2.1

5. result tables are:

Housing

| | 1 | 2 | 3 | 4 | 5 | 6 \ |
|---|---|---|---|---|---|---|
| train | 16.130689 | 16.248080 | 16.327717 | 16.170133 | 16.218028 | 16.206398 |
| test | 18.287521 | 16.676254 | 14.079052 | 16.718687 | 15.688264 | 16.245316 |

| | 7 | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|---|
| train | 16.125212 | 16.306939 | 16.370292 | 16.319315 | 16.242280 | 0.086329 |
| test | 17.582078 | 15.708135 | 15.314841 | 15.917910 | 16.221806 | 1.184452 |

Yacht

| | 1 | 2 | 3 | 4 | 5 | 6 \ |
|---|---|---|---|---|---|---|
| train | 10.654518 | 10.810268 | 10.871902 | 10.532604 | 10.388817 | 10.701055 |
| test | 11.181216 | 9.138587 | 7.354098 | 12.209803 | 14.263663 | 9.855132 |

| | 7 | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|---|
| train | 10.815432 | 10.436103 | 10.727657 | 10.679787 | 10.661814 | 0.162684 |
| test | 8.670470 | 13.830610 | 8.882607 | 10.639726 | 10.602591 | 2.277645 |

Concrete

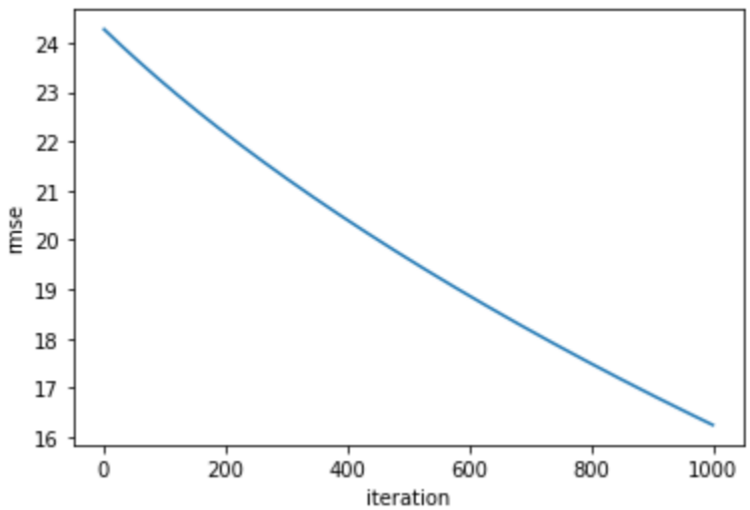
| | 1 | 2 | 3 | 4 | 5 | 6 \ |
|---|---|---|---|---|---|---|
| train | 21.508338 | 21.481594 | 21.690871 | 21.646483 | 21.581220 | 21.795581 |
| test | 23.764844 | 23.292755 | 21.477183 | 22.364117 | 21.651304 | 19.940856 |

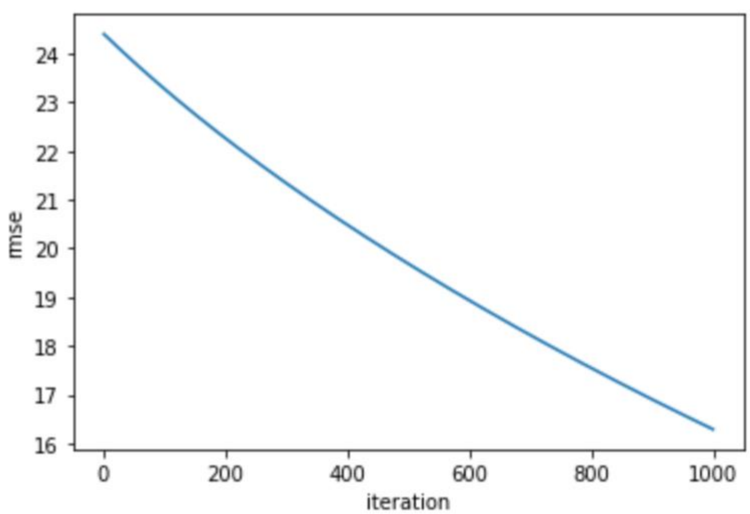
| | 7 | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|---|
| train | 21.702099 | 21.675394 | 21.717865 | 21.629788 | 21.642923 | 0.096557 |
| test | 20.527850 | 21.628093 | 20.014224 | 21.677305 | 21.633853 | 1.272173 |

6.

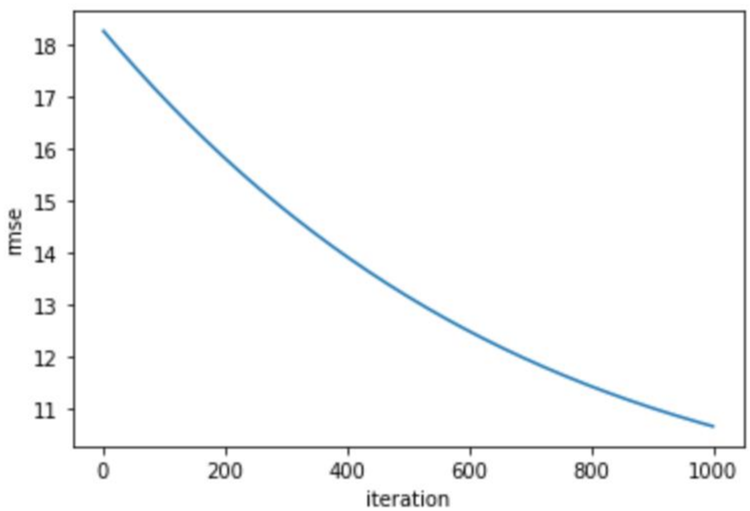
Housing

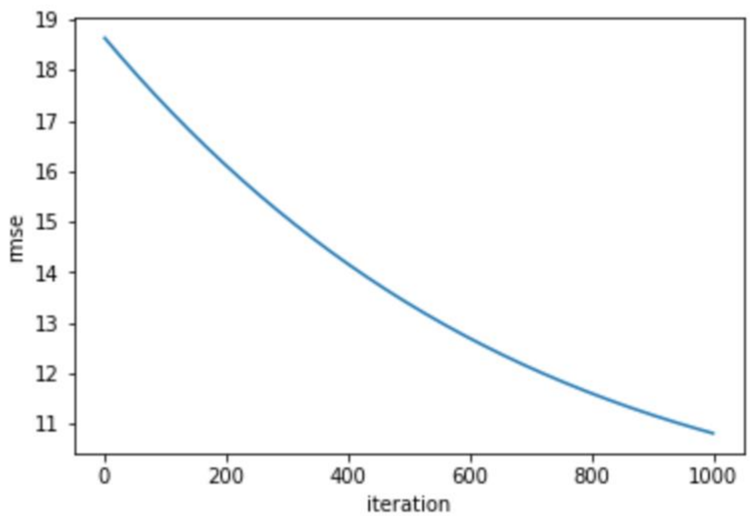
0.005943111485485980



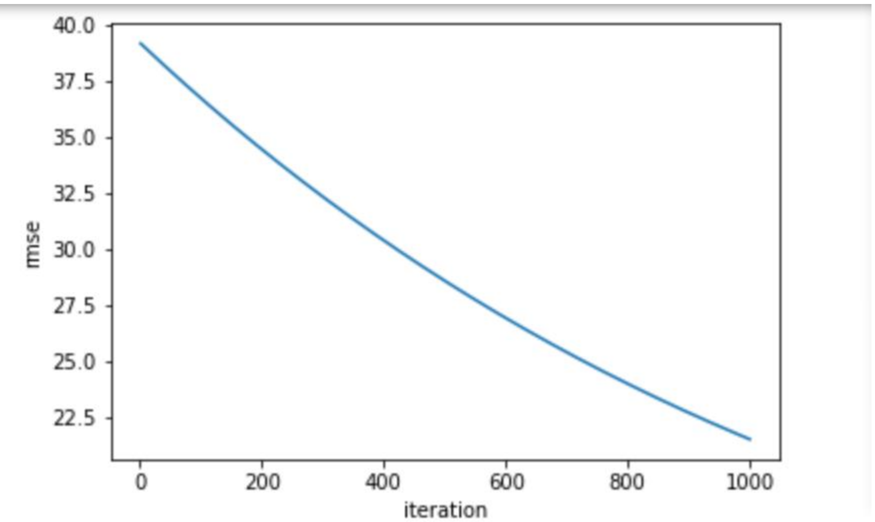



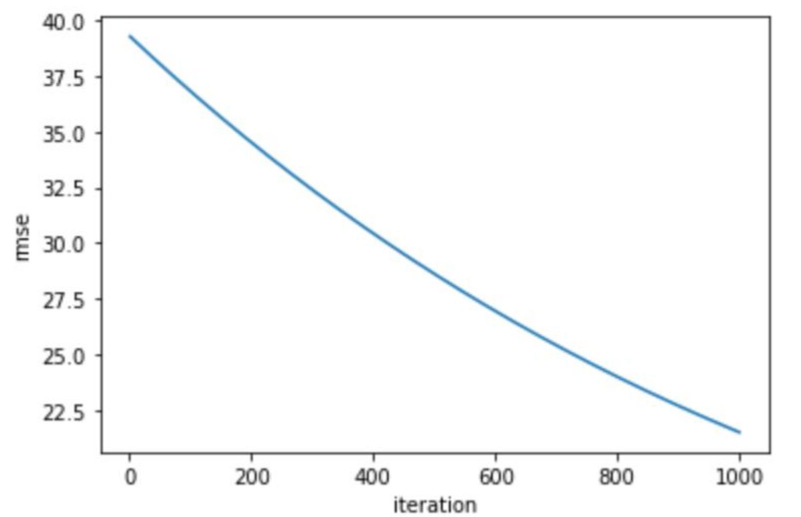

Yacht


mse
iteration
18
17
16
15
14
13
12
11
0
200
400
600
800
1000


mse
iteration
19
18
17
16
15
14
13
12
11
0
200
400
600
800
1000

Concrete




3.

Housing

use gradient descent

| | 1 | 2 | 3 | 4 | 5 | 6 \ |
|---|---|---|---|---|---|---|
| train | 16.130689 | 16.248080 | 16.327717 | 16.170133 | 16.218028 | 16.206398 |
| test | 18.287521 | 16.676254 | 14.079052 | 16.718687 | 15.688264 | 16.245316 |

| | 7 | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|---|
| train | 16.125212 | 16.306939 | 16.370292 | 16.319315 | 16.242280 | 0.086329 |

| | | | | | | |
|---|---|---|---|---|---|---|
| test | 17.582078 | 15.708135 | 15.314841 | 15.917910 | 16.221806 | 1.184452 |

use normal equations

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 \ |
|---|---|---|---|---|---|---|---|
| train | 4.510554 | 4.412320 | 4.798337 | 4.676975 | 4.685141 | 4.733414 | 4.665410 |
| test | 6.093142 | 6.781644 | 3.601935 | 4.798868 | 4.762457 | 4.291029 | 4.965709 |

| | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|
| train | 4.708447 | 4.769535 | 4.747251 | 4.670738 | 0.120167 |
| test | 4.456782 | 3.910719 | 4.123887 | 4.778617 | 0.982264 |

Yacht

use gradient descent

| | 1 | 2 | 3 | 4 | 5 | 6 \ |
|---|---|---|---|---|---|---|
| train | 10.759077 | 10.596337 | 10.300852 | 10.904817 | 10.900586 | 10.795771 |
| test | 9.009742 | 11.740571 | 14.868459 | 7.762121 | 6.650139 | 9.001725 |

| | 7 | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|---|
| train | 10.418659 | 10.825034 | 10.604874 | 10.535148 | 10.664115 | 0.206303 |
| test | 13.621566 | 9.034490 | 11.684818 | 11.789931 | 10.516356 | 2.632331 |

use normal equations

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 \ |
|---|---|---|---|---|---|---|---|
| train | 8.889671 | 8.859633 | 8.567945 | 9.057355 | 8.956526 | 8.938295 | 8.669535 |
| test | 8.629342 | 8.899324 | 11.295985 | 6.841947 | 8.090457 | 8.222574 | 10.519773 |

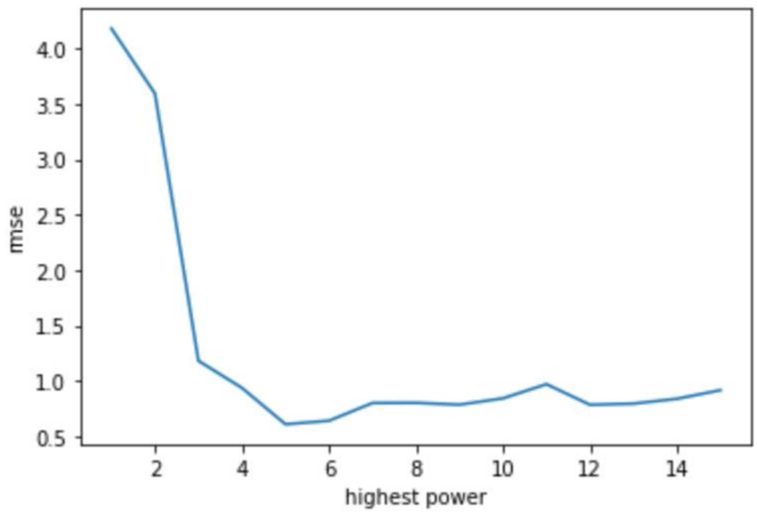
| | 8 | 9 | 10 | mean RMSE | std RMSE |
|---|---|---|---|---|---|
| train | 8.955624 | 8.836542 | 8.733864 | 8.846499 | 0.149306 |
| test | 7.935449 | 9.146025 | 9.968604 | 8.954948 | 1.328747 |

We can see that using normal equations has more accurate results than gradient descent. But it is probably because for this question we only continue our loop for 1000 times but as we increase the loop number, the result will get more accurate.

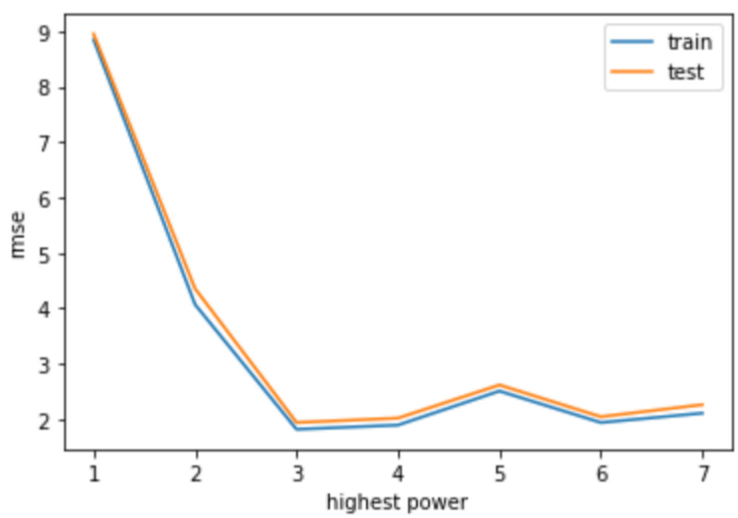
5.

5.1

Sinusoid



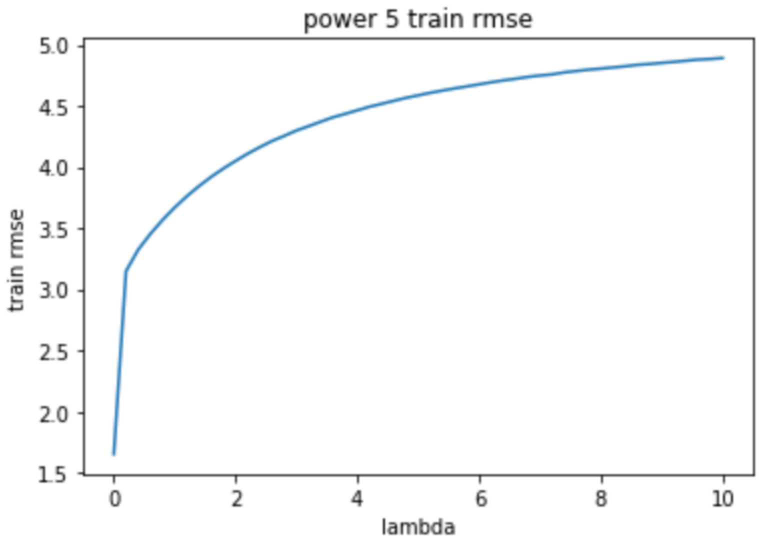

Yacht

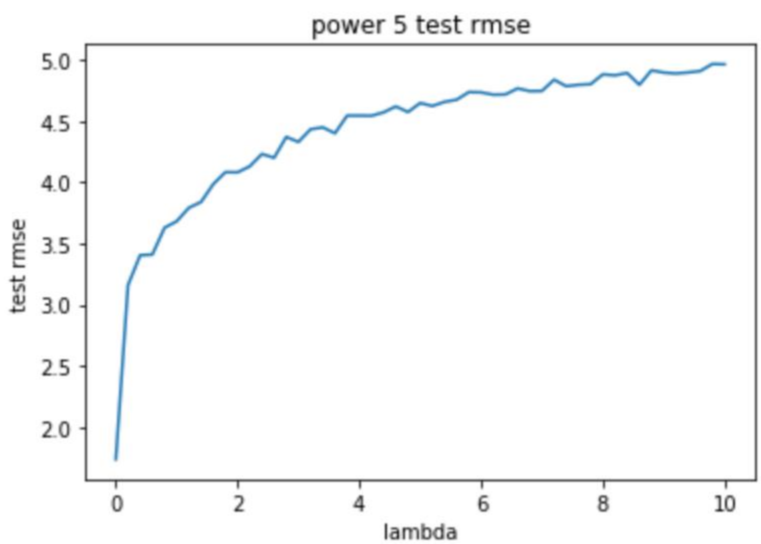


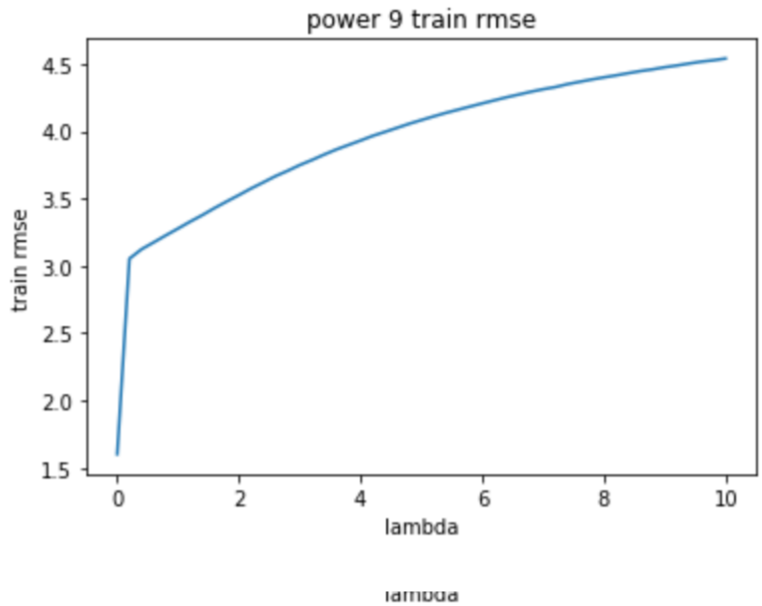

We can see that as we change the highest power, the fitting degree changes accordingly.

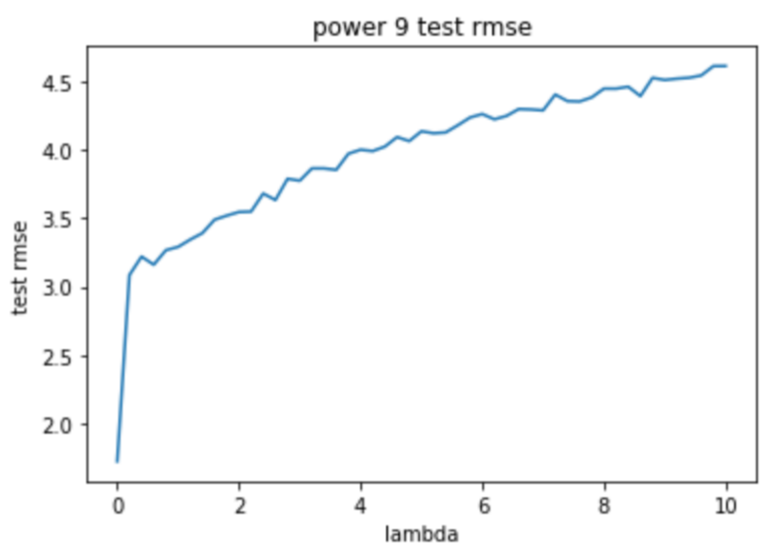
In most of the cases, validation set has higher rmse than train set because we use train to calculate theta vector and it can be understood that our theta fit train set better.

7.


power 5 train rmse
train rmse
lambda
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0
2
4
6
8
10


power 5 test rmse
test rmse
lambda
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
2
4
6
8
10




7.1

Interpretation:

It can be inferred from the graphs that overall, power 9 has more accurate estimations than power 5 because the highest rmse of power 5 reach near 5 but the highest rmse of power 9 only reach 4.5.

We can see other things too:

(1) Estimations fit train set more stably compared to test set.
(2) When lambda is 0, which means it is a liner regression, here comes the most accurate estimations. I think that means in this question we didn't produce overfitting issues so liner regression alone will make results accurate enough and by adding lambda item in the formula, we increase the final rmse.

(3) The value of lambda will significantly affect results. In this case, the final rmse increase as lambda increase.