

Week3 - Linear Regression with effect of standardization of data set

DS3010: Introduction to Machine Learning Lab

Timing: 02:00 PM to 04:45 PM

Max Marks: 5

Instructions

1. Submit one .ipynb file containing all answers. The name should be [student_name]_[rollno]_[lab].ipynb
 2. Write the questions in separate text blocks before the answers.
 3. Outputs for all sub-questions should be given and the code should be executable.
 4. Write justifications for your choices where needed.
 5. Ensure that all plots include clear labels and legends for better interpretation.
 6. Use of generative AI tools (such as ChatGPT, Gemini, etc.) is strictly prohibited. Any submission found to contain AI-generated or plagiarized content will receive a score of zero, and disciplinary action.
-

1. **You are provided with a dataset - Median house prices for California districts derived from the 1990 census provided in csv format - housing.csv**
 - (a). Load the dataset and split it into training and testing sets. **(0.5 Marks)**
 - (b). Train a Linear Regression model without any preprocessing. Record the coefficients, intercept, and model performance (e.g., Mean Squared Error, R^2 score). **(2 Marks)**
 - (c). Standardize the dataset features (zero mean, unit variance). Retrain the Linear Regression model using the standardized dataset. Again record coefficients, intercept, and performance. **(2 Marks)**
 - (d). Compare the results of both models and explain: **(0.5 Marks)**
 - How did standardization affect the coefficients?
 - Was there any improvement in performance metrics?
 - Why might standardization be more important when features are on very different scales?
-

Tasks to Submit:

- Python code with proper comments.
- Plots of predictions vs actual values for both cases.
- Printout of coefficients before and after standardization.
- A short report (5–10 lines) summarizing your observations on the effect of standardization in Linear Regression.