**ME698A**

**Assignment 4**

Name- Shivansh Pandey

Roll- 210988

Problem Statement:

The problem statement suggests that theta\_double\_dot is generally the function of theta, theta\_dot, sin(theta) and theta\_dot\_squared. Therefore, I found the correlation matrix between these entities and plotted the graphs to help propose a hypothesis.

1st hypothesis:

The correlation matrix showed a strong relation between the values of theta and sin(theta) with theta\_double\_dot. Essentially, the dependence of theta\_double\_dot on sin(theta) can be understood by reliance of sin(theta) on theta values. The correlation matrix showed significantly less correlation between the values of theta\_dot and theta\_dot\_squared with theta\_double\_dot.

2nd hypothesis:

The next hypothesis that I proposed was based on the maximum correlation of theta and sin(theta) with theta\_double\_dot.

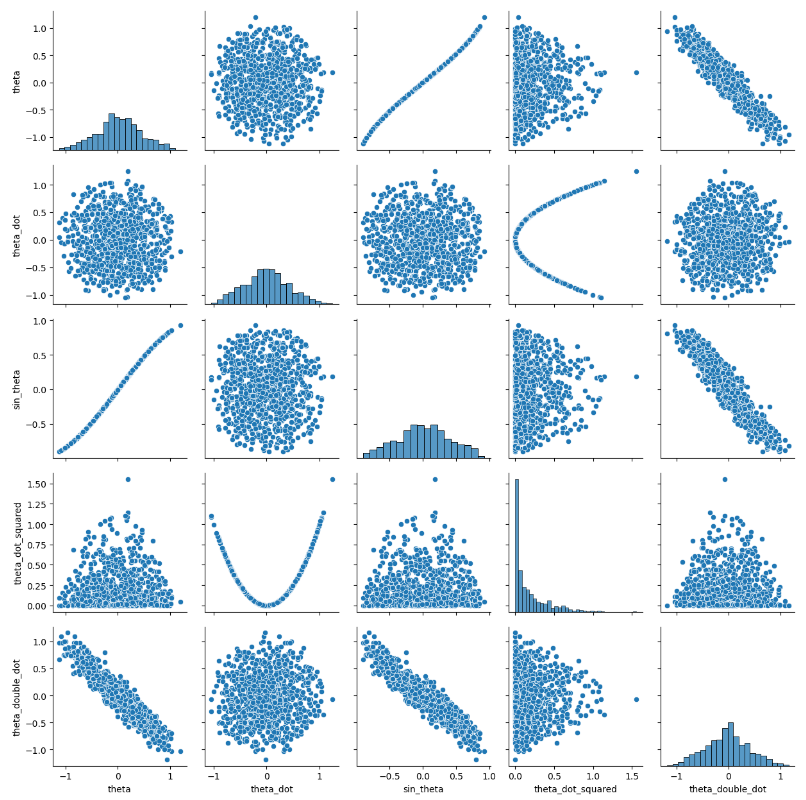
3rd hypothesis:

Theta values show maximum correlation with the values of theta\_double\_dot out of all quantities.

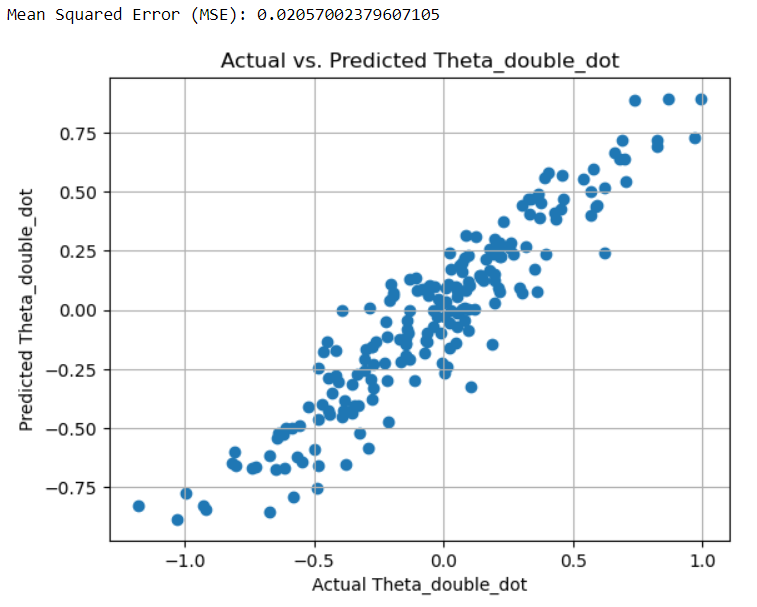
Next, I found the correlation between theta\_double\_dot and higher powers of theta. From the correlation matrix, I concluded that the values of theta\_double\_dot have a higher correlation with odd powers of theta. It showed very little correlation for even powers of theta. The correlation was approaching zero as I increased the power for both cases.

So the 4th hypothesis becomes:

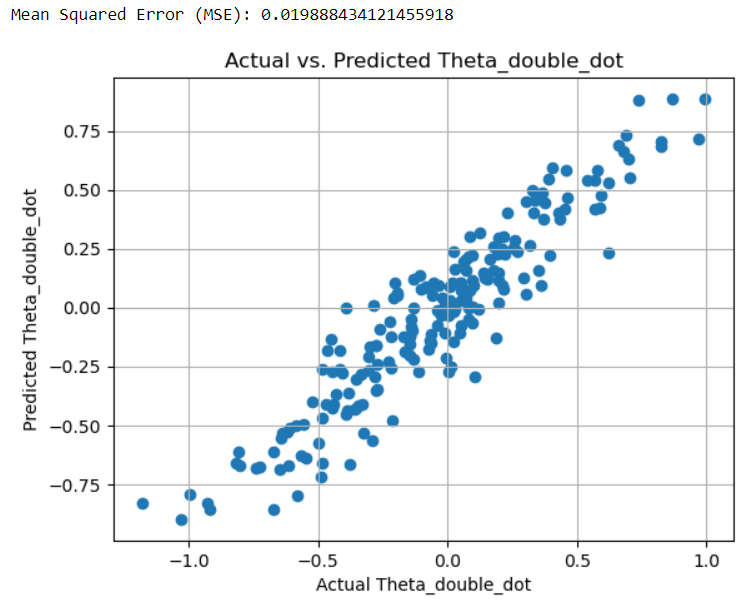
After proposing a certain hypothesis, we needed to implement ridge regularization along with the tuning of the regularization parameter using cross-validation. I have used 5-fold cross-validation. First I checked for discrete values of the regularization parameter from 10^-7 to 1000 to get an idea of where the cross-validation score is maximum. Then for hypertuning the parameter, I created an array using linspace from 0 to 10 as in all cases best values came around 1 and 0.1. Then I trained the model using the best value of the regularization parameter and obtained the mean squared error (MSE) using the test data. I finally made a scatter plot between actual values of theta and predicted values of theta. And mentioned the MSE for each plot with the image below.

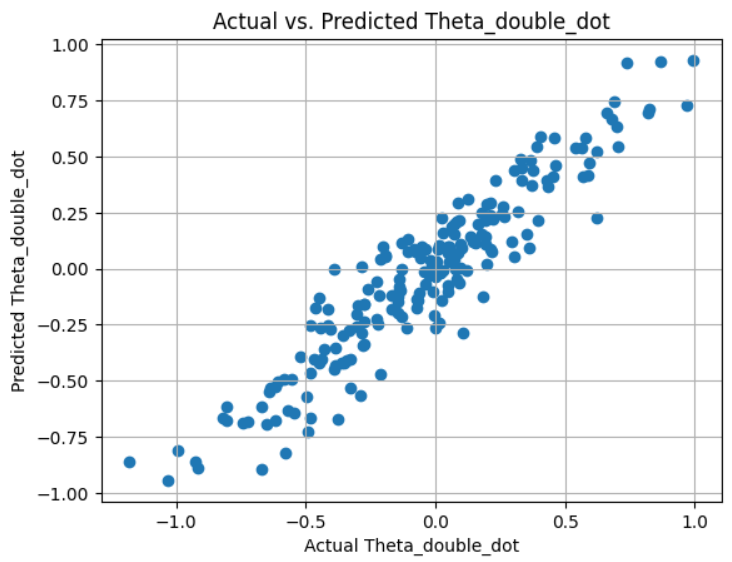


**Plot of correlations between all the given parameters including theta\_double\_dot**





****

****

****

**Conclusion:**

**By the hypotheses we made, we can conclude that the governing equations for our simple pendulum are much more dependent on theta and sin(theta). Among the three hypotheses made, the governing equation which best fits the experimental data(minimum mean square error) is:**

****

**General method:**

**A general way would have been considering theta\_double\_dot as a function of polynomial of n order of theta + polynomial of m order of theta\_dot.**

**We won’t take polynomial of kth derivative of theta where k>=2 as theta\_double\_dot is the label. N and m would be taken by seeing the trend of error between some large n and n+1, similarly for m.**

**This would cover both sin and cos and other functions by considering taylor series.**

**I tried to implement it and saw the trend of n and coefficient of terms of greater order came very less, nearly zero, indicating less dependency on them. Also, terms with odd orders had greater coefficients indicating dependency on sin\_theta and verifying our above hypothesis . But due to time constraints I was not able to code it fully and show in my code. Nevertheless, it would be fun to try it out without any time constraints.**

**Thank You, it was fun doing this.**