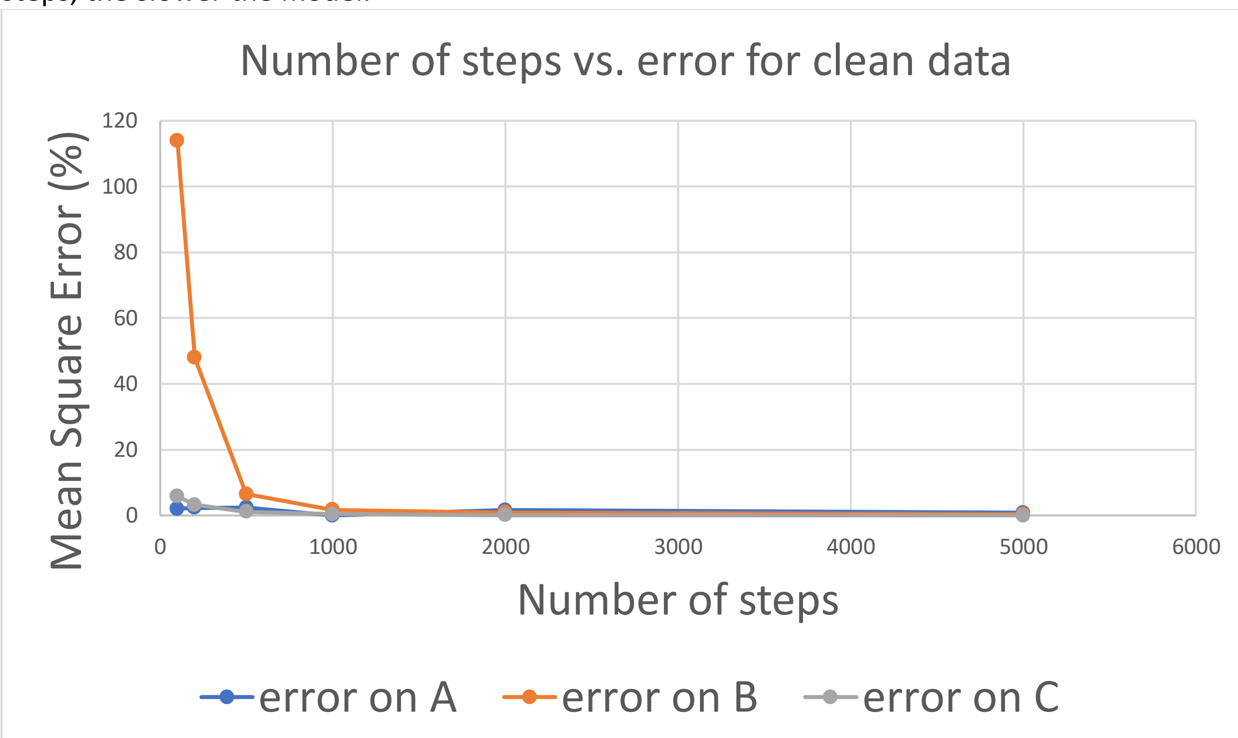
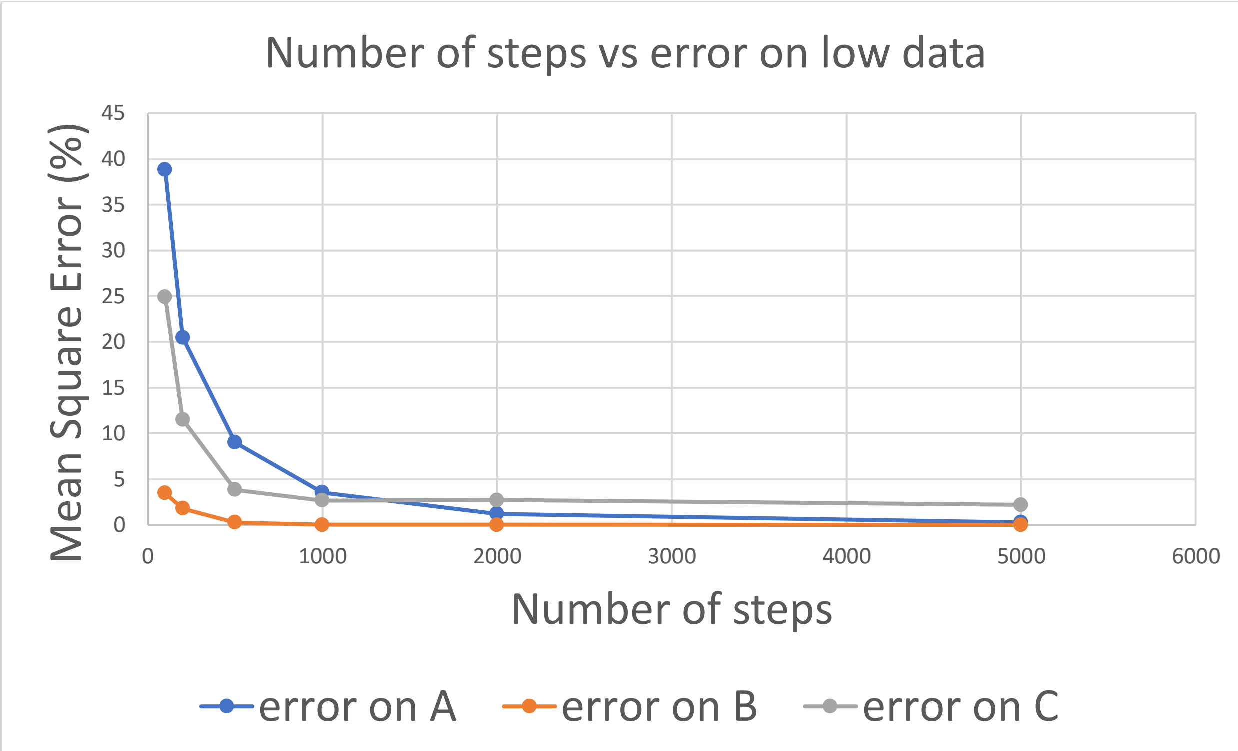
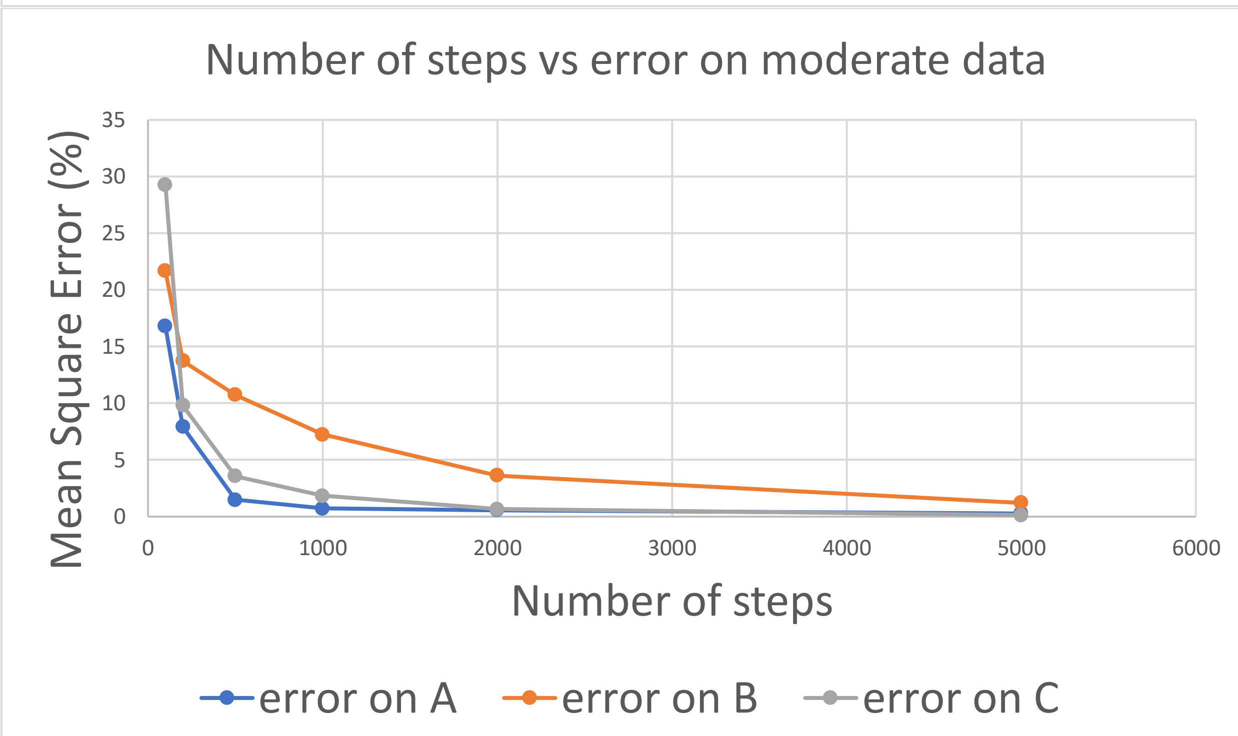
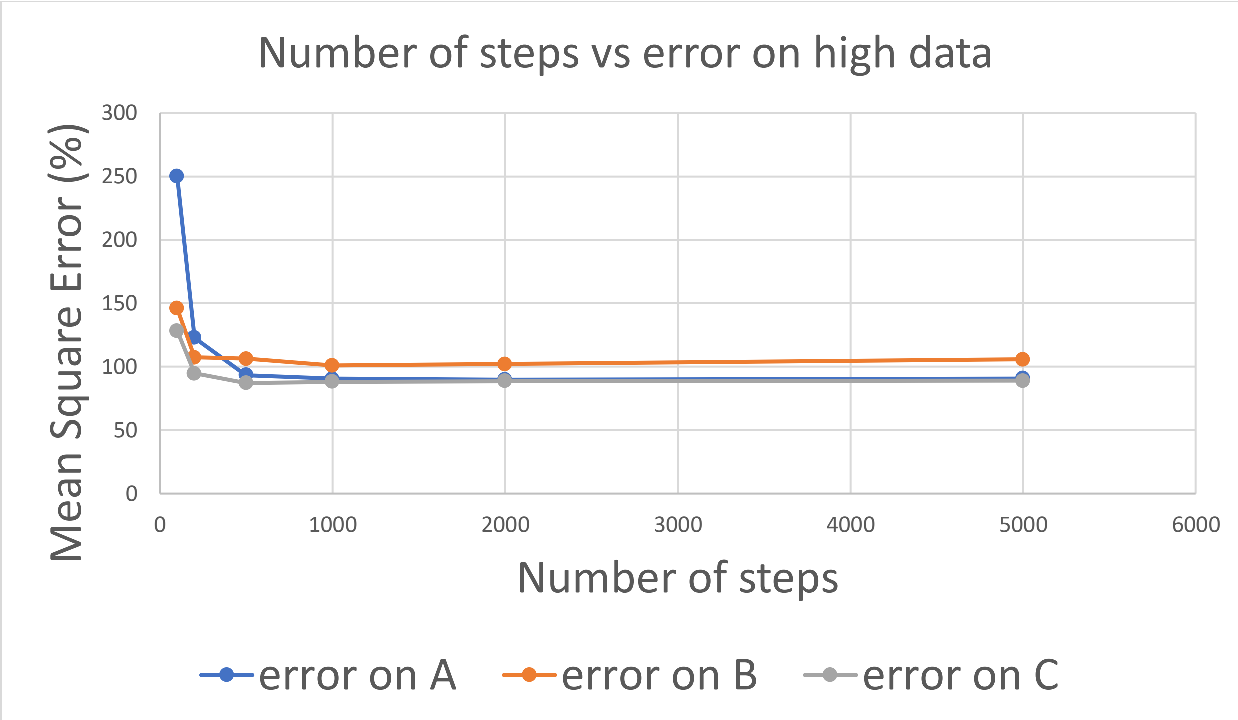
Name: Yousra Awad
Email: yawad2@u.rochester.edu

Experiment #1:
This experiment was conducted using batch gradient descent with a learning rate a=0.01 and a lambda value of 0.01. Since the datasets are relatively small, batch gradient descent was used as it provides better predictions. If the dataset was large, stochastic gradient descent would be a better option for faster weight updates. After trying different values for the learning rate and lambda, this combination seemed to give the best results across datasets with different noise levels. The graphs were generated after experimenting with 100, 200, 500, 1000, 2000, and 5000 steps for each of the datasets provided. The more steps, the less error. However, there's a tradeoff between the number of steps and the speed of generating a prediction. The more the steps, the slower the model.



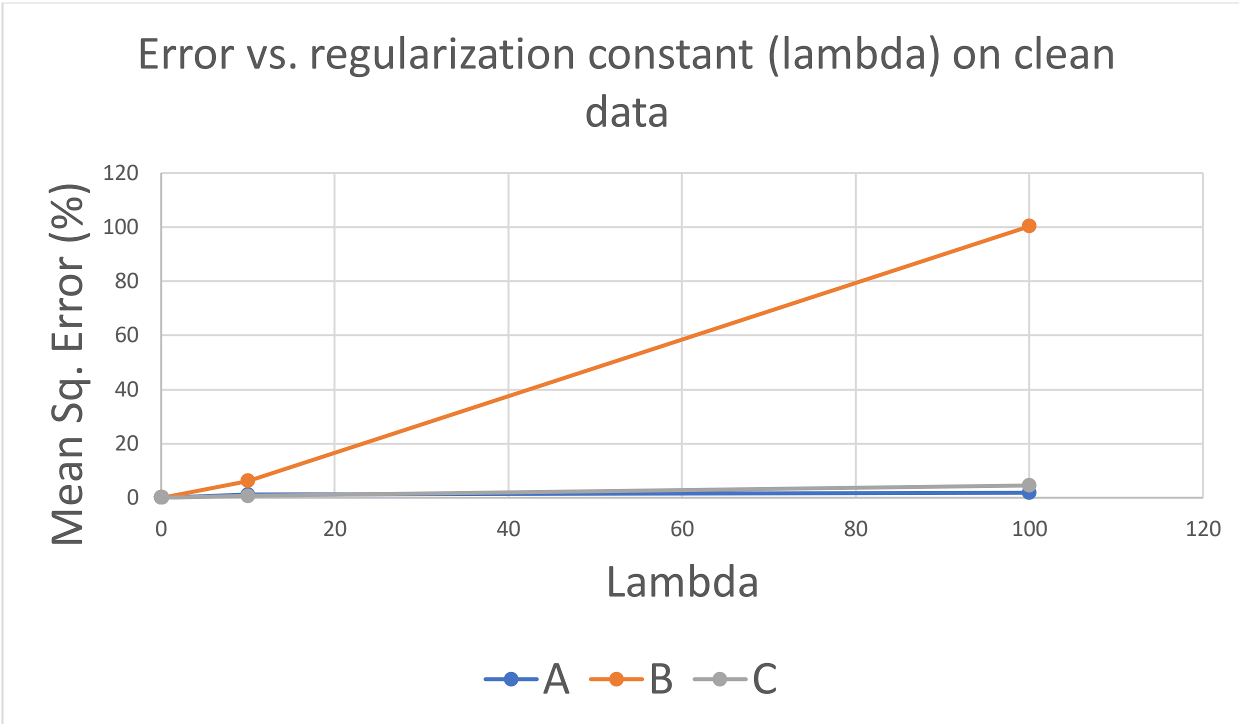

Number of steps vs error on low data
Mean Square Error (%)
45
40
35
30
25
20
15
10
5
0
0
1000
2000
3000
4000
5000
6000
Number of steps
error on A
error on B
error on C

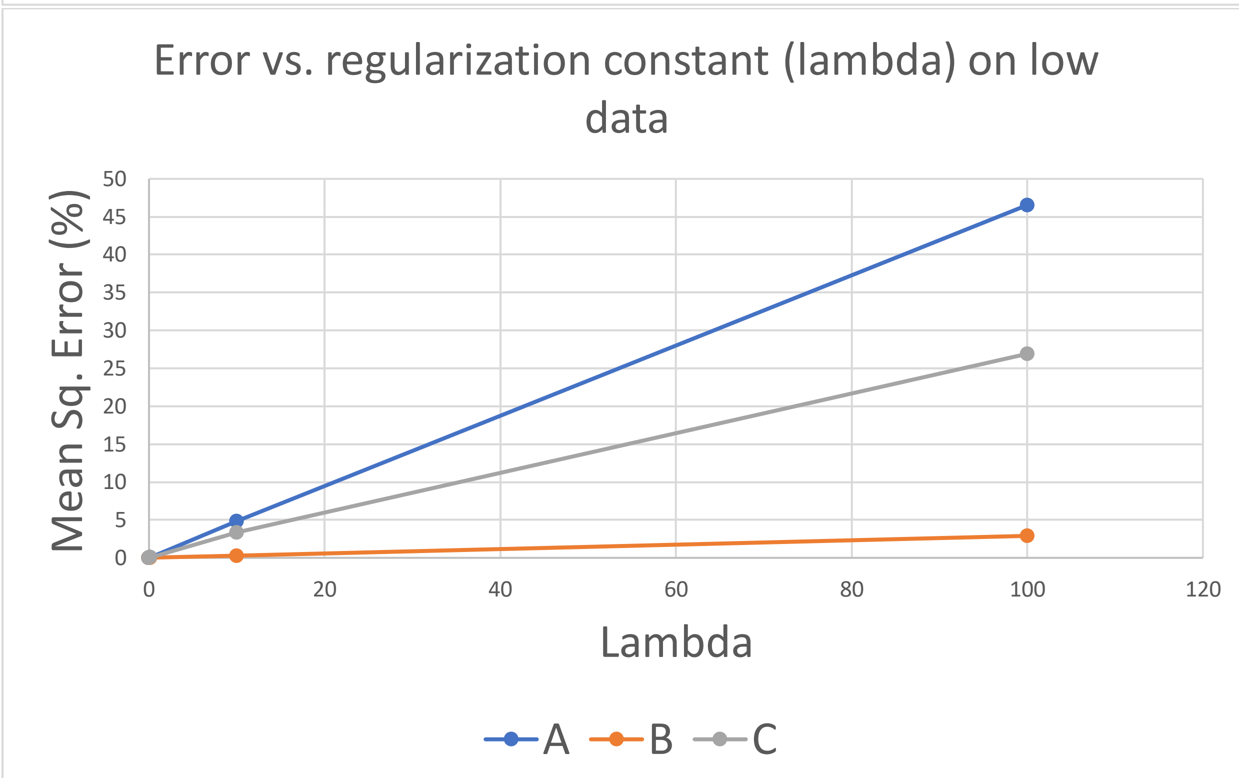

Number of steps vs error on moderate data
Mean Square Error (%)
35
30
25
20
15
10
5
0
0
1000
2000
3000
4000
5000
6000
Number of steps
error on A
error on B
error on C

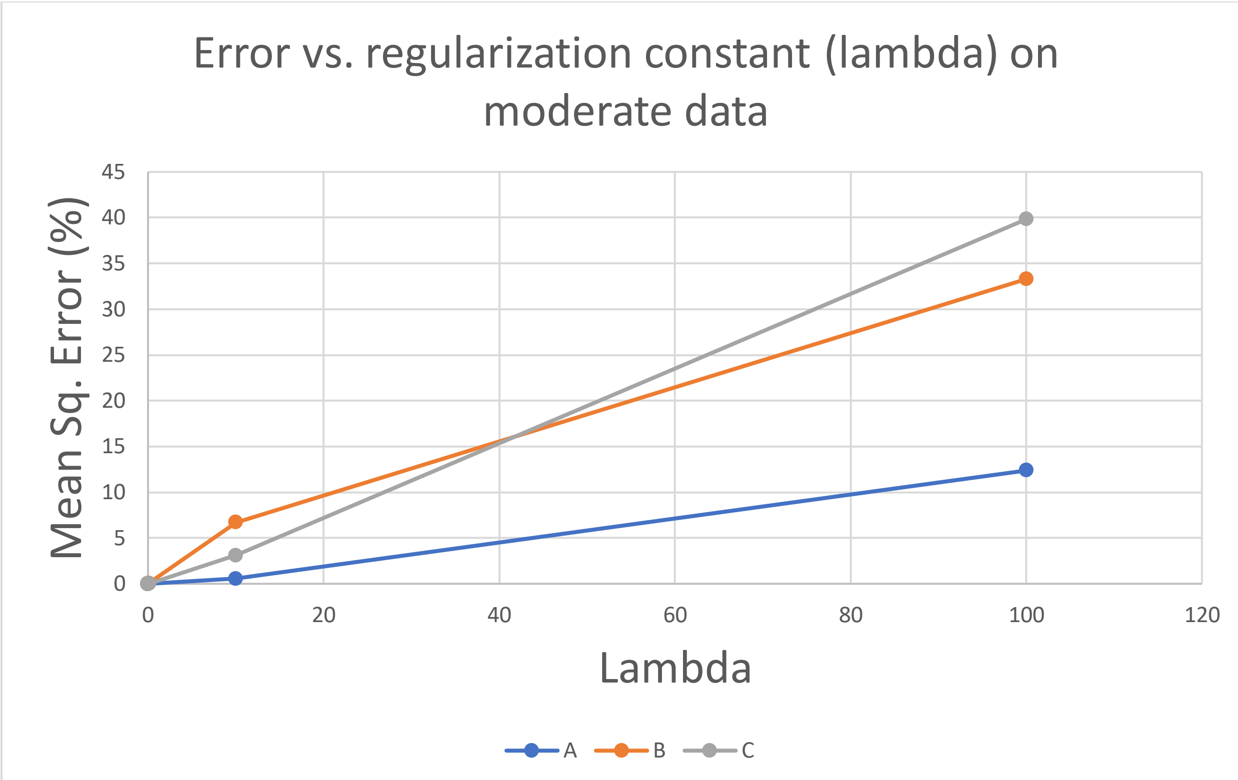


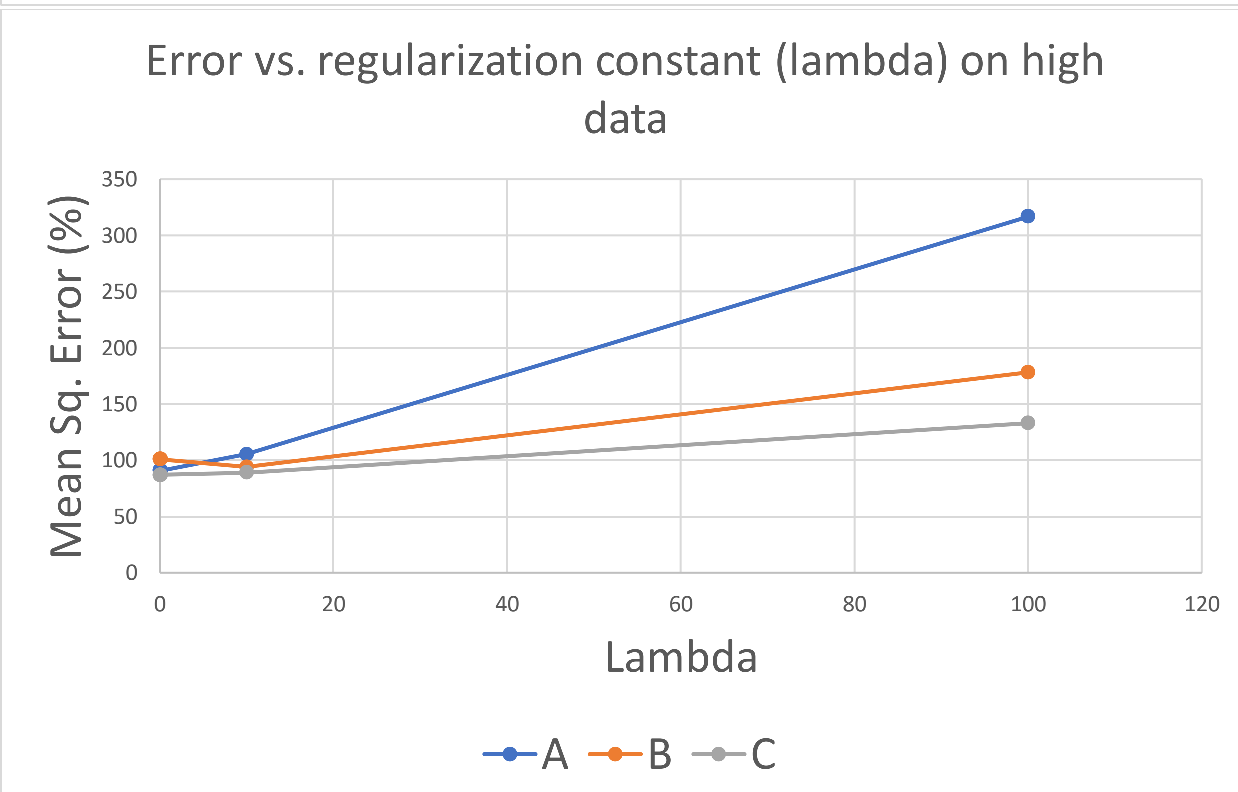
Experiment #2:

This experiment was conducted using different values of regularization constant (lambda). Values of 0, 0.01, 0.1, 10, 100 were used to generate the graphs. For most of the datasets, it seems like the higher the lambda, the higher the error. However, for the high noise data, the error decreased from lambda of 0.1 to lambda of 10. After that point, the error kept increasing. Trying values that are closer to zero with more decimals seemed to generate better predictions.

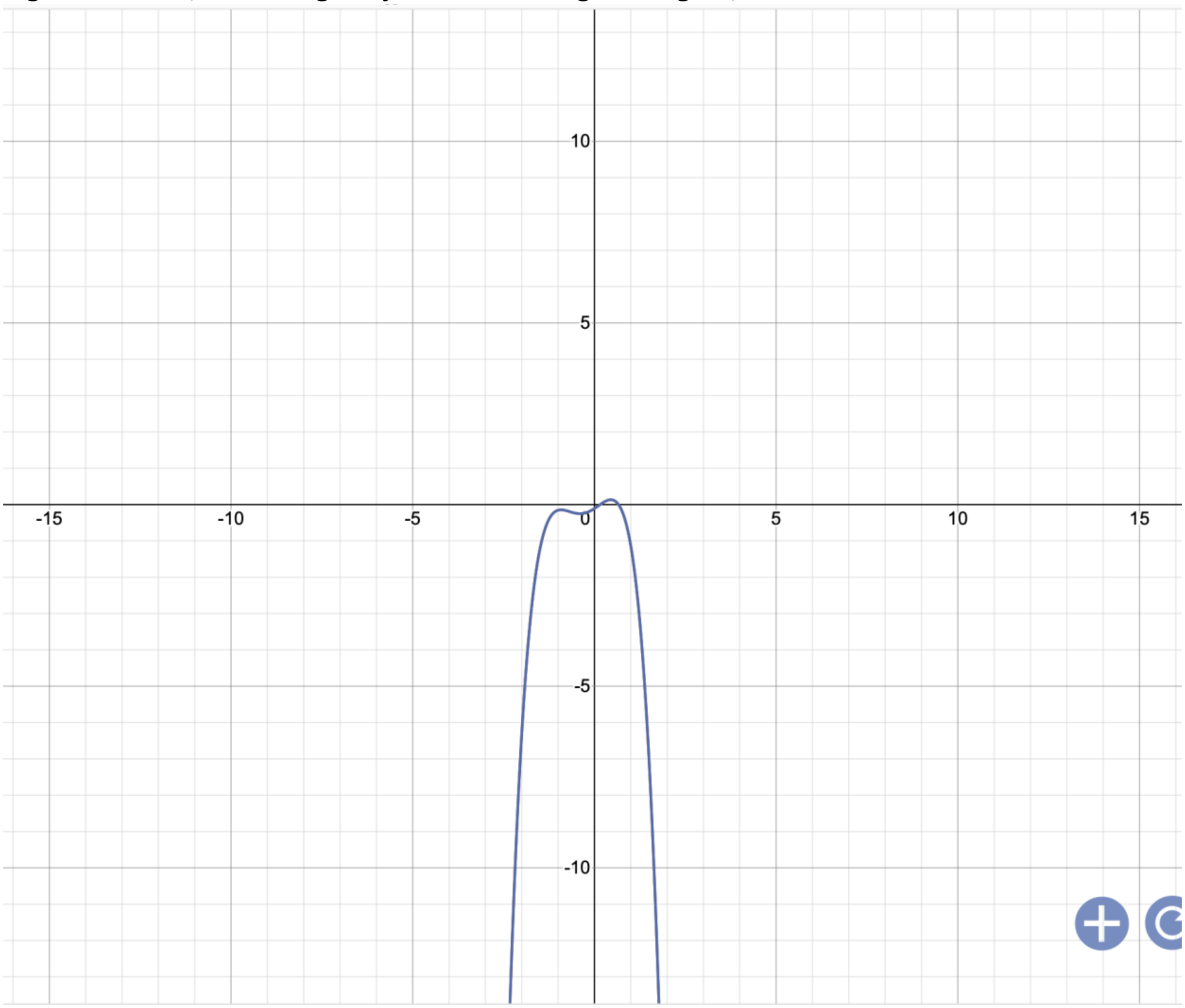

Error vs. regularization constant (lambda) on clean data
Mean Sq. Error (%)
0
20
40
60
80
100
120
0
20
40
60
80
100
120
Lambda
A
B
C


Error vs. regularization constant (lambda) on low data
Mean Sq. Error (%)
0
5
10
15
20
25
30
35
40
45
50
0
20
40
60
80
100
120
Lambda
A
B
C

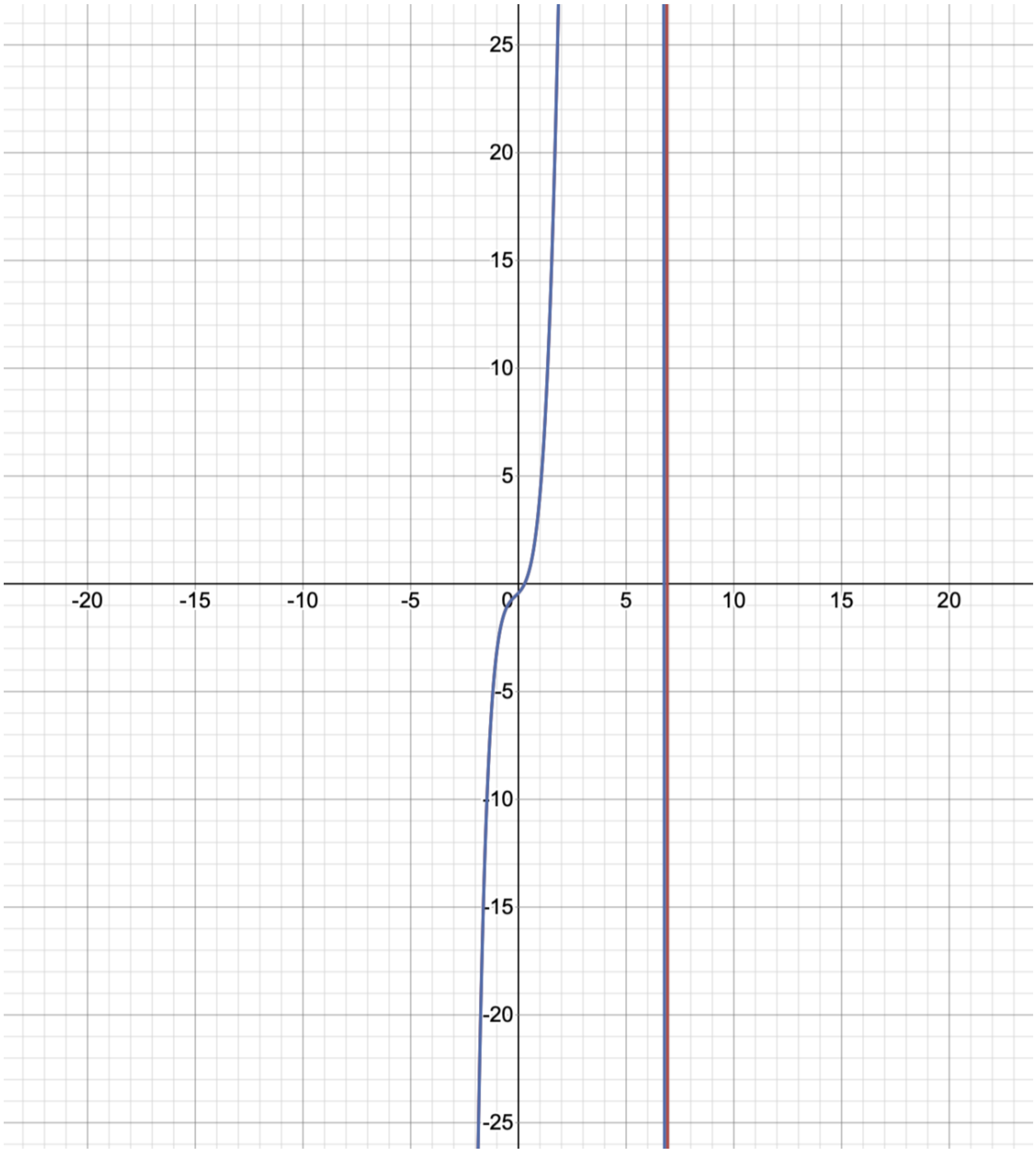

Error vs. regularization constant (lambda) on moderate data
Mean Sq. Error (%)
45
40
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
Lambda
A
B
C


Error vs. regularization constant (lambda) on high data
Mean Sq. Error (%)
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
Lambda
A
B
C

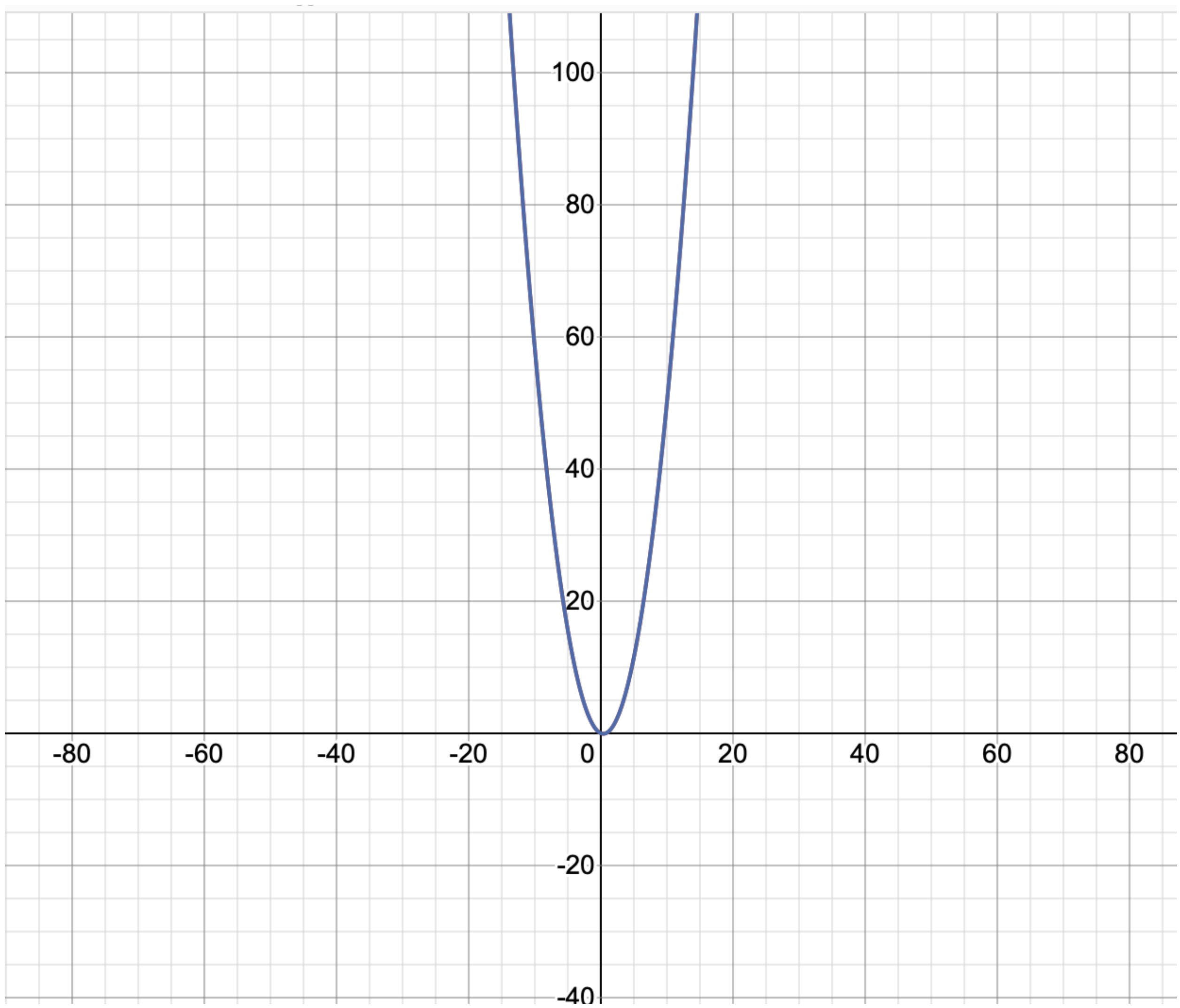
Experiment #3 (Target [red curve] vs predicted polynomial [blue curve]:
The model makes accurate predictions for clean data and data with relatively low noise. For high noise data, it does a good job at estimating the degree, but it's not as accurate.



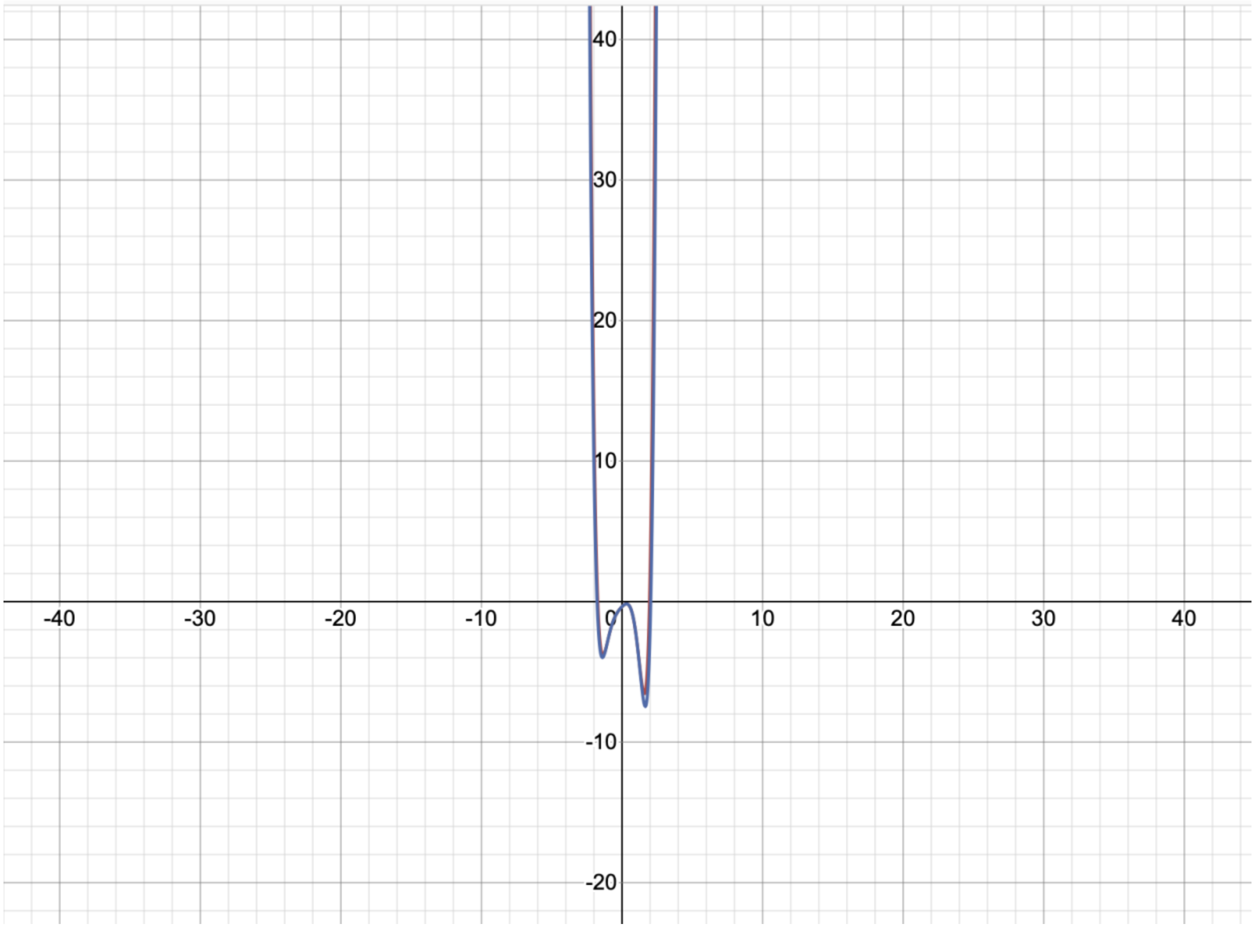
clean data- A, real degree = 4, predicted = 4



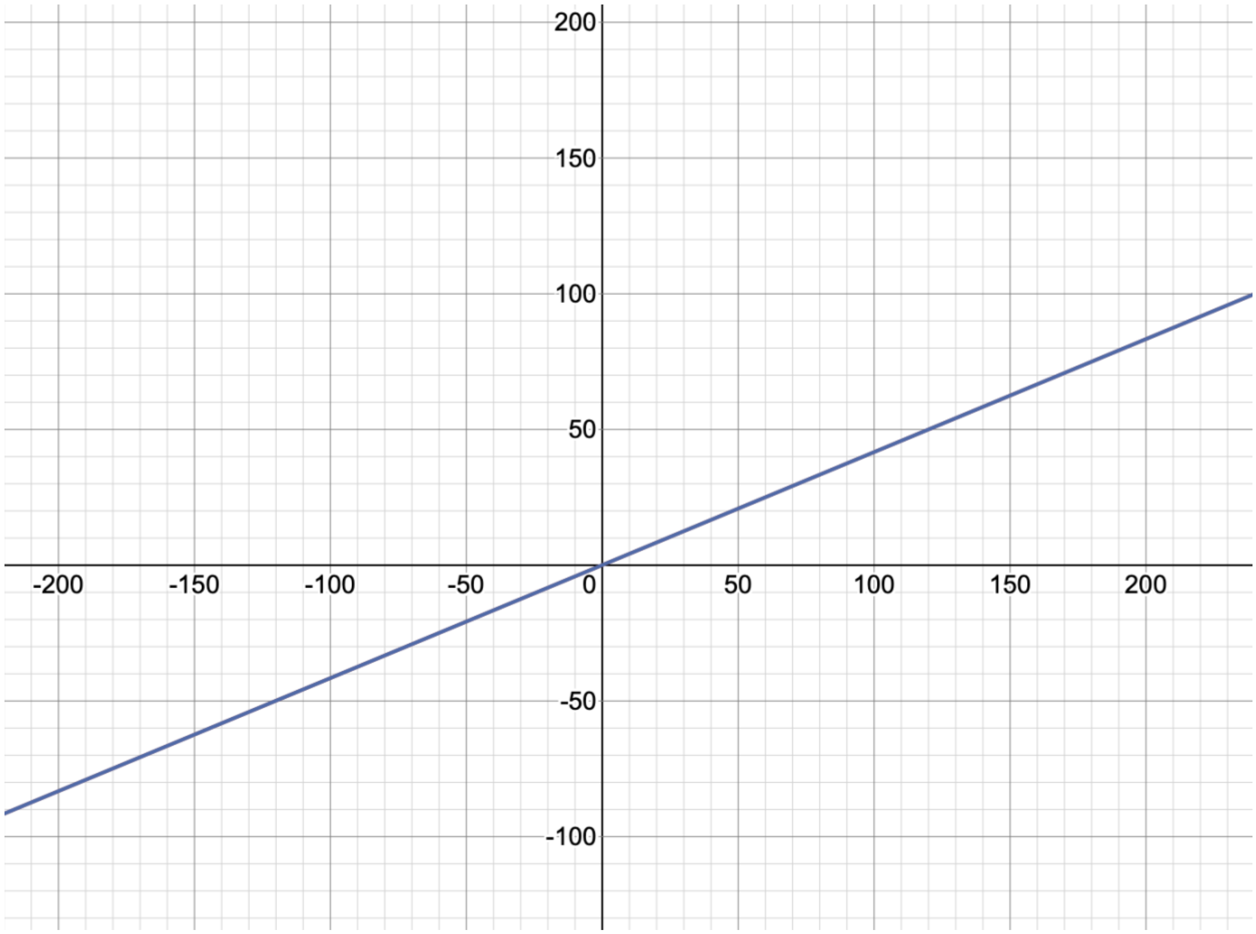
clean data- B, real degree = 6, predicted = 6



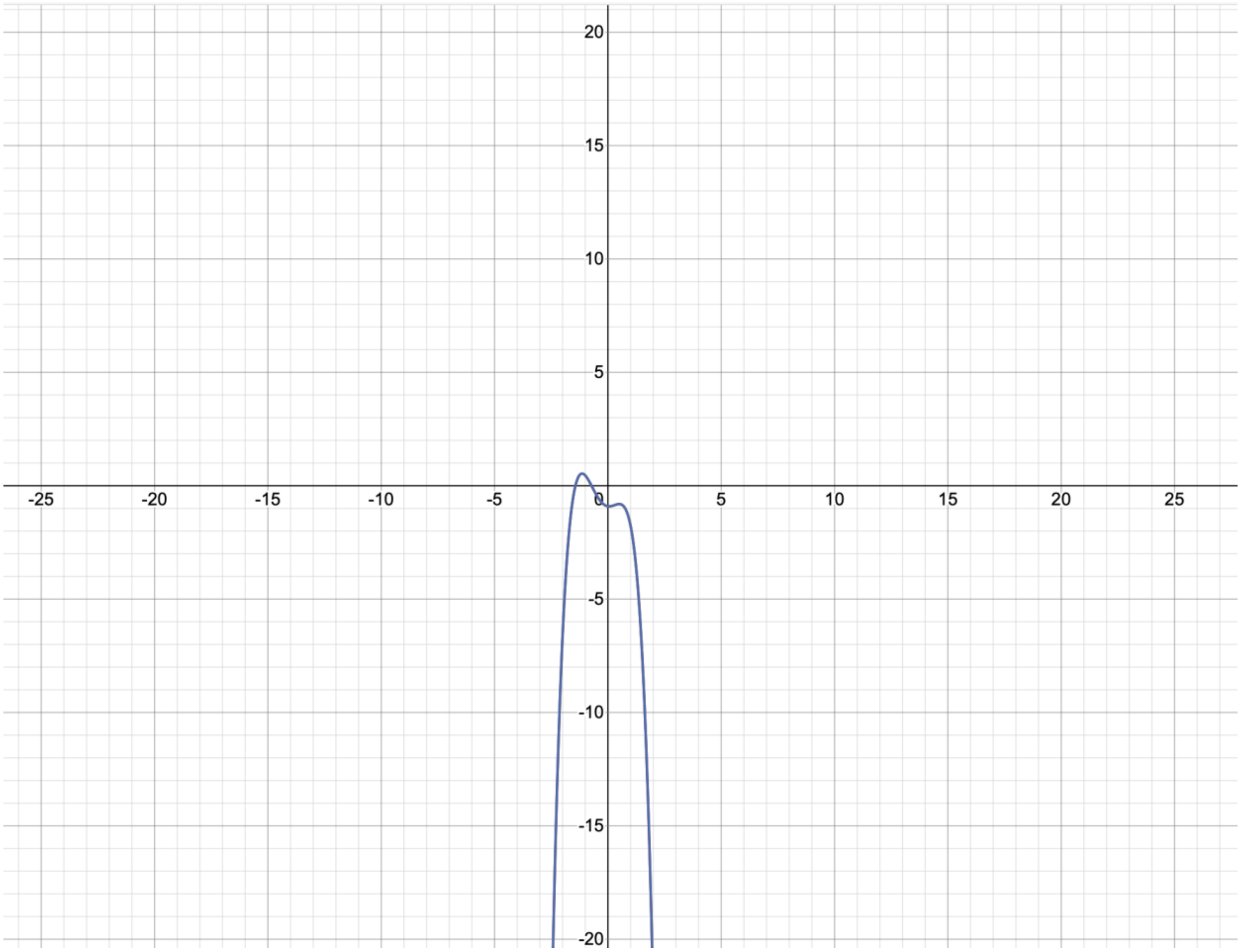
clean data- C, real degree = 2, predicted = 2



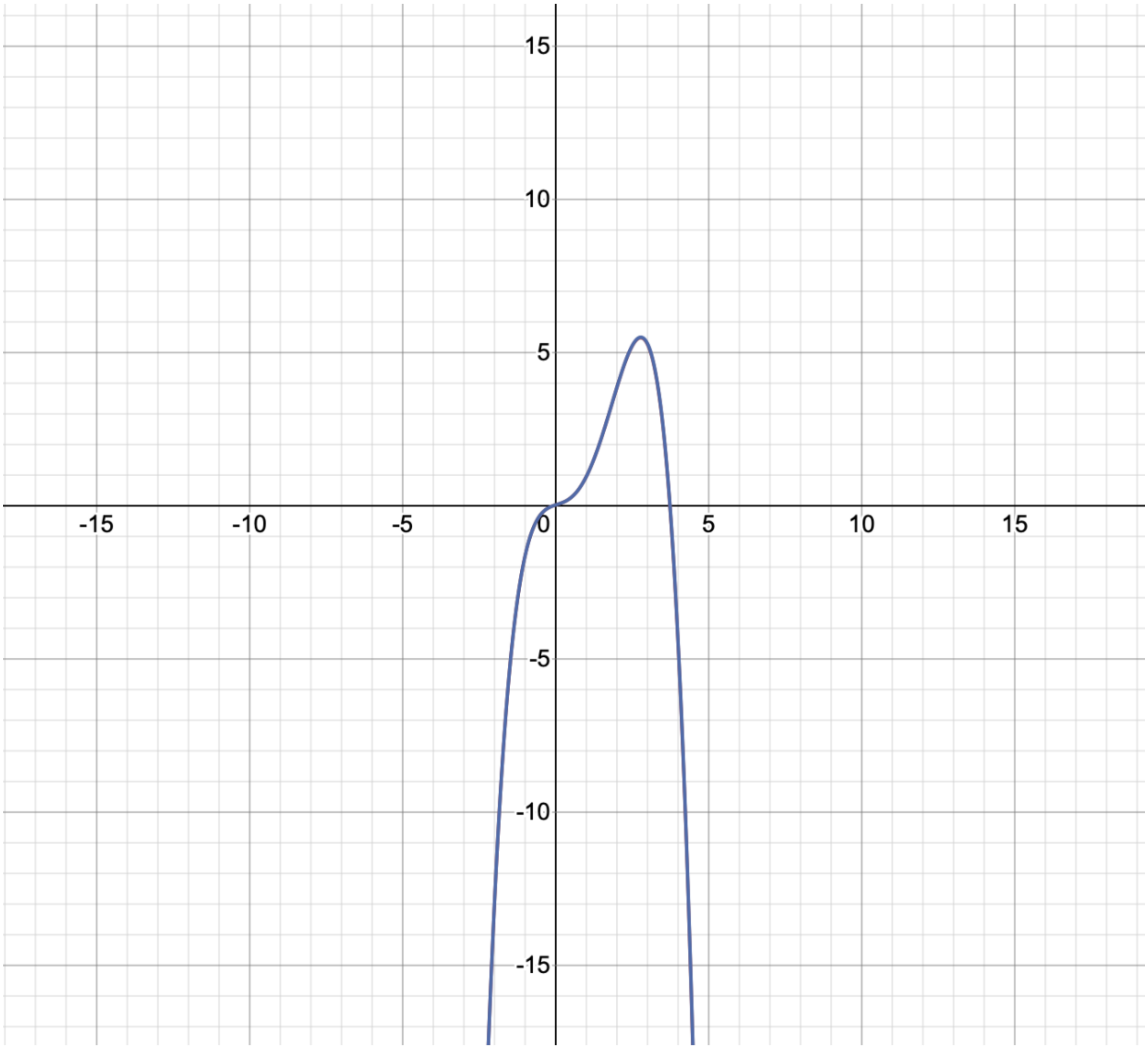
low data- A, real degree = 6, predicted = 6



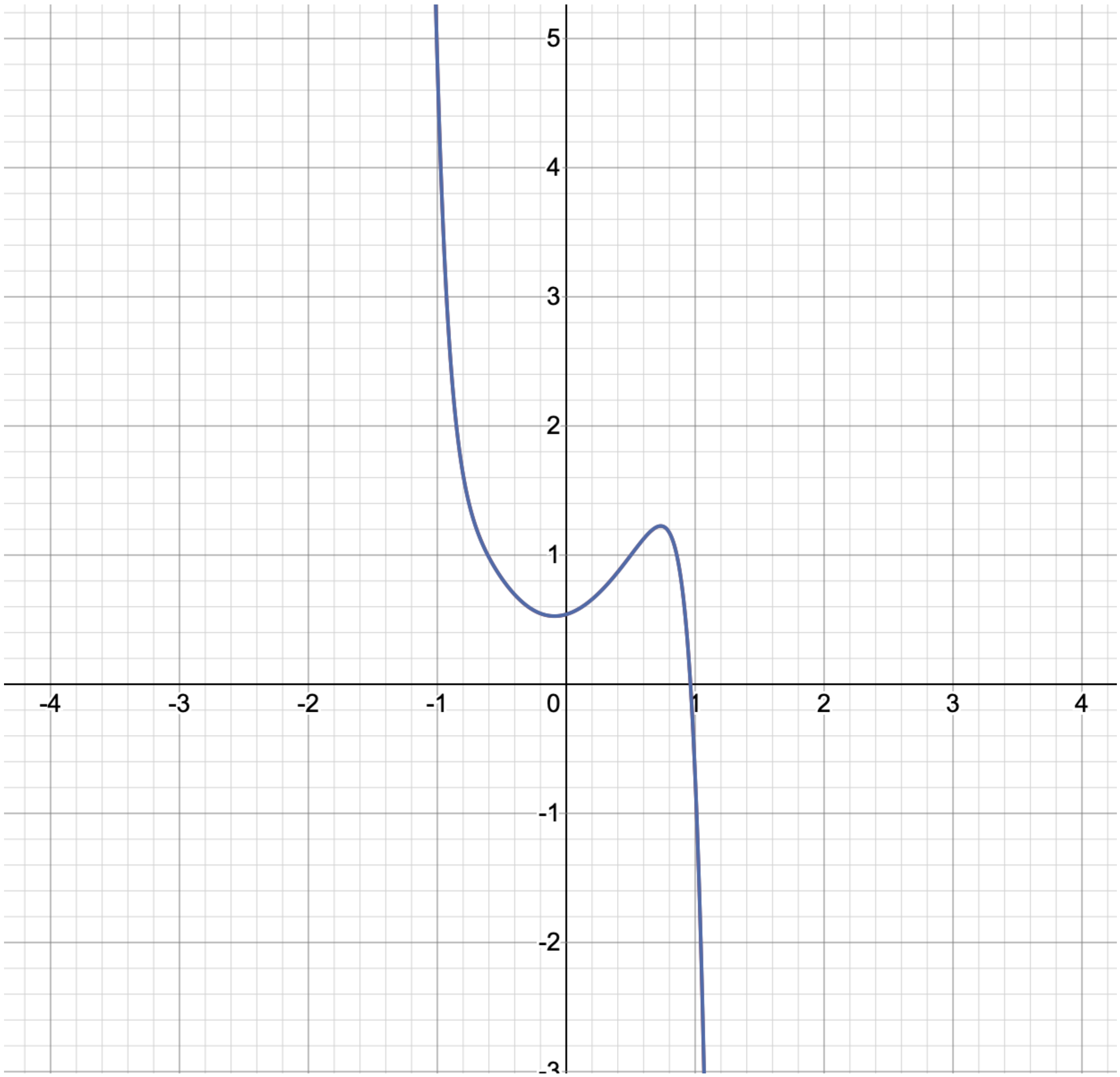
low data- B, real degree = 1, predicted = 1



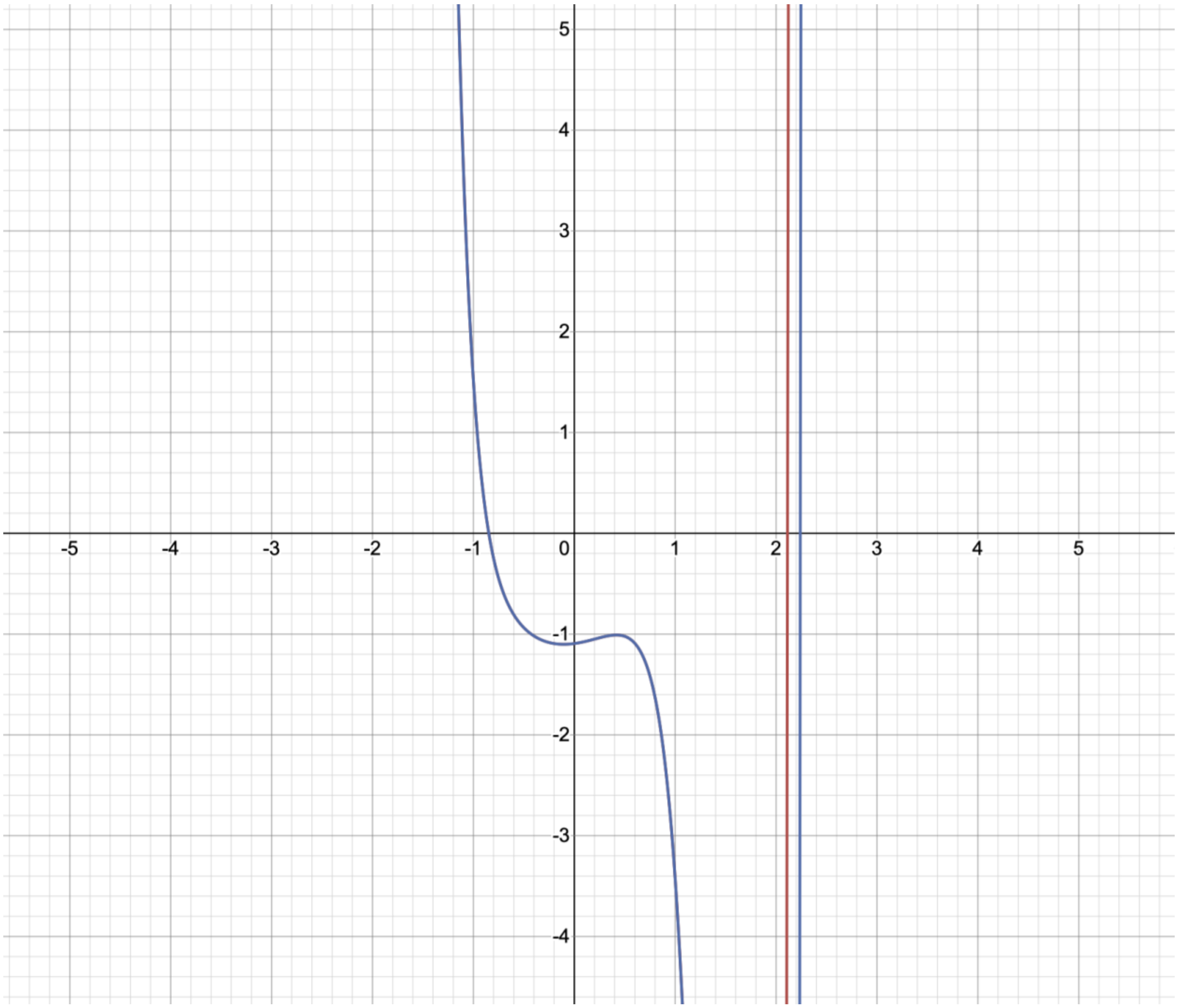
low data- C, real degree = 4, predicted = 4



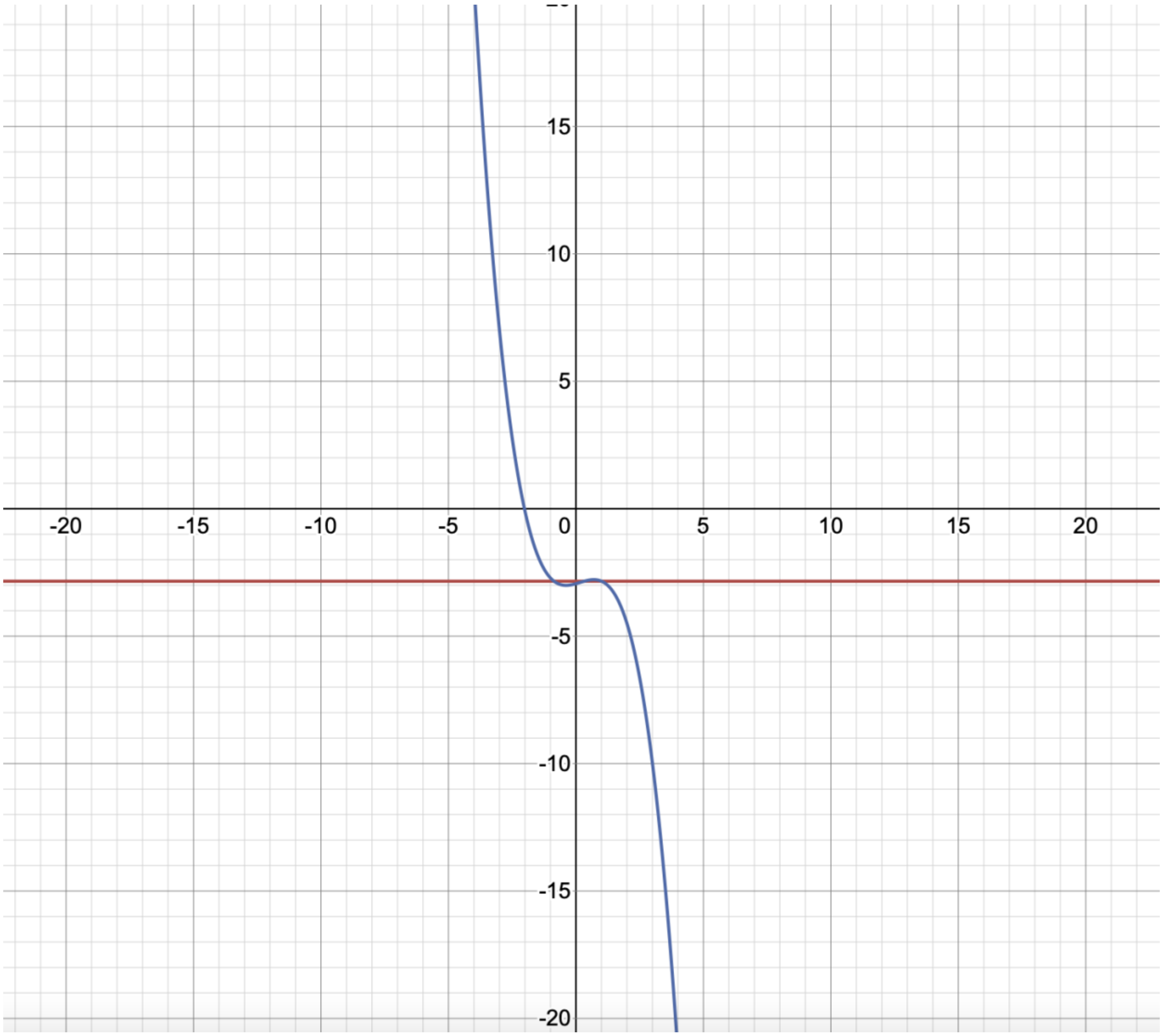
moderate data- A, real degree = 4, predicted = 4



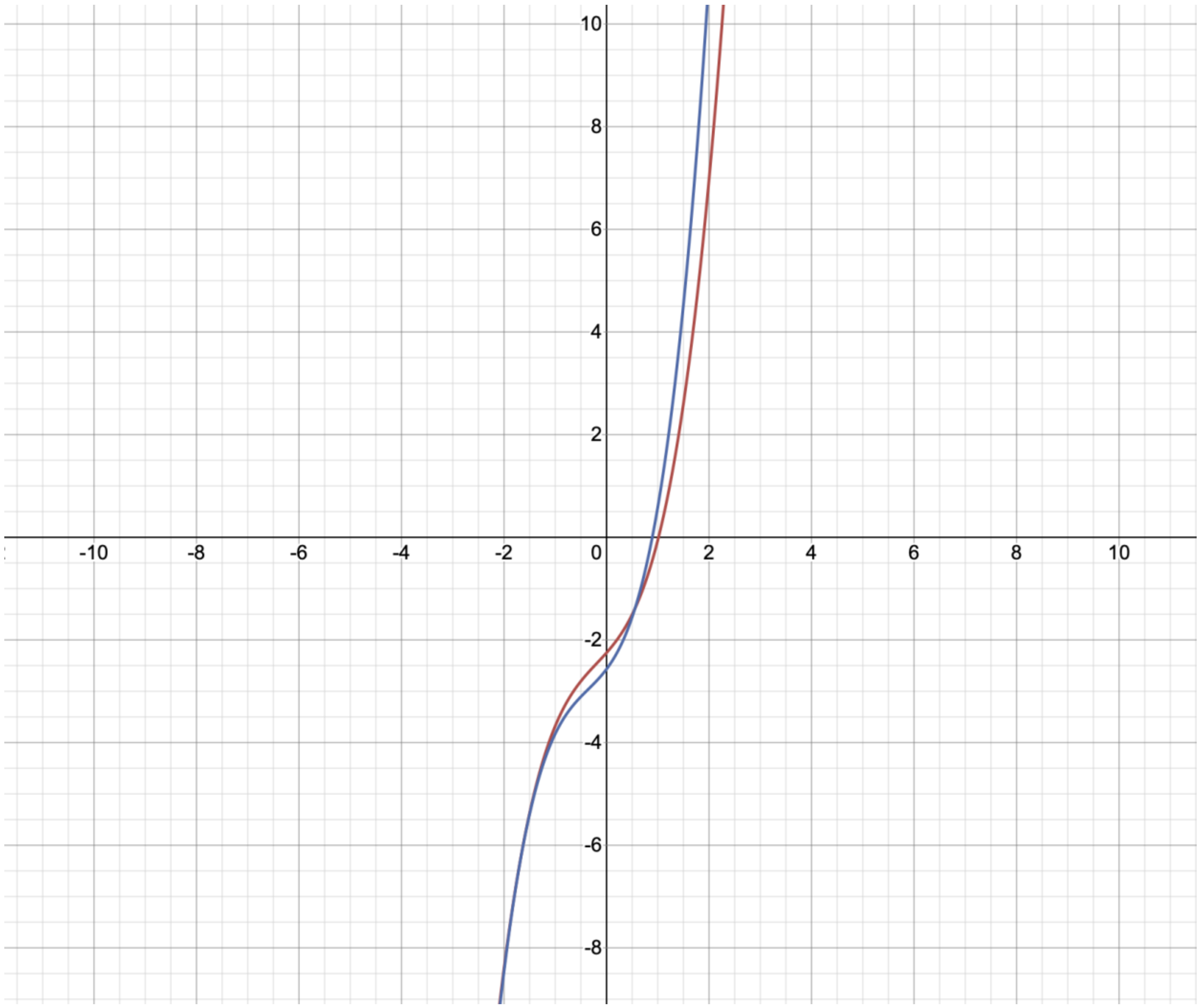
moderate data- B, real degree = 11, predicted = 11



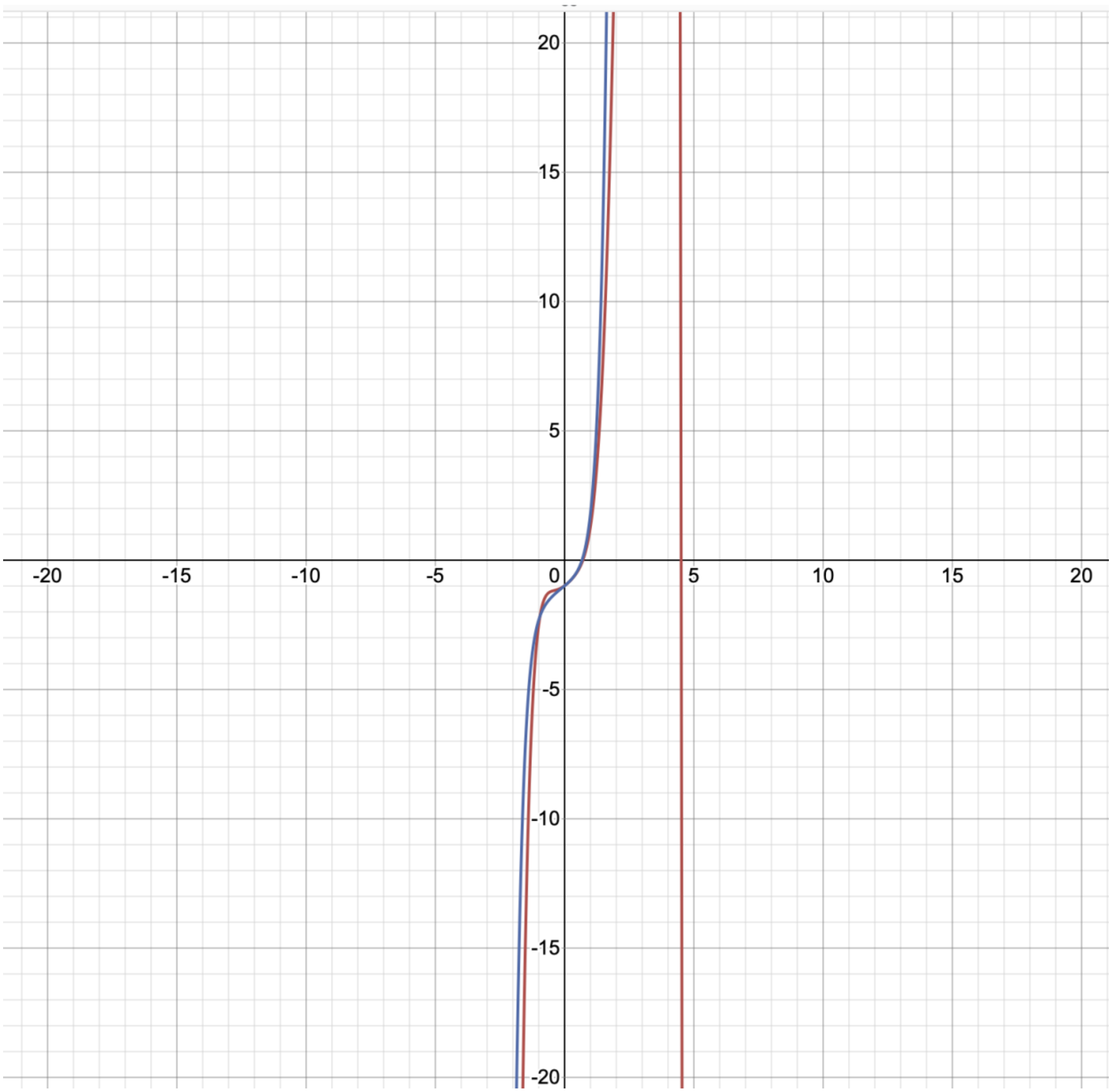
moderate data- C, real degree = 10, predicted = 10



high data- A, real degree = 0, predicted = 3



high data- B, real degree = 3, predicted = 3



high data- C, real degree = 6, predicted = 7