**The description of the model and algorithm**

The number of features is 2: ;

and the degree of the equation can vary from 1 to 9.

Hence the general model can be written as

We define the loss function as –

To introduce regularization

We need to keep a constraint on the size of the coefficients which can be done in 2 ways –

Ridge regression –

Lasso regression –

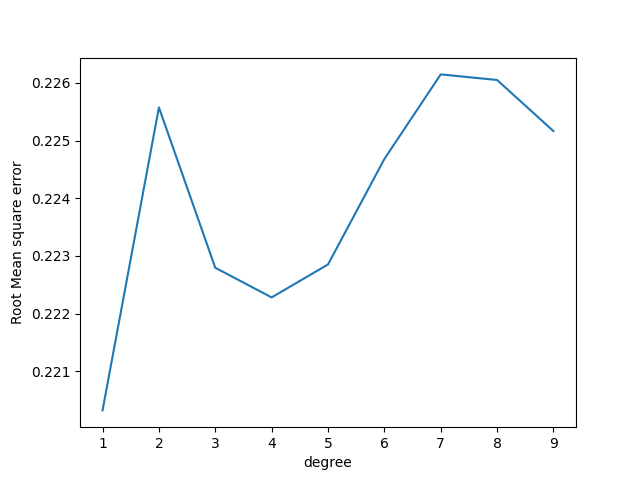
Stochastic gradient descent – we take just a single datapoint at a time to calculate the error function

Gradient descent – keep moving in the direction opposite to the gradient in every iteration until the error becomes low enough.

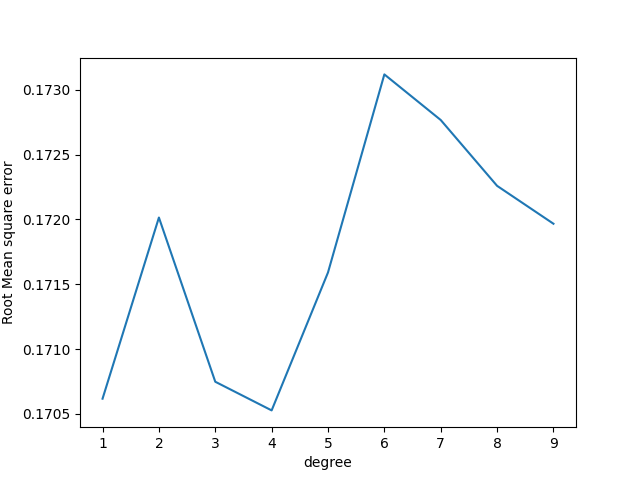
Minimum error achieved by training data and testing data without regression –

|  |  |  |
| --- | --- | --- |
| Degree | Training | Testing |
| 1 | 0.17061694130006516 | 0.22032448784492475 |
| 2 | 0.17201389750017146 | 0.22557449362133308 |
| 3 | 0.1707476134126093 | 0.22279621628262866 |
| 4 | 0.17052665029040592 | 0.2222822730585573 |
| 5 | 0.17159041748118278 | 0.22285310059051633 |
| 6 | 0.17311772753712315 | 0.2246804857404248 |
| 7 | 0.1727660014059834 | 0.22614663241836477 |
| 8 | 0.172258332634627 | 0.2260498371936418 |
| 9 | 0.1719661011733597 | 0.22516363302105152 |

Test error –



Train error –

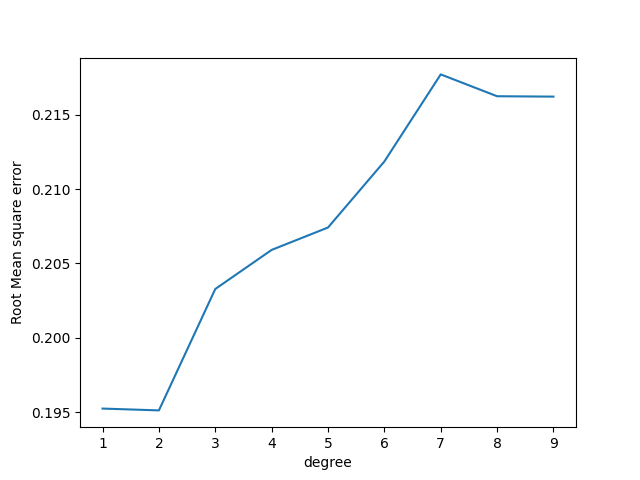
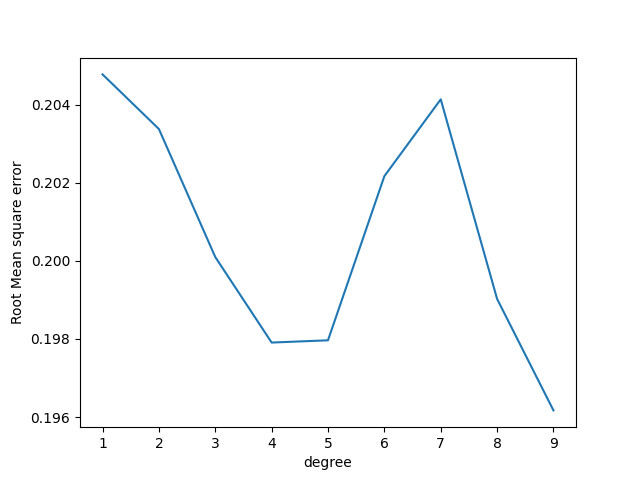


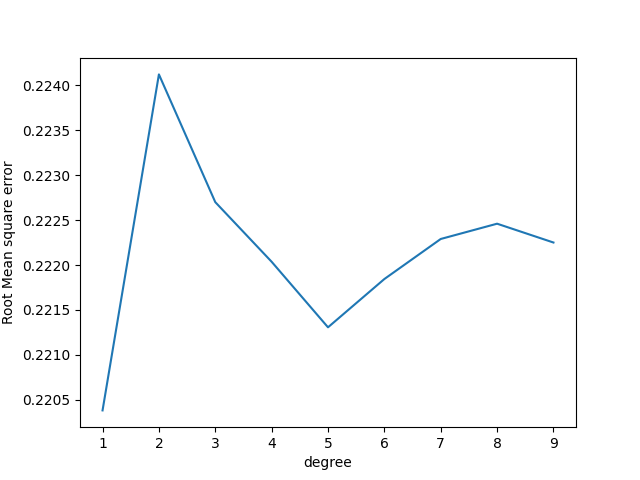
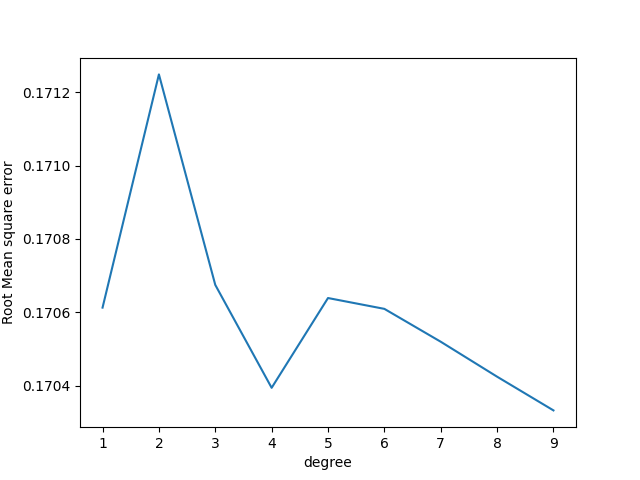
**Overfitting** – when the number of parameters is much less compared to the data points, the parameters become increasingly fine tuned to the noise in the data. This results in a very small error in the training data, but a very large error in the test data.

This can be overcome by regression.

The maximum number of parameters in our model is 55 when the degree is 9. The number of datapoints in our training set is around 900, thus the data is about 20 times larger than our parameters, so overfitting is not an issue for us.

However, if we decrease the training set to around 135, then it can be observed –



The training error decreases with increase in degree, but this is not the case with the test error.

