Name: Krutarth Deshmukh Roll No.: 210535

**ME698A: Home Assignment-4  
Report**

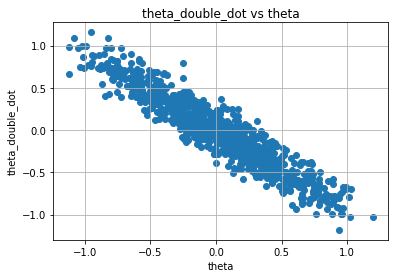
**Problem Statement**  
A data file contains three columns: theta (angular displacement), theta\_dot (angular velocity), and theta\_double\_dot (angular acceleration). A row of the data file indicates the angular displacement (θ), angular velocity (), and angular acceleration () of the pendulum at a given time instant. The goal is to discover the governing equation of the pendulum. In general, could be a function of θ, sinθ, and 2.

**Solution Procedure**

1. Qualitative analysis:

sinθ and 2 are calculated at each data point and added to the data file.

is plotted against θ, sinθ, and 2 separately. At first glance, the relation with θ and sinθ appears linear. There does not seem to be an obvious relation with and 2.

** A blue dot pattern on a white grid

Description automatically generated**

**A screen shot of a blue dot

Description automatically generated A screen shot of a graph

Description automatically generated**

A correlation matrix is used to evaluate the relations between the different variables in the data set. If the correlation coefficient between two variables is 1, it means the relationship between them is strong; 0 implies a neutral relationship and -1 implies a negative relationship. For better visualisation, a heatmap of the correlation matrix is generated using the ‘seaborn’ library in python.



From the correlation matrix, it is clear that has a strong negative relation with θ as well as sinθ and an almost neutral relation with and 2. Hence, it is reasonable to assume that is independent of and 2. The scatter plots of vs θ and vs sinθ plotted earlier hint towards a linear relation.

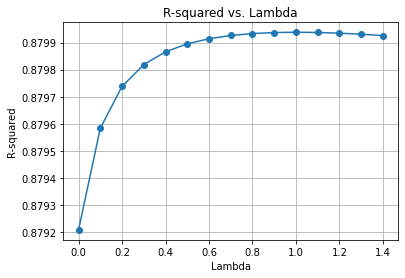
∴ Hypothesis:  **= w0 + w1θ + w2sinθ**

1. Linear regression with ridge regularization:

Ridge regularisation is implemented to control overfitting, otherwise, the error will be high with testing data. 80% of the original data set is used for training and the rest 20% for testing. Regression is carried out for λ ranging from 0 to 1.4 with a step of 0.1. R2 and coefficients of the hypothesis are calculated for each case.

1. Cross validation to find optimum λ:

R2 vs λ is plotted to find out the optimum value of λ for which R2 is maximum, i.e., the regression model is most accurate.



**Results**Optimum lambda = 1.0R-squared = 0.8799384868175187w0 = -0.008836076133253708w1 = -0.533341060956239w2 = -0.4065148867027883

**Conclusion:**  = -0.01 – 0.53θ – 0.41sinθ