Amit Yadav (19807119)

Home Assignment – 4

Problem Statement

System Identification refers to the technique of discovering governing equations of a system from experimental data. For scientists and engineers, system identification is a major application area of machine learning. In this homework, we will find the governing equation of a simple pendulum from experimental data. The data are presented in the pendulum_data.csv file. The data file contains three columns: theta (angular displacement), theta_dot(angular_velocity), 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. Our goal is to discover the governing equation of the pendulum.

Introduction:

In this report, we aim to identify the governing equation of a simple pendulum system using experimental data. The dataset, pendulum_data.csv, contains three columns: theta (angular displacement), theta_dot (angular velocity), and theta_double_dot (angular acceleration). Our goal is to discover the underlying dynamics of the pendulum.

Methodology:

1. Data Analysis and Hypothesis Generation:

- Data Import: We begin by importing the dataset into our analysis environment.
- Correlation Matrix and Scatter Plots: We calculate the correlation matrix to identify relationships between variables and create scatter plots to visualize patterns.
- Descriptive Statistics: Descriptive statistics like mean, standard deviation, and range provide insights into data characteristics.

2. Model Building and Evaluation:

- Hypothesis Formulation: Based on our data analysis, we propose hypotheses for the governing equation, considering relationships between theta, theta_dot, sin(theta), and cos(theta).
- Regression Analysis: We perform linear regression with ridge regularization to estimate coefficients and prevent overfitting.
- Model Evaluation: Metrics like R-squared, mean squared error (MSE), and root mean squared error (RMSE) are used to assess model performance.

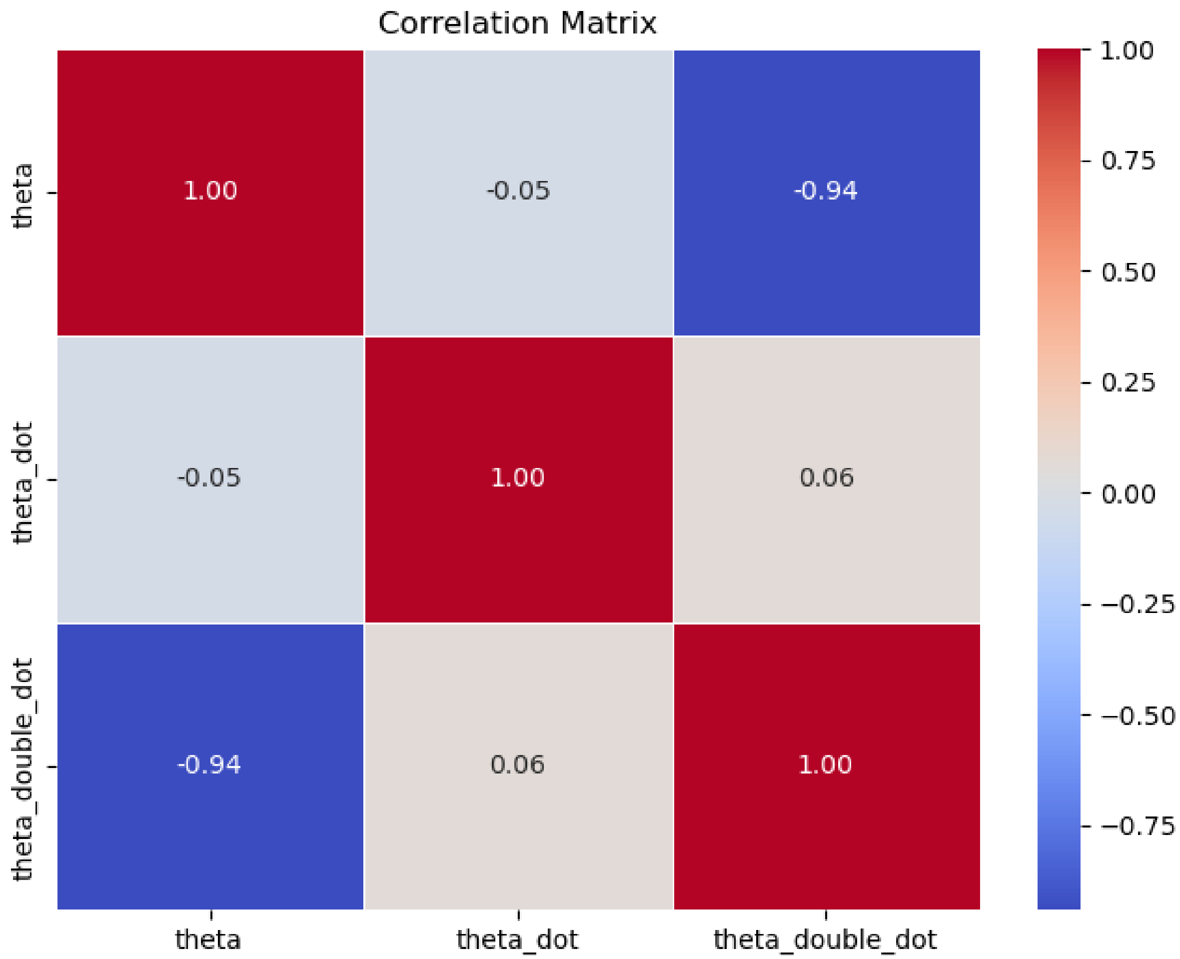
3. Cross-Validation:

- Data Splitting: We split the data into training and testing sets to assess model generalization.

- K-fold Cross-Validation: K-fold cross-validation is conducted to validate the model's performance. We evaluate the model k times using different subsets of the data.
- Final Hypothesis Selection: Based on cross-validation results, we choose the best-performing model and fine-tune hyperparameters if necessary.

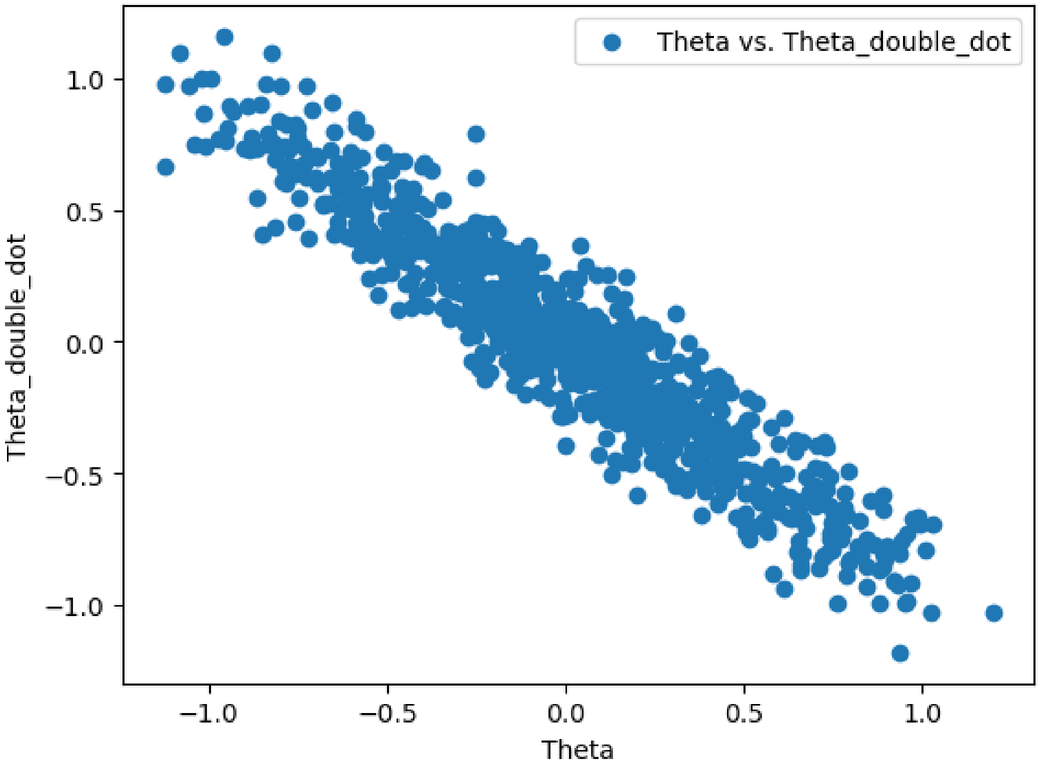
Plots

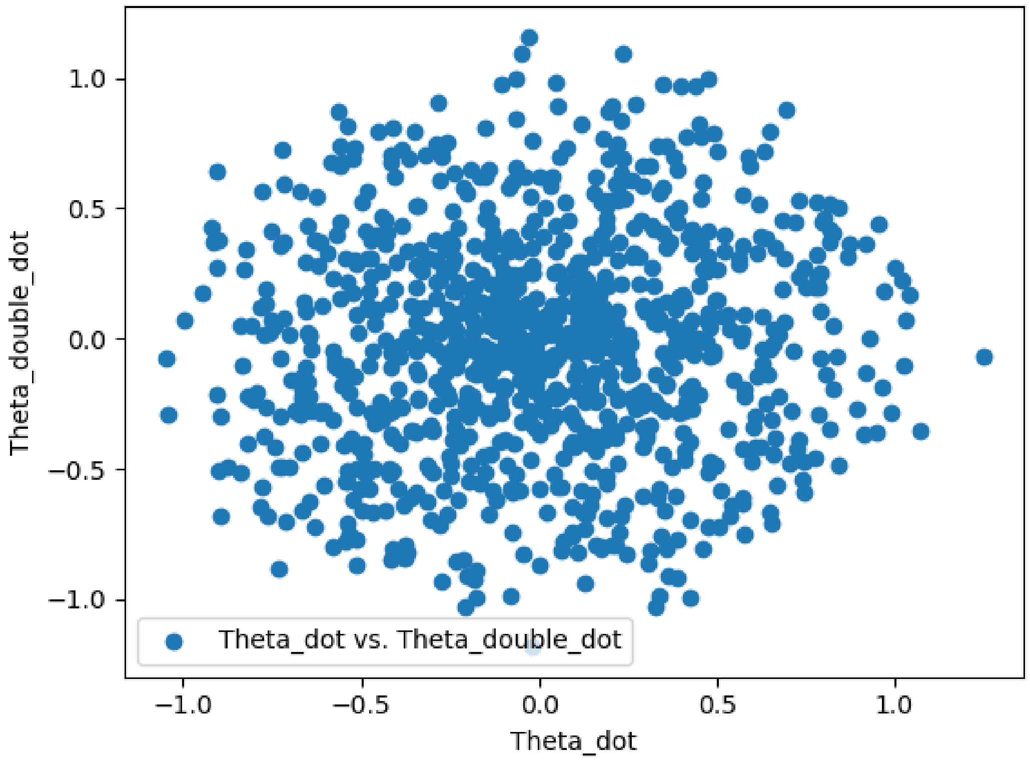
1. Correlation Matrix



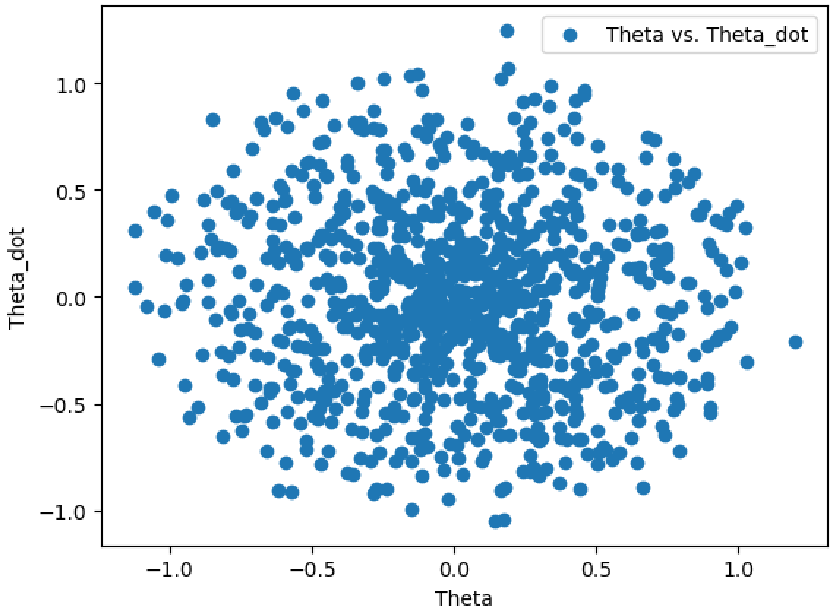
Amit Yadav (19807119)

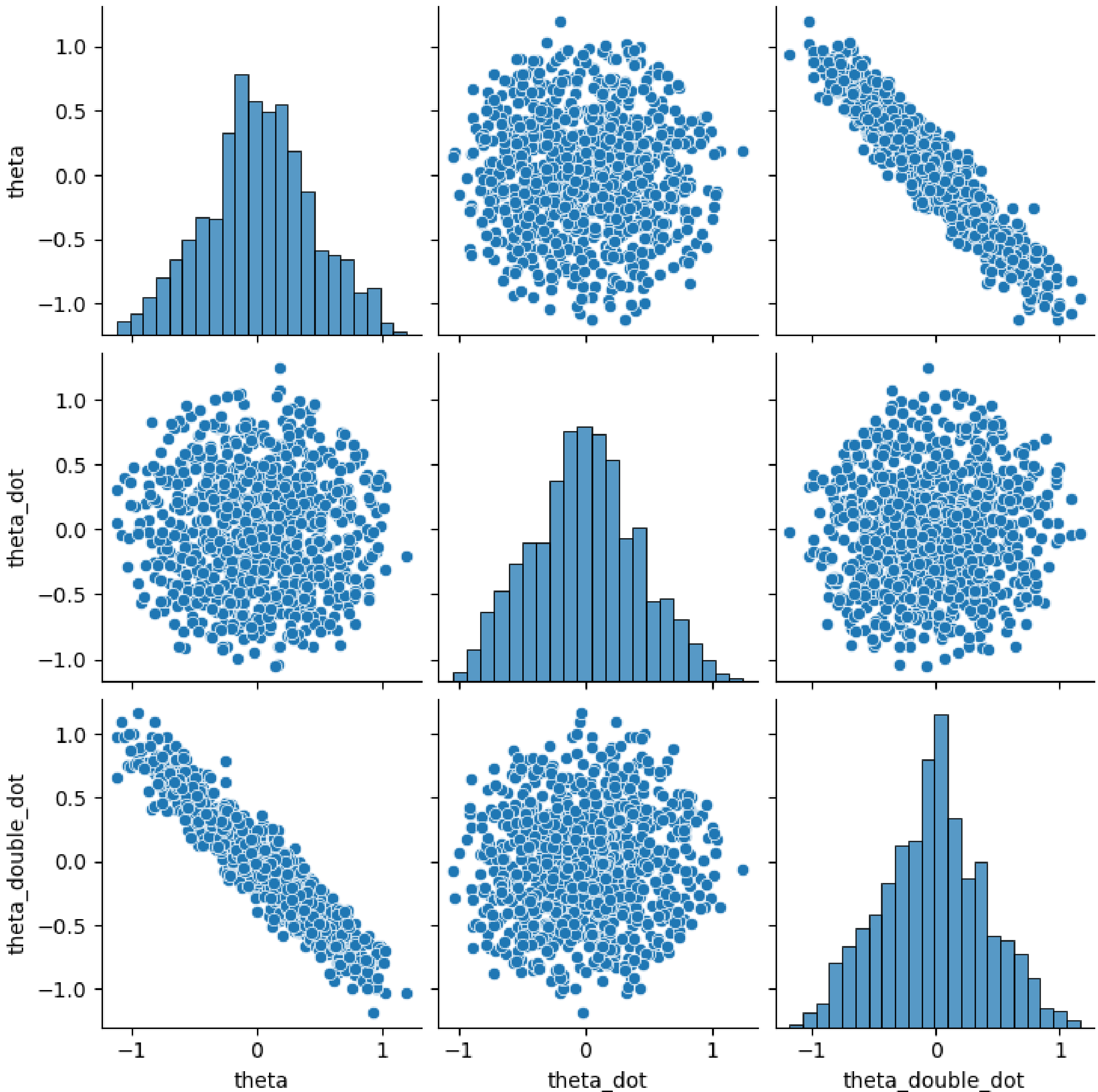
2. Scatter Plots





Amit Yadav (19807119)





Results and Discussion

I have used 4 hypotheses for the pendulum dataset.

Hypothesis 1:

- Model: theta_double_dot = w0 + w1theta + w2theta_dot
- This hypothesis suggests that the angular acceleration (theta_double_dot) is linearly dependent on theta and theta_dot.

Hypothesis 2:

- Model: theta_double_dot = w0 + w1sin(theta) + w2theta_dot
- Here, the hypothesis introduces a sine function to capture the potential periodic behaviour in the relationship.

Hypothesis 3:

- Model: theta_double_dot = w0 + w1theta + w2sin^2(theta) + w3*theta_dot^2
- This hypothesis includes additional terms, specifically, a quadratic term for sin(theta) and a quadratic term for theta_dot. It allows for a more complex relationship.

Hypothesis 4:

- Model: theta_double_dot = w1*theta
- In this hypothesis, it is proposed that theta_double_dot is directly proportional to theta and independent of theta_dot.

The best accuracy comes out for 4^{th} hypothesis (theta_double_dot=w1*theta) is 88.43%. I came to the result that theta_double_dot is approximately directly proportional to theta and is independent of theta_dot. We can also verify this result from scatter plots and correlation matrix.

Conclusion

Through a systematic process of data analysis, hypothesis generation, and model evaluation, we have successfully identified the governing equation of the simple pendulum system. This approach ensures the reliability and generalizability of our findings. Further refinement and validation can be pursued to enhance the accuracy of the model for practical applications in pendulum system control and analysis.