

Solution Algorithm

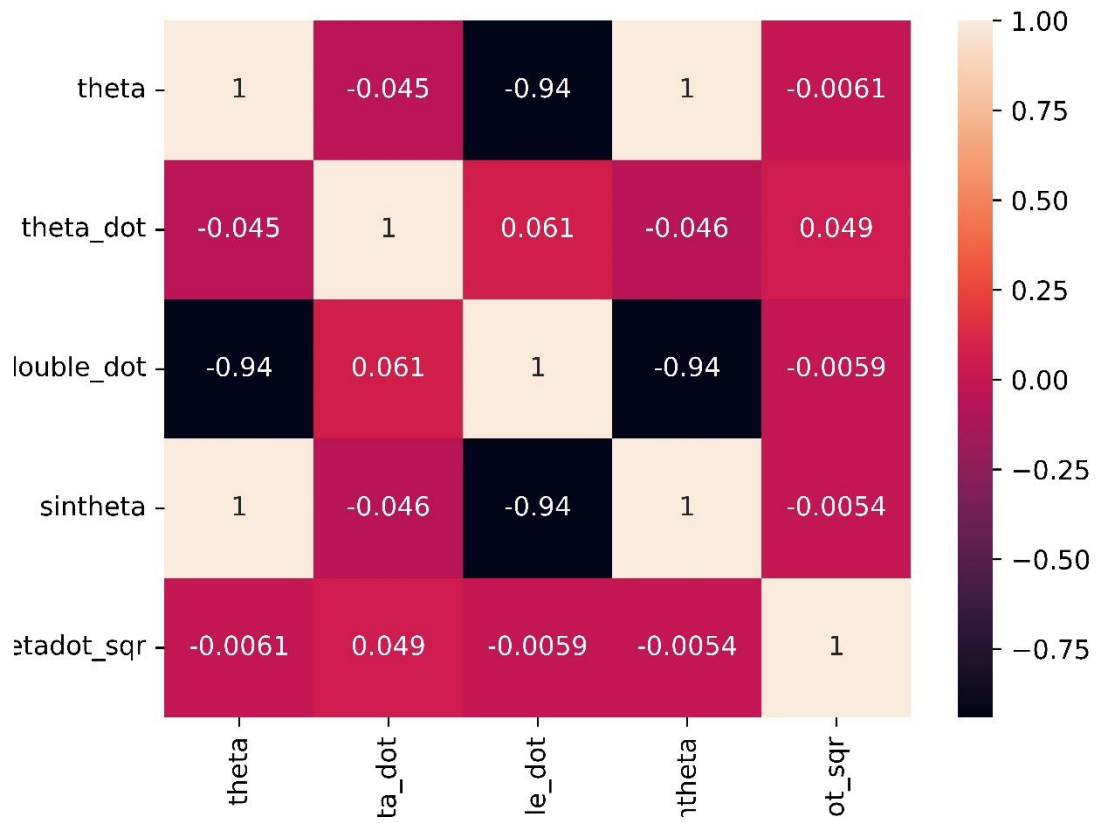
- 1) Correlation matrix was obtained for the given data set and heatmap was plotted.
- 2) Scatter plots for (θ) , $(\dot{\theta})$, $\sin(\theta)$ and $\dot{\theta}^2$ with respect to $(\ddot{\theta})$ were plotted.
- 3) Data was sorted according to their Correlation score with respect to $(\ddot{\theta})$. So our pandas data frame was of the form.

	theta_double_dot	theta	sintheta	theta_dot	thetadot_sqr
0	0.029322	0.040431	0.040420	-0.069829	0.004876
1	0.138360	-0.151797	-0.151214	-0.662335	0.438688
2	0.446742	-0.584669	-0.551923	0.797156	0.635458
3	0.016372	-0.151016	-0.150442	-0.772319	0.596477
4	-0.298235	0.256469	0.253667	0.188715	0.035613
...
995	0.369823	-0.406553	-0.395446	0.367425	0.135001
996	-0.048264	0.058292	0.058259	-0.137930	0.019025
997	-0.021672	-0.006571	-0.006571	0.127600	0.016282
998	-0.302841	0.277242	0.273704	0.196108	0.038458
999	-0.391495	0.704613	0.647739	0.732455	0.536491

- 4) Then our hypothesis space was formed taking 4 hypotheses containing
 - Hypothesis₁: (θ) , $\sin(\theta)$, $(\dot{\theta})$, and $\dot{\theta}^2$
 - Hypothesis₂: (θ) , $\sin(\theta)$ and $(\dot{\theta})$
 - Hypothesis₃: (θ) , $\sin(\theta)$
 - Hypothesis₄: (θ)
- 5) For each of the hypotheses, found split data for training and testing which was then iteratively processed in Linear Ridge regression model for 2000 different $\lambda \in [0.001, 10]$.
- 6) λ with minimum mean square error was used for 4-cross validation.
- 7) Mean square for each of the fold was then averaged out to give error for the hypotheses.
- 8) Mean square errors for each hypothesis were compared and the one with minimum error can be chosen for fitting the data.

a. Plots

Correlation matrix



b. Results and Discussion

Lambda for different hypothesis

$$\lambda_1 = 3.3023206603301656$$

$$\lambda_2 = 2.902160580290145$$

$$\lambda_3 = 2.7020805402701353$$

$$\lambda_4 = 1.2515002501250625$$

Error Values for different Hypothesis are

$$e_1 = 0.019832048733834685$$

$$e_2 = 0.019834610116315925$$

$$e_3 = 0.01985435369785135$$

$$e_4 = 0.01989220127440026$$

c. Conclusions

Least error was found in Hypothesis₁ = **0.019832048733834685**

Therefore, Hypothesis one with parameters (θ) , $\sin(\theta)$, $(\dot{\theta})$, and $(\dot{\theta}^2)$ would give the best linear relation

Governing equation will be given by

$$\ddot{\theta} = w_0 + w_1 * \theta + w_2 * \dot{\theta} + w_3 * \dot{\theta}^2 + w_4 * \sin(\theta) + w_5 * \lambda$$