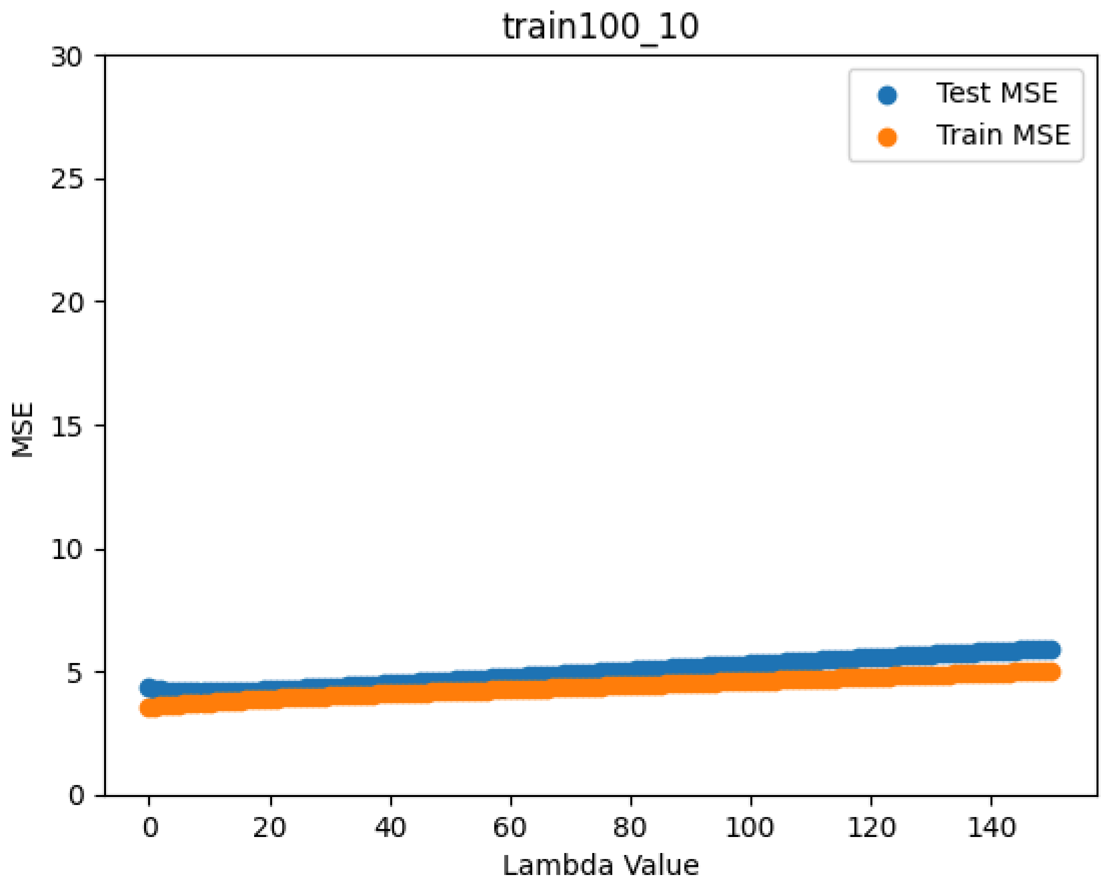
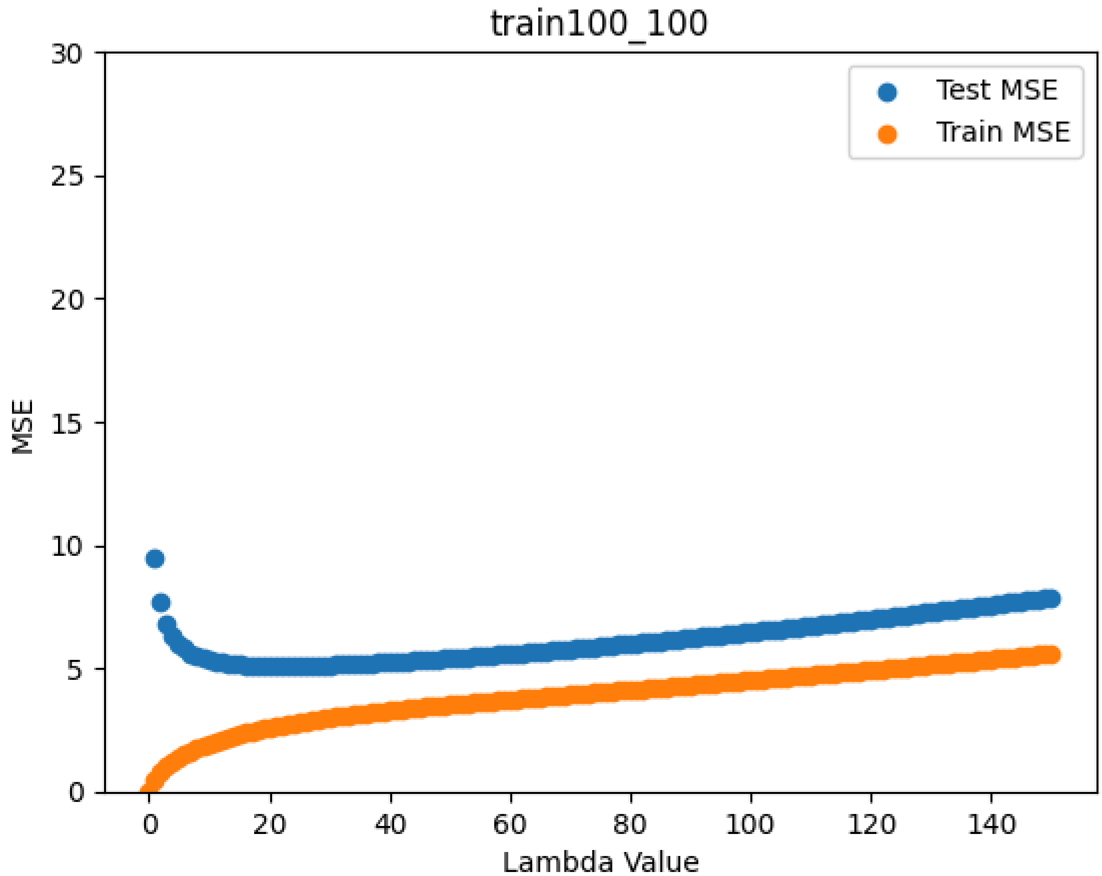
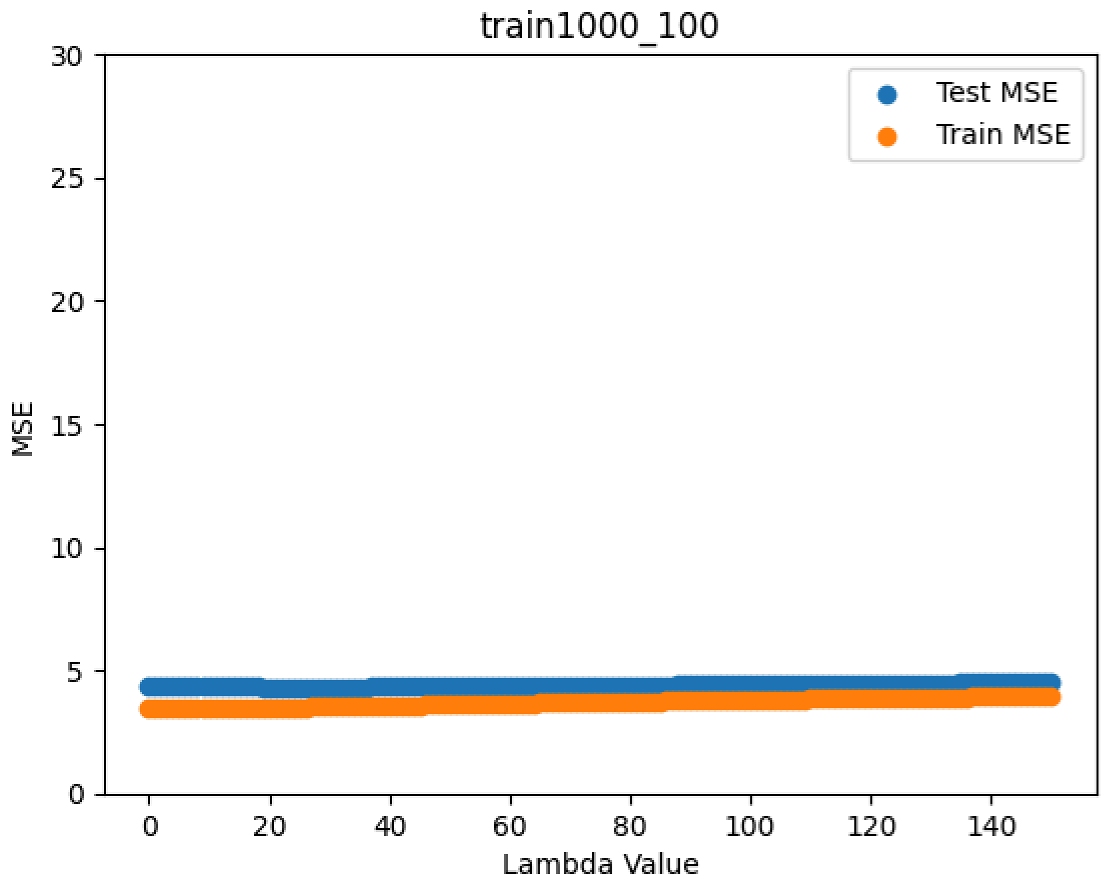
1. **Graphs**

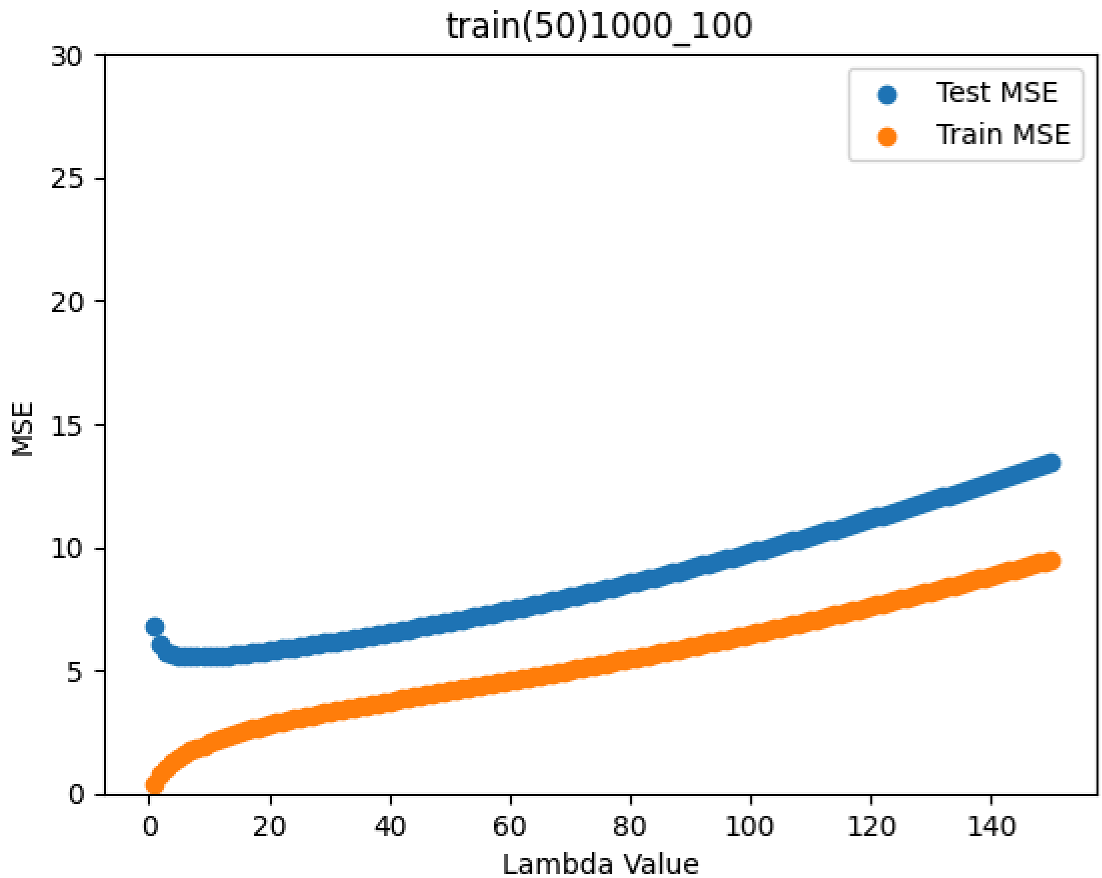


train100_10
Test MSE
Train MSE
MSE
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140
Lambda Value

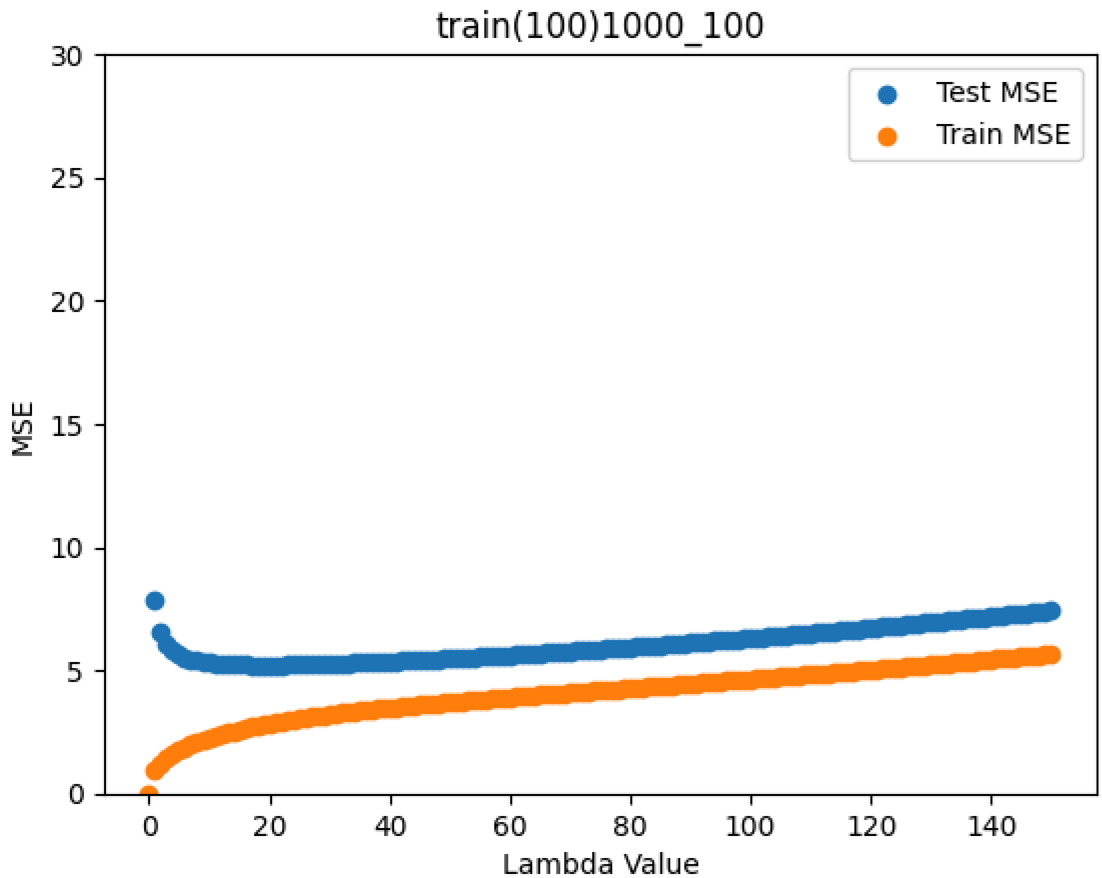

train100_100
Test MSE
Train MSE
MSE
Lambda Value
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140

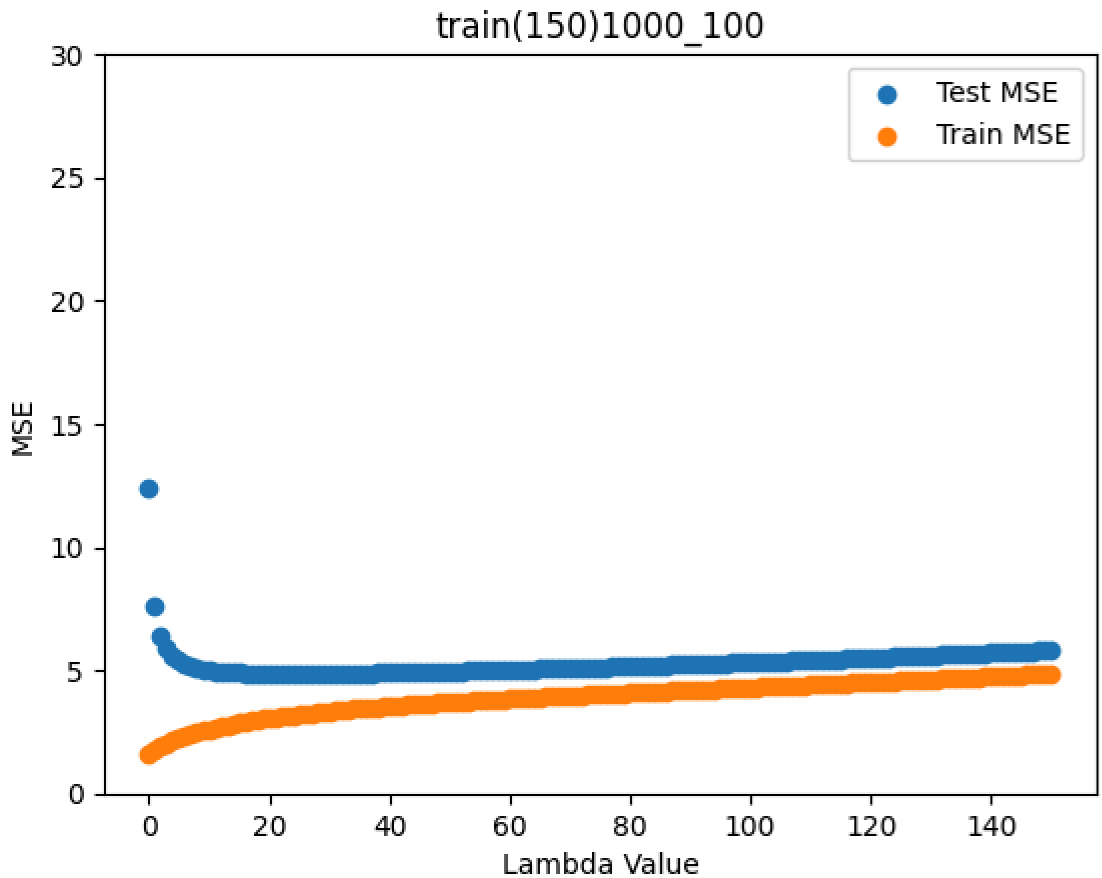


train1000_100
Test MSE
Train MSE
MSE
Lambda Value



train(50)1000_100
Test MSE
Train MSE
MSE
Lambda Value
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140


train(100)1000_100
Test MSE
Train MSE
MSE
Lambda Value
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140



1A.

The values of λ that give the least test set MSE

"train100_10", λ = 8, MSE = 4.159679

"train100_100", λ = 22, MSE = 5.0783

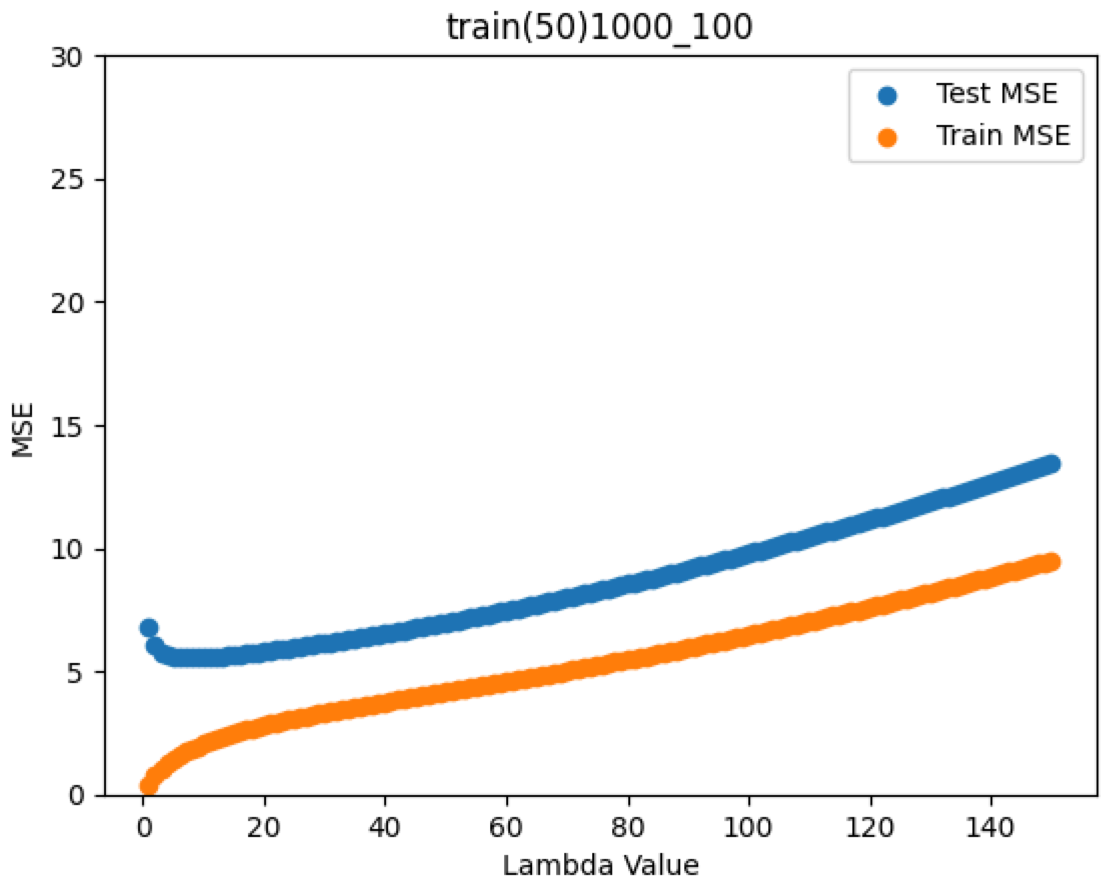
"train1000_100", λ = 27, MSE = 4.315571

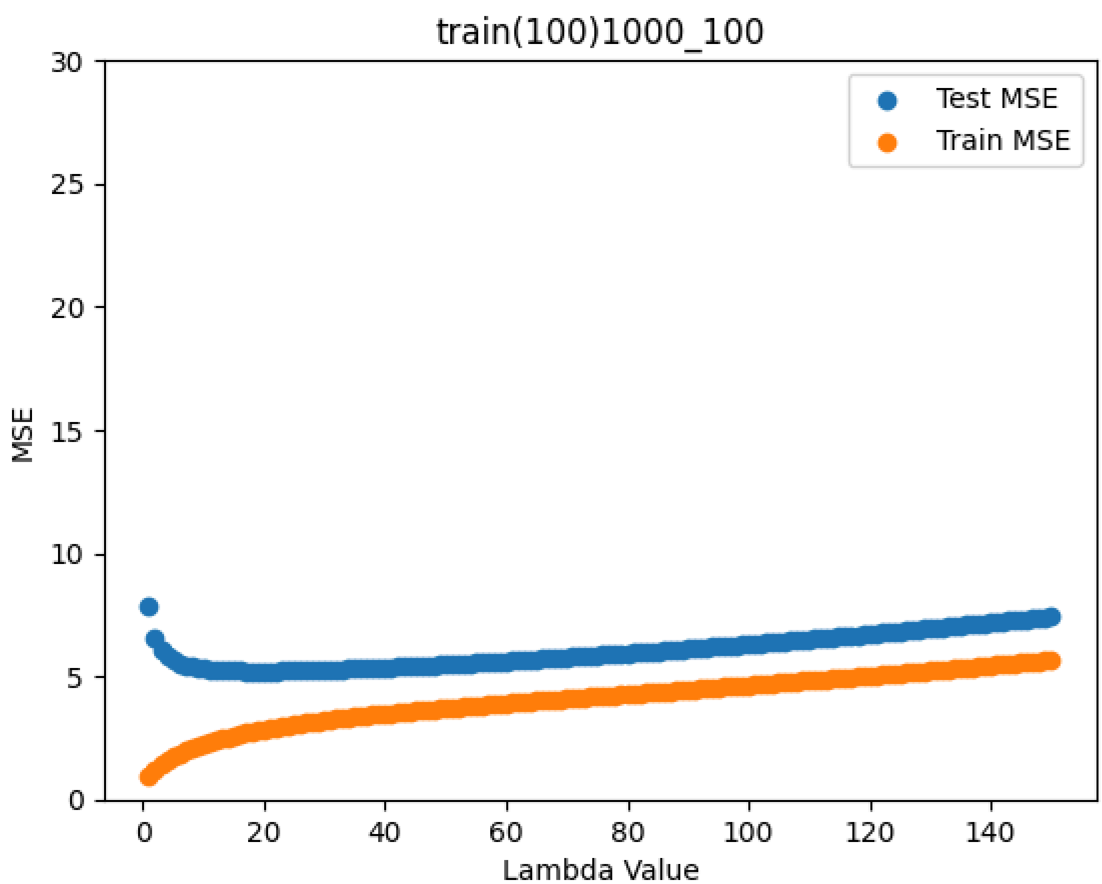
"train(50)1000_100", λ = 8, MSE = 5.540902

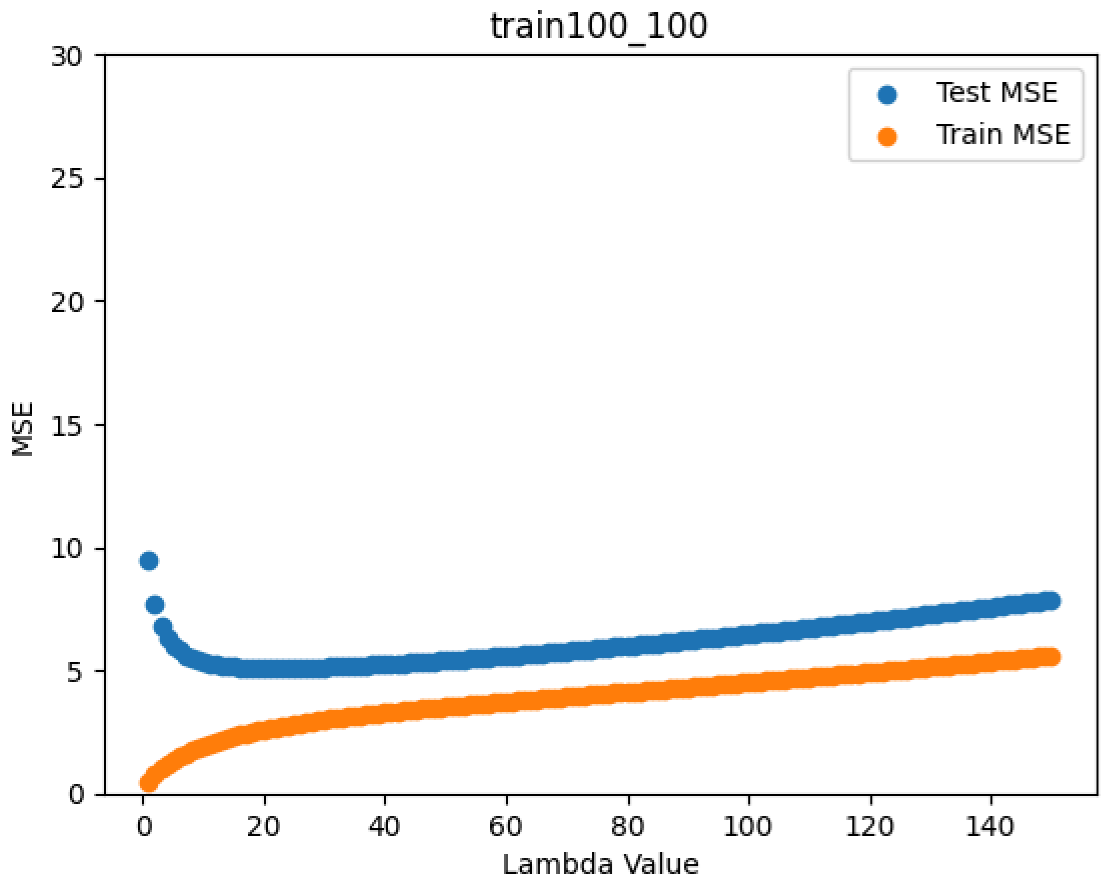
"train(100)1000_100", λ = 19, MSE = 5.205912

"train(150)1000_100", λ = 23, MSE = 4.848943

1B. Graphs on next page(s)


train(50)1000_100
Test MSE
Train MSE
MSE
Lambda Value
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140


train(100)1000_100
Test MSE
Train MSE
MSE
Lambda Value
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140



1C.

The reason why no regularization gives abnormally large MSEs for those datasets is because the size of the datasets is small. With limited amounts of data to train on, testing the data will be more likely to have errors. This is amplified by the lambda value being 0 as there is no regularization being applied within the ridge regression formula. Basically, there are less values and higher variance. Because of these two factors the error will be more extreme because of a lower initial bias.

2A.

The values of λ that give the least test set MSE using CV technique

"train100_10", λ = 12, MSE = 4.158702

"train100_100", λ = 20, MSE = 4.494045

"train1000_100", λ = 39, MSE = 4.137141

"train(50)1000_100", λ = 24, MSE = 5.300696

"train(100)1000_100", λ = 30, MSE = 4.839469

"train(150)1000_100", λ = 46, MSE =4.868577

2B.

The lambda values seem larger for the training datasets that are tested on larger testing sets and the MSE values are all smaller if not relatively equal

"train100_10", pre-cv: λ = 8, MSE = 4.159679

Post-cv:λ = 12, MSE = 4.158702

"train100_100", pre-cv: λ = 22, MSE = 5.0783

Post-cv:λ = 20, MSE = 4.494045

"train1000_100", pre-cv: λ = 27, MSE = 4.315571

Post-cv:λ = 39, MSE = 4.137141

"train(50)1000_100",pre-cv: λ = 8, MSE = 5.540902

Post-cv:λ = 24, MSE = 5.300696

"train(100)1000_100", pre-cv: λ = 19, MSE = 5.205912

Post-cv:λ = 30, MSE = 4.839469

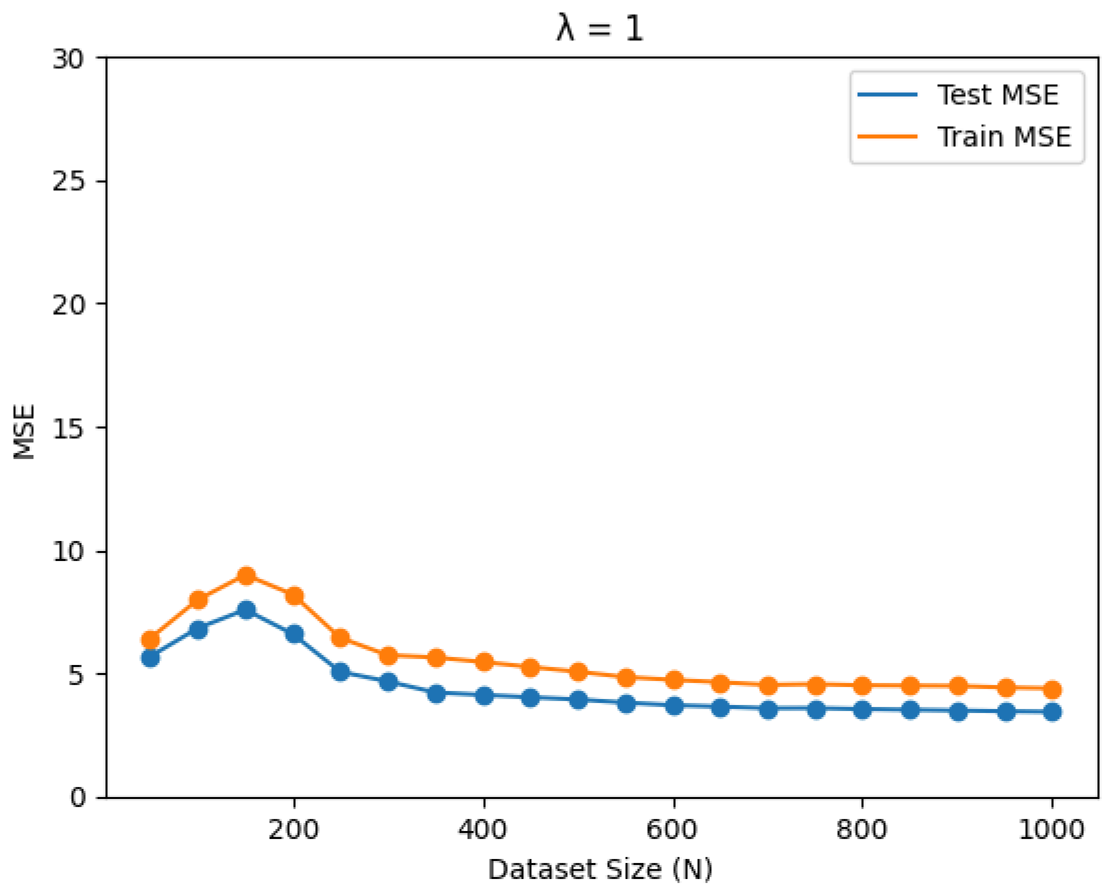
"train(150)1000_100", pre-cv: λ = 23, MSE = 4.848943

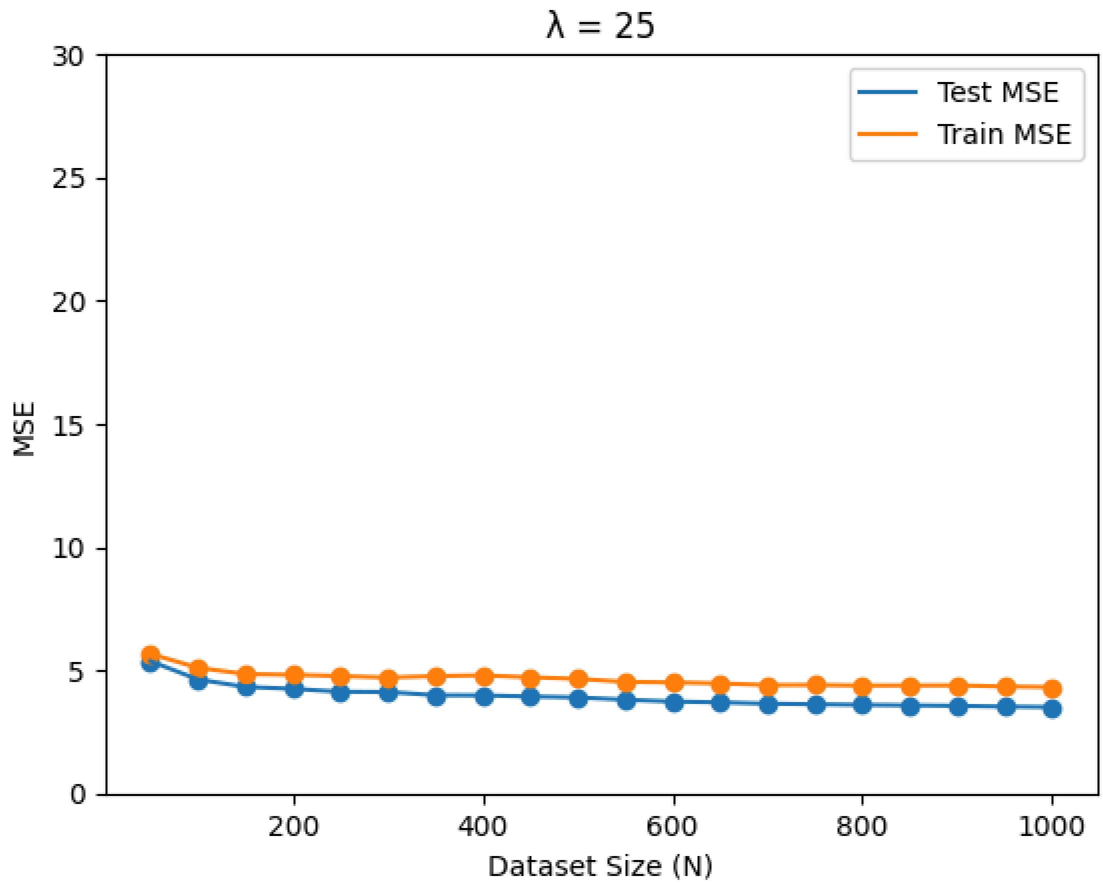
Post-cv: λ = 46, MSE =4.868577

2C. The drawbacks of cross validation are an increase in computational cost and computational time.

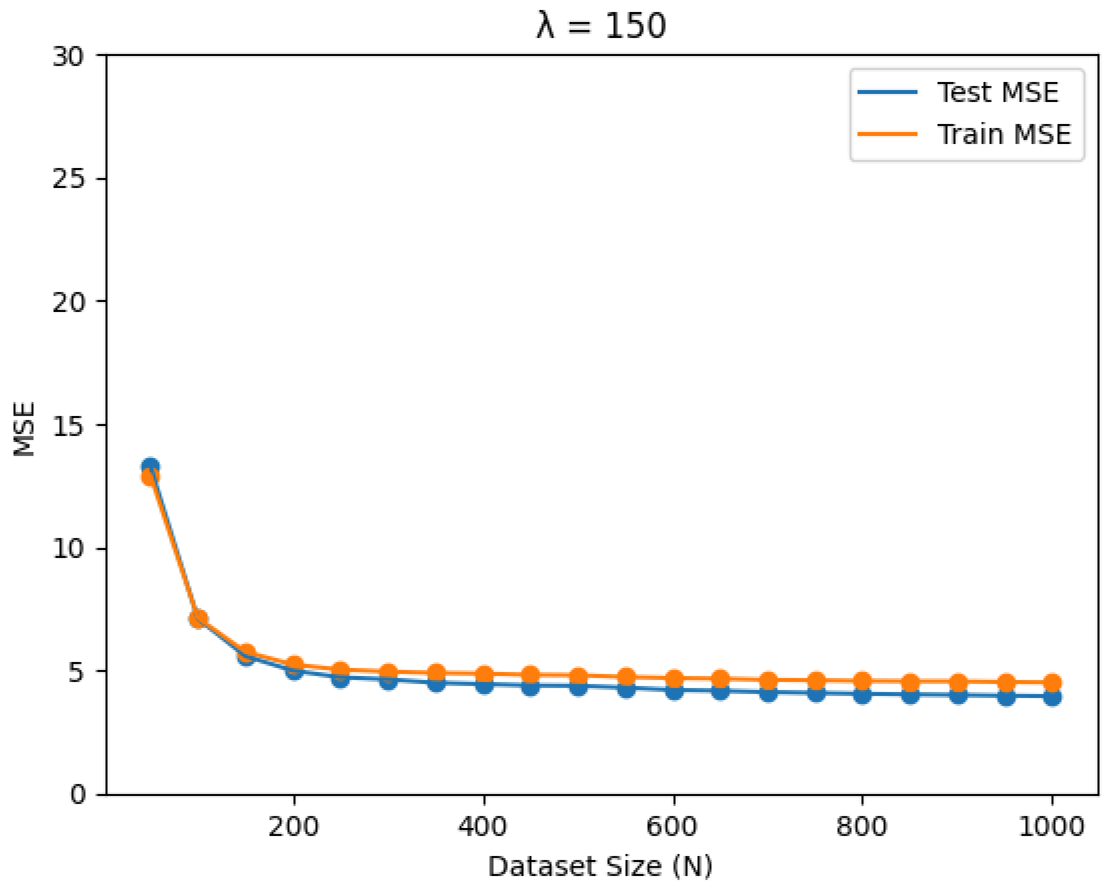
2D. The number of folds and number of data points within a fold. The more folds there are the more models there are to train and test. The points there are in a fold the more points you have to train/test a model on.

3. Learning Curve Graphs





λ = 25
Test MSE
Train MSE
MSE
Dataset Size (N)
0
5
10
15
20
25
30
200
400
600
800
1000


λ = 150
Test MSE
Train MSE
MSE
Dataset Size (N)
0
5
10
15
20
25
30
200
400
600
800
1000