Introduction to Machine Learning
CSCE 478/878

Programming Assignment 2

Fall 2020

Linear Regression

Part A (Model Code): 478 (68 pts) & 878 (78 pts)
Part B (Data Processing): 478 & 878 (7 pts)
Part C (Model Evaluation): 478 (25 pts) & 878 (35 pts)
Part D (Written Report): 478 & 878 (25 pts)
Extra credit (BONUS) tasks for both 478 & 878: 30 pts

Total: 478 (125 pts) & 878 (145 pts)

Last Name 1: Kim

First Name 1: Changsu

NUID 1: 53634841

Last Name 2: Reyes-Cortes

First Name 2: Jesse

NUID 2: 32637124

Last Name 3: Steiner

First Name 3: Devan

NUID 3: 99064664

Obtained Score:

Introduction to Machine Learning
Assignment Two

Changsu Kim
University of Nebraska-Lincoln
Email: changsu.kim@huskers.unl.edu

Jesse Reyes Cortes
University of Nebraska-Lincoln
Email: jesserc.2@gmail.com

Devan Steiner
University of Nebraska-Lincoln
Email: devan@thesteiners.org

1. Introduction

This assignment guided the group through learning the following four skills: implementing the iterative optimization gradient descent algorithms, implementing various regularization techniques, polynomial regression, and learning curves. The assignment also promoted further growth of the machine learning and python programming skills gained from Assignment One.

1.1. Team Member Contributions

Jesse and Changsu were the primary programmers of the project. Devan was the primary writer of the report. The group worked together to finalize each part.

2. Data Summary

2.1. Dataset and Variables

The dataset used in this assignment consists of several different attributes of red wine. These attributes include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality.

2.2. Variable Descriptive Statistics

Variable Descriptive Statistics

Variable	Mean	Standard Deviation	Quartiles (25%/50%/75%)
Fixed Acidity	6.855	0.844	6.300/6.800/7.300
Volatile Acidity	0.278	0.101	0.210/0.260/0.320
Citric Acid	0.334	0.121	0.270/0.320/0.390
Residual Sugar	6.391	5.072	1.700/5.200/9.900
Chlorides	0.046	0.022	0.036/0.043/0.050
Free Sulfur Dioxide	35.308	17.007	23.000/34.000 /46.000
Total Sulfur Dioxide	138.361	42.498	108.000/134.000 /167.000
Density	0.994	0.003	0.992/0.994/0.996
pH	3.188	0.151	3.090/3.180/3.280
Sulphates	0.490	0.114	0.410/0.470/0.550
Alcohol	10.514	1.231	9.500/10.400 /11.400
Quality	5.878	0.886	5.000/6.000/6.000

2.3. Feature Scaling

Feature scaling was performed in the form on standardization. This was to ensure that each feature would have equal weights when computing using a distance formula.

2.4. Dropped Features

No features were dropped during this assignment. This was mostly due to an instruction in "Part C: Model Evaluation" to not augment any features and to use the data as it is. We decided to keep the data the same throughout the rest of the assignment.

3. Methods

3.1. Closed Form vs. Iterative Optimization

For this dataset, iterative optimization is more suitable than the closed-form solution. This is due to iterative optimization being able to account for when the dataset is non-invertible, as well as producing lower error rates in our experiments.

3.2. Batch vs. Stochastic - Gradient Descent

Batch Gradient Descent uses chunks of samples at a time, while Stochastic Gradient Descent looks at each sample one by one to find values. They would both end up descending to similar weights. If a large learning rate was used, both algorithms could struggle to find optimal weights.

4. Results

4.1. First Degree Model Evaluation

First Degree Model Evaluation

Index	Lambda	Learning Rate	Regularizer	Average Error
0	1.0000	0.1000	l1	0.697243
1	1.0000	0.1000	l2	0.697083
2	1.0000	0.0100	l1	0.704654
3	1.0000	0.0100	l2	0.704273
4	1.0000	0.0010	l1	4.444975
5	1.0000	0.0010	l2	4.443295
6	1.0000	0.0001	l1	20.503144
7	1.0000	0.0001	l2	20.502910
8	0.0000	0.1000	l1	0.697590
9	0.0000	0.1000	l2	0.697590
10	0.0000	0.0100	l1	0.704754
11	0.0000	0.0100	l2	0.704754
12	0.0000	0.0010	l1	4.445160
13	0.0000	0.0010	l2	4.445160
14	0.0000	0.0001	l1	20.503175
15	0.0000	0.0001	l2	20.503175
16	0.1000	0.1000	l1	0.697555
17	0.1000	0.1000	l2	0.697539
18	0.1000	0.0100	l1	0.704744
19	0.1000	0.0100	l2	0.704706
20	0.1000	0.0010	l1	4.445141
21	0.1000	0.0010	l2	4.444973
22	0.1000	0.0001	l1	20.503172
23	0.1000	0.0001	l2	20.503148
24	0.0100	0.1000	l1	0.697587

First Degree Model Evaluation (Continued)

Index	Lambda	Learning Rate	Regularizer	Average Error
25	0.0100	0.1000	l2	0.697585
26	0.0100	0.0100	l1	0.704753
27	0.0100	0.0100	l2	0.704749
28	0.0100	0.0010	l1	4.445158
29	0.0100	0.0010	l2	4.445141
30	0.0100	0.0001	l1	20.503174
31	0.0100	0.0001	l2	20.503172
32	0.0010	0.1000	l1	0.697590
33	0.0010	0.1000	l2	0.697590
34	0.0010	0.0100	l1	0.704754
35	0.0010	0.0100	l2	0.704754
36	0.0010	0.0010	l1	4.445159
37	0.0010	0.0010	l2	4.445158
38	0.0010	0.0001	l1	20.503175
39	0.0010	0.0001	l2	20.503174
40	0.0001	0.1000	l1	0.697590
41	0.0001	0.1000	l2	0.697590
42	0.0001	0.0100	l1	0.704754
43	0.0001	0.0100	l2	0.704754
44	0.0001	0.0010	l1	4.445159
45	0.0001	0.0010	l2	4.445159
46	0.0001	0.0001	l1	20.503175
47	0.0001	0.0001	l2	20.503175

4.2. First Degree Learning Curve

Based on the model evaluation above and the learning curve below, our first degree model seems to be overfitting. This is due to the surprisingly low average error in most scenarios.

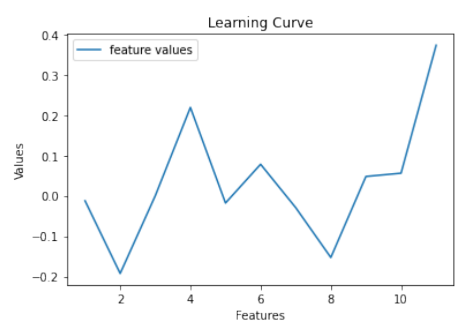

Figure 1. Learning Curve (Degree 1)

4.3. Third Degree Model Evaluation

Third Degree Model Evaluation				
Index	Lambda	Learning Rate	Regularizer	Average Error
0	1.0000	0.1000	l1	NaN
1	1.0000	0.1000	l2	NaN
2	1.0000	0.0100	l1	0.650979
3	1.0000	0.0100	l2	0.650444
4	1.0000	0.0010	l1	4.965780
5	1.0000	0.0010	l2	4.960907
6	1.0000	0.0001	l1	22.066614
7	1.0000	0.0001	l2	22.064489
8	0.0000	0.1000	l1	NaN
9	0.0000	0.1000	l2	NaN
10	0.0000	0.0100	l1	0.651029
11	0.0000	0.0100	l2	0.651029
12	0.0000	0.0010	l1	4.962958
13	0.0000	0.0010	l2	4.962958
14	0.0000	0.0001	l1	22.064753
15	0.0000	0.0001	l2	22.064753
16	0.1000	0.1000	l1	NaN
17	0.1000	0.1000	l2	NaN
18	0.1000	0.0100	l1	0.651013
19	0.1000	0.0100	l2	0.650970
20	0.1000	0.0010	l1	4.963239
21	0.1000	0.0010	l2	4.962753
22	0.1000	0.0001	l1	22.064940
23	0.1000	0.0001	l2	22.064727
24	0.0100	0.1000	l1	NaN
25	0.0100	0.1000	l2	NaN
26	0.0100	0.0100	l1	0.651027
27	0.0100	0.0100	l2	0.651023
28	0.0100	0.0010	l1	4.962986
29	0.0100	0.0010	l2	4.962937
30	0.0100	0.0001	l1	22.064772
31	0.0100	0.0001	l2	22.064751
32	0.0010	0.1000	l1	NaN
33	0.0010	0.1000	l2	NaN
34	0.0010	0.0100	l1	0.651029
35	0.0010	0.0100	l2	0.651028
36	0.0010	0.0010	l1	4.962961
37	0.0010	0.0010	l2	4.962956
38	0.0010	0.0001	l1	22.064755
39	0.0010	0.0001	l2	22.064753
40	0.0001	0.1000	l1	NaN
41	0.0001	0.1000	l2	NaN
42	0.0001	0.0100	l1	0.651029
43	0.0001	0.0100	l2	0.651029
44	0.0001	0.0010	l1	4.962958
45	0.0001	0.0010	l2	4.962958
46	0.0001	0.0001	l1	22.064754
47	0.0001	0.0001	l2	22.064753

4.4. Third Degree Learning Curve

Similarly to the first degree model, the third degree model evaluation and learning curve seem to be overfitting. This is due to most of the average errors being low.

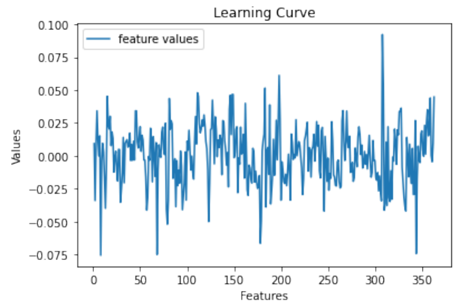

Figure 2. Learning Curve (Degree 3)

5. Conclusion

Overall, we are happy with the results of this assignment. We successfully completed the learning goals previously mentioned: implementing the iterative optimization gradient descent algorithms, implementing various regularization techniques, polynomial regression, and learning curves. We are also much more comfortable with the general topics presented after having a chance to work hands-on with them.