

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 ($\dot{\theta}$), and angular acceleration ($\ddot{\theta}$) of the pendulum at a given time instant.

Our goal is to discover the governing equation of the pendulum. In general, $\ddot{\theta}$ could be a function of $\theta, \sin \theta, \dot{\theta}, \dot{\theta}^2$

- (a) Calculate correlation matrix, scatter plots and any other related metrics to qualitatively propose possible hypotheses and create the hypotheses space.
- (b) Using linear or nonlinear regression with ridge regularization find appropriate parameters.
- (c) Use cross validation to finalize your hypothesis.

Submission:

1. Report (maximum 5 pages): .docx (or .tex) file that includes (a) problem statement and solution procedure (~ 1 page), (b) plots, (d) results and discussion, (e) conclusion
2. Computer program