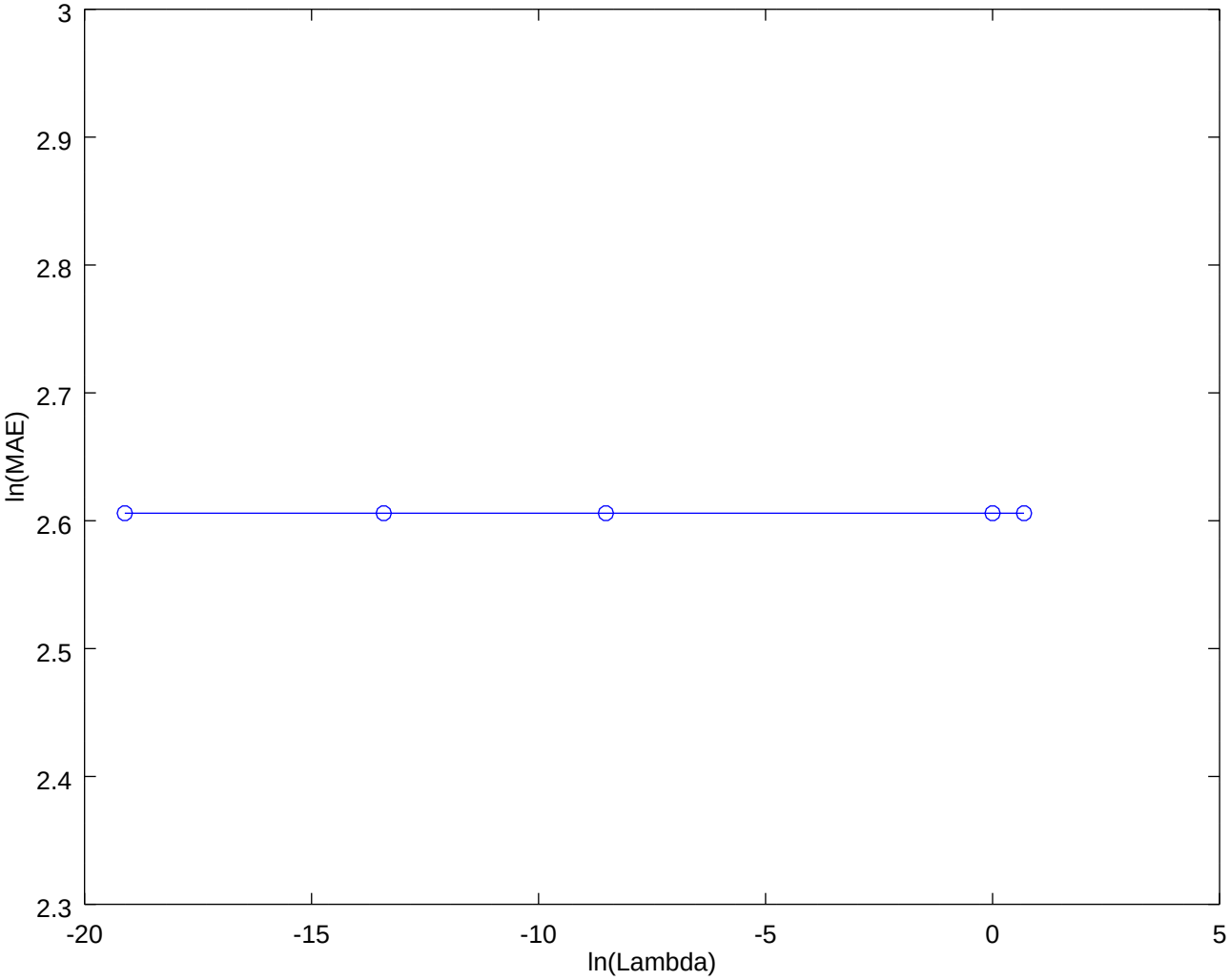


Graph 12: Lambda vs. MAE for Model 9 (12 Features)

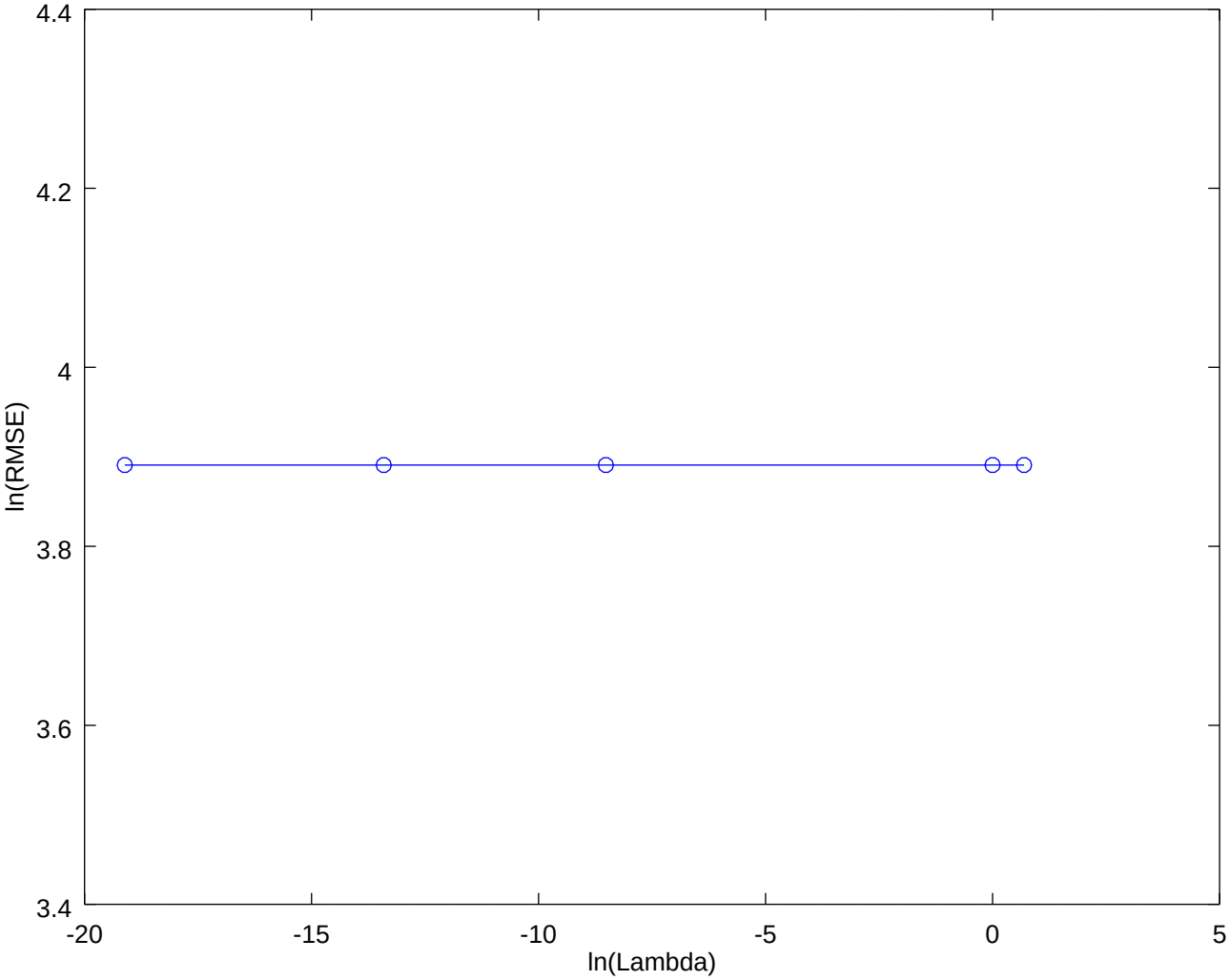



Graph 13: Lambda vs. RMSE for Model 9 (12 Features)
3.89109
3.89109
3.89109
3.89109
3.89109
ln(RMSE)
-20
-15
-10
ln(Lambda)
-5
0
5

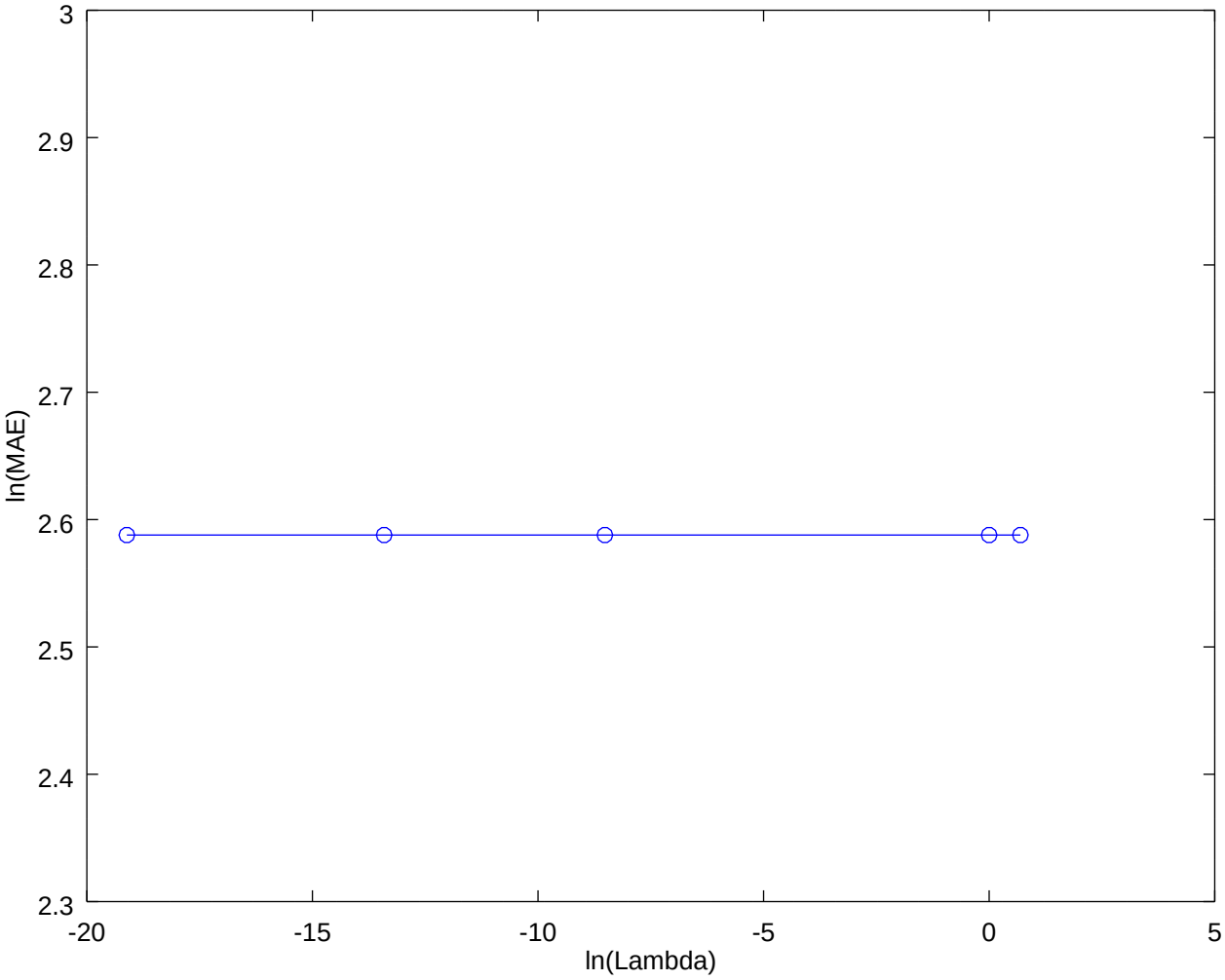
Graph 14: Lambda vs. MAE for Model 6 (Top 5 Features)



Graph 15: Lambda vs. RMSE for Model 6 (Top 5 Features)



Graph 16: Lambda vs. MAE for Model 9 (Top 5 Features)



Graph 17: Lambda vs. RMSE for Model 9 (Top 5 Features)