Multiple linear Regression Report

Introduction

Multiple Linear Regression is a statistical method used to model the relationship between a dependent variable (target) and two or more independent variables (predictors). This report analyzes the performance of a Multiple Linear Regression model applied to a given dataset.

Methodology

The Multiple Linear Regression model follows the equation:

$Y=b_0+b_1X_1+b_2X_2+...+b_nX_n$

where:

- Y is the dependent variable (target).
- $X_1, X_2, ..., X_n$ are independent variables (predictors).
- b_0 is the intercept.
- $b_1, b_2, ..., b_n$ are the regression coefficients (slopes).

Steps Taken:

1. Data Preprocessing:
    - Checked for missing values.
    - Performed exploratory data analysis (EDA) to understand feature distributions and correlations.
    - Encoded categorical variables (if any) using one-hot encoding.
    - Split data into training (70%) and testing (30%) sets.
2. Model Training:
    - Used sklearn.linear_model.LinearRegression to fit the model.
    - Checked model coefficients to interpret feature importance.
3. Performance Evaluation:
    - Measured using R^2 Score, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Multiple Linear Regression Model Analysis

No	Model Type	Parameters	Results
01	Standard Linear Regression	Basic model	0.9358
		fit_intercept=True	0.9358
		fit_intercept=False	0.7389
		normalize	-19155898675.7897
		n_jobs	-21724334601.2341
02	Ridge Regression (L2 Regularization)	alpha=1.0	0.9357
		solver='saga'	0.9353
		tol =0.001	0.9357
03	Lasso Regression (L1 Regularization)	max_iter=1000	0.9357
		selection='cyclic'	0.9358
04	Elastic Net Regression (Combination of L1 & L2)	l1_ratio=0.5	0.9355
		max_iter=1000	0.9355
05	Feature Scaled Regression	with_mean=True	0.9355
		with_std=True	0.9355

Conclusion

The analysis of multiple linear regression models reveals several key insights:

1. Standard Linear Regression: The basic model performs well with a result of 0.9358 when the intercept is included. However, excluding the intercept significantly reduces performance to 0.7389. Normalization and the use of n_jobs parameter lead to extremely poor results, indicating potential issues with these settings.

2. Ridge Regression (L2 Regularization): This model shows consistent performance with slight variations based on the solver and tolerance settings. The results are very close to the standard linear regression, indicating that L2 regularization does not significantly impact performance in this case.

3. Lasso Regression (L1 Regularization): Similar to Ridge Regression, Lasso Regression performs well with results close to the standard model. The selection method (cyclic) does not significantly alter the outcome.

4. Elastic Net Regression: Combining L1 and L2 regularization, this model also performs well, with results slightly lower than the standard model but still robust.

5. Feature Scaled Regression: Scaling features with mean and standard deviation does not significantly impact the model's performance, maintaining results similar to other models.

Recommendations

1. Standard Linear Regression: Use this model with fit_intercept=True for optimal performance. Avoid normalization and the n_jobs parameter as they degrade results.
2. Regularization Techniques: Both Ridge and Lasso regressions are viable alternatives if regularization is needed. They provide similar performance to the standard model and can help prevent overfitting.
3. Elastic Net Regression: Consider this model if a combination of L1 and L2 regularization is desired. It offers a balanced approach with minimal performance trade-off.
4. Feature Scaling: While feature scaling does not significantly impact performance in this case, it is generally good practice to scale features, especially when using regularization techniques.

Final Recommendation

Given the analysis, the Standard Linear Regression model with fit_intercept=True is recommended for its simplicity and high performance. If regularization is a concern, Ridge Regression or Lasso Regression can be used as they provide similar results with added benefits of regularization. Elastic Net Regression is also a good option if a combination of regularization techniques is preferred. Feature scaling should be considered as a standard preprocessing step, although it does not significantly impact the results in this specific scenario.