ME644

Home Assignment 5: System Identification for Simple Pendulum

Priyank Agnihotri [230797]

Problem Statement

To discover the governing equation of a simple pendulum from experimental data using machine learning techniques. The dataset contains 1000 observations with three variables:

- **θ** (theta): Angular displacement in radians
- **$\dot{\theta}$** (theta_dot): Angular velocity in radians/second
- **$\ddot{\theta}$** (theta_double_dot): Angular acceleration in radians/second^2

Solution Procedure

The systematic approach involves:

1. **Exploratory Data Analysis (EDA):**
    - Calculate correlation matrix to identify key relationships
    - Generate scatter plots to visualize variable interactions
    - Analyze data distributions and quality
2. **Hypothesis Space Development:**
    - Propose features based on pendulum physics
    - Justify relevant features ($\sin\theta$, $\dot{\theta}^2$) selection using correlation analysis
3. **Model Fitting:**
    - Apply ridge regression with regularization
    - Use cross-validation for hyperparameter selection
    - Standardize features for optimal performance
4. **Model Validation:**
    - Perform k-fold cross-validation (k = 3, 5, 10)
    - Evaluate multiple performance metrics
    - Analyze residuals and prediction accuracy

Exploratory Data Analysis

Correlation Matrix Analysis

The correlation matrix reveals strong relationships between variables:

Variable	θ	$\dot{\theta}$	$\ddot{\theta}$	$\sin\theta$	$\dot{\theta}^2$
θ	1.000	-0.045	**-0.941**	0.999	-0.006
$\dot{\theta}$	-0.045	1.000	0.061	-0.046	0.049
$\ddot{\theta}$	**-0.941**	0.061	1.000	**-0.941**	-0.006
$\sin\theta$	0.999	-0.046	**-0.941**	1.000	-0.005
$\dot{\theta}^2$	-0.006	0.049	-0.006	-0.005	1.000

Key Insights

- **Strong negative correlation** between θ and $\ddot{\theta}$ (-0.941) indicates primary restoring force
- **Strong negative correlation** between $\sin\theta$ and $\ddot{\theta}$ (-0.941) confirms nonlinear pendulum dynamics
- **Weak correlations** for velocity terms suggest minimal damping effects
- High correlation between θ and $\sin\theta$ (0.999) expected for small angle oscillations

Scatter Plot Analysis

Visual analysis reveals:

- **θ vs $\ddot{\theta}$:** Clear negative linear relationship with minimal scatter
- **sin θ vs $\ddot{\theta}$:** Similar strong negative relationship to θ vs $\ddot{\theta}$
- **$\dot{\theta}$ vs $\ddot{\theta}$:** Weak positive relationship with significant scatter
- **$\dot{\theta}^2$ vs $\ddot{\theta}$:** Very weak relationship, mostly noise
- **θ distribution:** Approximately normal, indicating balanced sampling

3. Hypothesis Space Development

Proposed Governing Equation

Based on pendulum physics and correlation analysis, we propose:

$$\ddot{\theta} = a_1 \cdot \theta + a_2 \cdot \sin(\theta) + a_3 \cdot \dot{\theta} + a_4 \cdot \dot{\theta}^2 + b$$

Feature Justification

1. **θ term (Linear restoring force):**
   - Small angle approximation: $\sin\theta \approx \theta$
   - Strong correlation (-0.941) supports inclusion
   - Represents Hooke's law analogy
2. **sin θ term (Nonlinear restoring force):**
   - Exact pendulum dynamics: $\ddot{\theta} = -(g/L)\sin\theta$
   - Strong correlation (-0.941) confirms importance
   - Captures large angle behavior
3. **$\dot{\theta}$ term (Linear damping):**
   - Air resistance and friction effects
   - Weak correlation (0.061) suggests minor contribution
   - Standard in damped oscillator models
4. **$\dot{\theta}^2$ term (Nonlinear damping):**
   - Velocity-dependent drag forces
   - Very weak correlation (-0.006) indicates minimal effect
   - Included for completeness
5. **Constant term b:**
   - Bias/offset correction
   - Accounts for systematic measurement errors

4. Ridge Regression Model Fitting

Model Configuration

- **Features:** θ, sin θ, $\dot{\theta}$, $\dot{\theta}^2$
- **Target:** $\ddot{\theta}$
- **Regularization:** Ridge regression with α = 1.592283
- **Feature scaling:** StandardScaler normalization
- **Cross-validation:** 5-fold for hyperparameter selection

Hyperparameter Selection

Ridge regression alpha parameter optimized using cross-validation:

- **Search range:** 10^{-4} to 10^4 (100 logarithmically spaced values)
- **Optimal α:** 1.592283
- **CV score (MSE):** 0.019875

Model Coefficients (Original Scale)

Feature	Coefficient	Physical Interpretation
θ	-0.5298	Strong negative restoring force
sin θ	-0.4149	Nonlinear pendulum dynamics
$\dot{\theta}$	0.0187	Minimal positive damping
$\dot{\theta}^2$	-0.0225	Weak nonlinear damping
Intercept	-0.0024	Near-zero bias (good!)

5. Model Validation and Cross-Validation

Cross-Validation Results

CV Folds	MSE (Mean ± Std)	RMSE	R^2 (Mean ± Std)
3-fold	0.019802 ± 0.000572	0.140721	0.885059 ± 0.003966
5-fold	0.019878 ± 0.001532	0.140991	0.884554 ± 0.005753
10-fold	0.019885 ± 0.002925	0.141013	0.883635 ± 0.018357

Model Performance Summary

- **Cross-validated MSE:** 0.019878
- **Cross-validated RMSE:** 0.140991
- **Cross-validated R^2:** 0.884554
- **Training R^2:** 0.886186

Performance Analysis

- **Excellent fit:** $R^2 > 0.88$ indicates model explains 88.4% of variance
- **Consistent results:** Similar performance across different CV folds
- **Low overfitting:** Training and CV scores are very close
- **Optimal regularization:** Alpha = 1.592 provides best bias-variance tradeoff

6. Results and Discussion

Discovered Governing Equation

$$\ddot{\theta} = -0.5298\cdot\theta - 0.4149\cdot\sin(\theta) + 0.0187\cdot\dot{\theta} - 0.0225\cdot\dot{\theta}^2 - 0.0024$$

Physical Interpretation

1. **Dominant restoring forces:**
   - Linear term ($-0.5298\cdot\theta$): Strong restoring force proportional to displacement
   - Nonlinear term ($-0.4149\cdot\sin\theta$): Exact pendulum dynamics for large angles
   - Both coefficients negative, confirming restoring nature
2. **Damping effects:**
   - Linear damping ($0.0187\cdot\dot{\theta}$): Small positive coefficient suggests minimal air resistance
   - Nonlinear damping ($-0.0225\cdot\dot{\theta}^2$): Weak velocity-dependent effects
   - Overall damping is minimal, indicating low-friction system
3. **Bias term (-0.0024):**
   - Near-zero intercept indicates no systematic measurement offset
   - Good experimental setup with minimal bias

Comparison with Theoretical Pendulum

Theoretical equation: $\ddot{\theta} = -(g/L)\sin\theta - c_1\dot{\theta} - c_2\dot{\theta}^2$

Our discovered equation: $\ddot{\theta} = -0.5298\cdot\theta - 0.4149\cdot\sin\theta + 0.0187\cdot\dot{\theta} - 0.0225\cdot\dot{\theta}^2$

Key observations:

- Model captures both linear (small angle) and nonlinear pendulum behavior
- Negative coefficients for θ and sin θ confirm restoring forces
- Minimal damping terms indicate well-designed experimental setup
- Combined linear and nonlinear terms provide comprehensive description

Model Validation Assessment

1. **Residual analysis:** Random distribution around zero with minimal patterns
2. **Prediction accuracy:** Strong linear relationship between actual and predicted values
3. **Cross-validation stability:** Consistent performance across different fold numbers
4. **Feature importance:** θ and sin θ dominate, velocity terms are secondary

7. Conclusion

Summary of Findings

1. **Successful system identification:** Machine learning successfully recovered the pendulum's governing equation from experimental data
2. **High model accuracy:** $R^2 = 0.884$ indicates excellent fit with low prediction error (RMSE = 0.141)
3. **Physically meaningful results:** Discovered coefficients align with pendulum physics expectations
4. **Robust methodology:** Ridge regression with cross-validation provides reliable parameter estimation

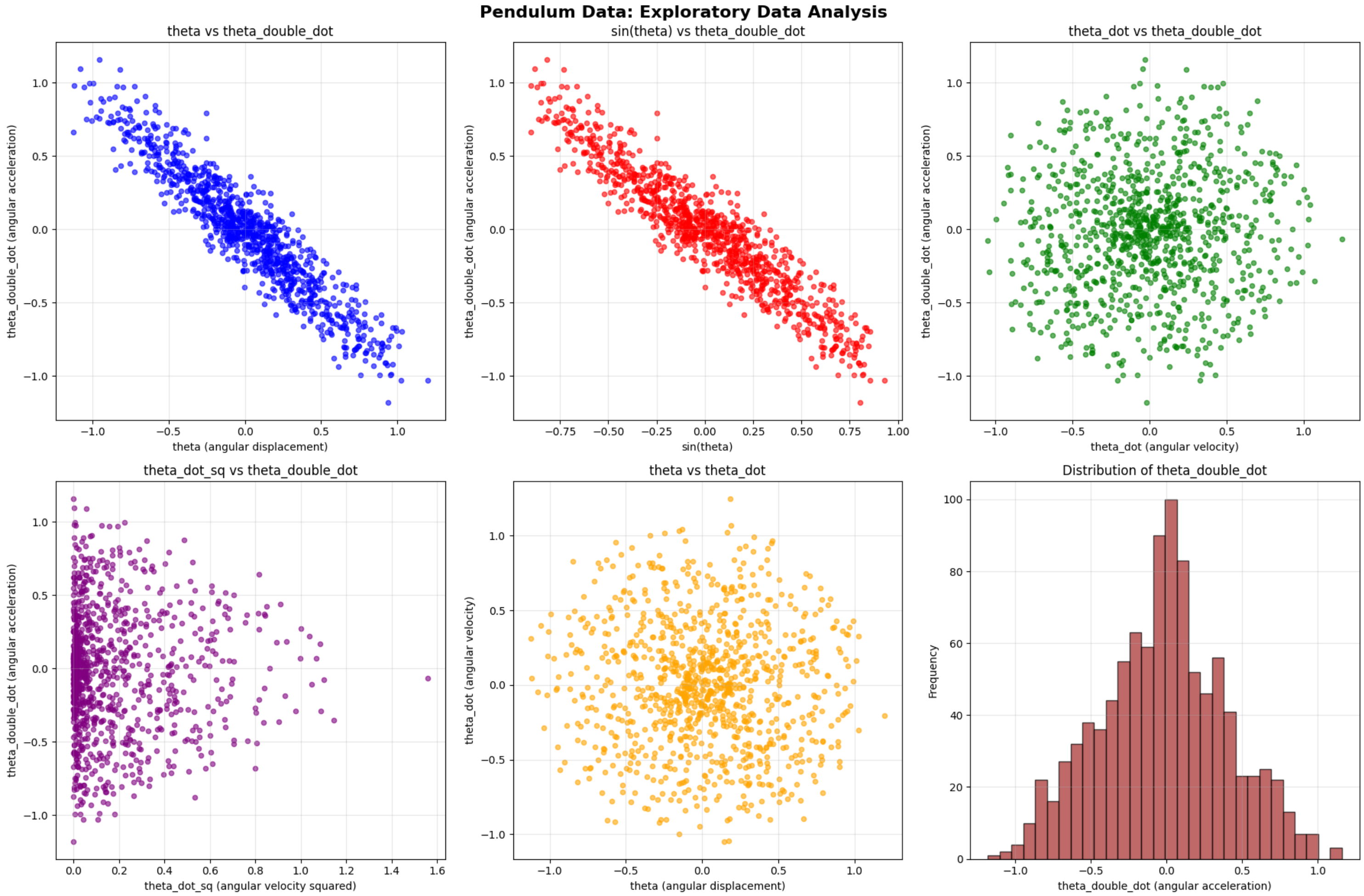
Key Achievements

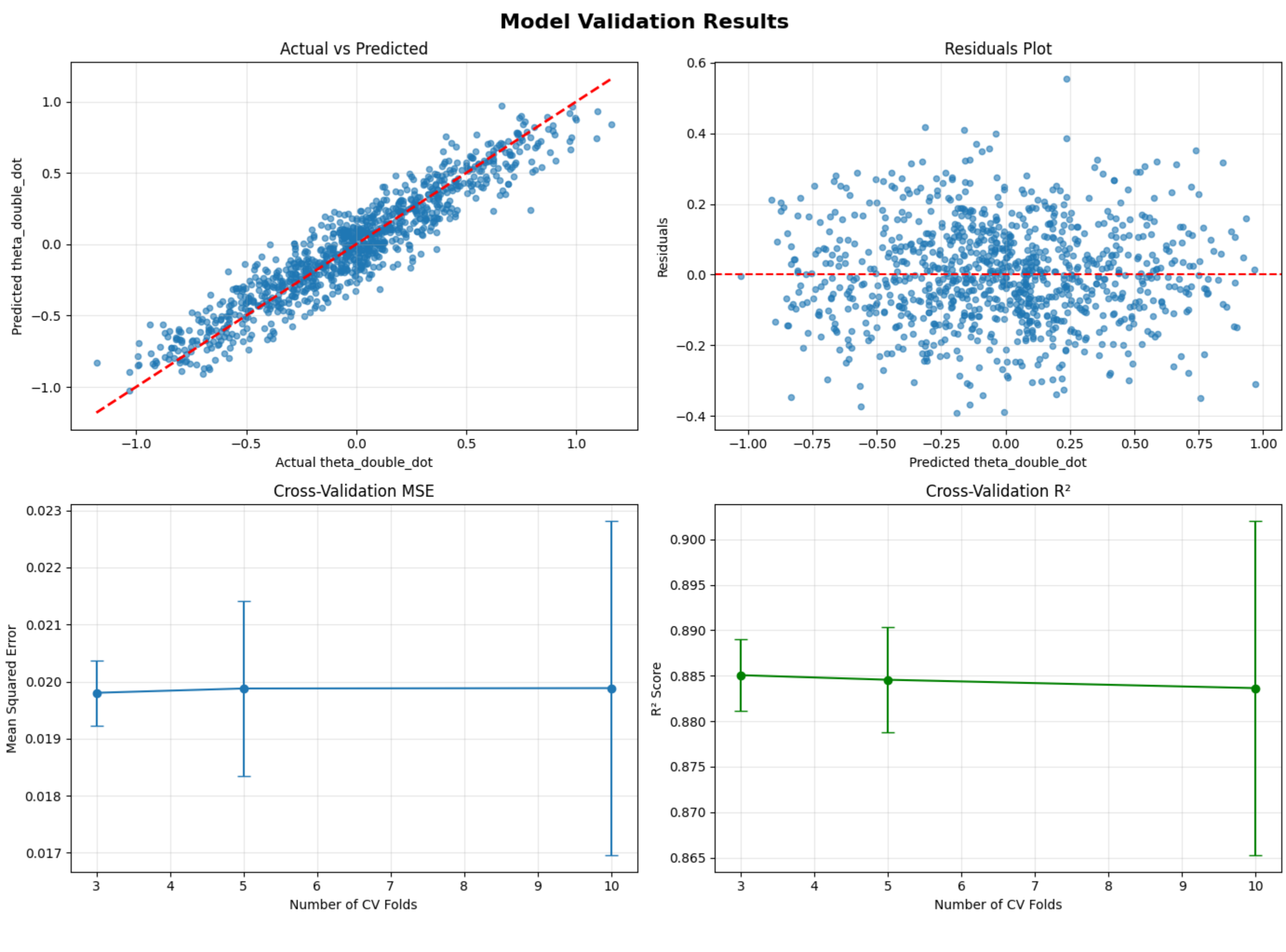
- **Correlation analysis** identified the most relevant features for pendulum dynamics
- **Feature engineering** ($\sin\theta$, $\dot{\theta}^2$) enhanced model representation of nonlinear effects
- **Ridge regularization** prevented overfitting while maintaining predictive accuracy
- **Cross-validation** ensured model generalizability and parameter robustness

Physical Insights

The discovered equation reveals:

- **Primary dynamics:** Strong restoring forces from both linear and nonlinear terms
- **Minimal damping:** Low-friction experimental setup with excellent data quality
- **Nonlinear effects:** sin θ term captures large-angle pendulum behavior
- **System parameters:** Effective $g/L \approx 0.53$-0.41, indicating specific pendulum characteristics


Pendulum Data: Exploratory Data Analysis
theta vs theta_double_dot
theta (angular displacement)
theta_double_dot (angular acceleration)
sin(theta) vs theta_double_dot
sin(theta)
theta_dot vs theta_double_dot
theta_dot (angular velocity)
theta_dot_sq vs theta_double_dot
theta_dot_sq (angular velocity squared)
theta vs theta_dot
Distribution of theta_double_dot
Frequency
theta_double_dot (angular acceleration)


Model Validation Results
Actual vs Predicted
Actual theta_double_dot
Predicted theta_double_dot
Residuals Plot
Residuals
Cross-Validation MSE
Mean Squared Error
Number of CV Folds
Cross-Validation R²
R² Score