Support Vector Regression (SVR) Report

Introduction

Support Vector Regression (SVR) is a powerful machine learning technique used for regression tasks. Unlike traditional linear regression models, SVR utilizes Support Vector Machines (SVM) principles to minimize errors while maintaining model complexity. This report evaluates the performance of an SVR model applied to a given dataset.

Methodology

SVR is based on the following equation:

$Y = wX + b$

where:
- Y is the dependent variable (target).
- X represents the independent variables (predictors).
- w represents the weight vector.
- b is the bias term.

Steps Taken:

1. 1. Data Preprocessing:

- - Checked for missing values.
- - Conducted exploratory data analysis (EDA) to examine feature distributions and correlations.
- - Encoded categorical variables (if any) using one-hot encoding.
- - Scaled features using StandardScaler for optimal performance.
- - Split data into training (70%) and testing (30%) sets.

2. 2. Model Training:

- - Used sklearn.svm.SVR to train the model.
- - Tested different kernel functions (linear, polynomial, radial basis function, and sigmoid).
- - Tuned hyperparameters such as C, epsilon, and gamma to optimize model performance.

3. 3. Performance Evaluation:

- - Measured using R^2 Score, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Support Vector Regression Model Analysis

No	Model Type	Parameters	Hyper Parameters	Results
01	Linear SVR	kernel="linear"	-	0.8950
02	Polynomial SVR (Degree 3)	kernel="poly"	C=1.0	-0.0508
03	RBF SVR (Default Gamma)	kernel="rbf"	C=10	-0.0558
04	Sigmoid SVR	kernel="sigmoid"	C=1.0	-0.0574
05	Custom Epsilon-SVR with RBF Kernel	kernel="rbf"	C=100	-0.0574
06	Optimized RBF SVR	kernel="rbf" / epsilon=0.1	C=100	-0.0302

Conclusion

The Support Vector Regression (SVR) analysis reveals key insights into the effectiveness of different kernel functions and hyperparameter settings:

1. **Linear SVR** achieved the highest result (0.8950), indicating that the data might have a strong linear relationship.
2. **Polynomial SVR** and **Sigmoid SVR** performed poorly, suggesting that these kernels do not generalize well for the given dataset.
3. **RBF SVR (Default Gamma)** with **C=10** had a negative result (-0.0558), meaning the default settings were not optimal.
4. **Custom Epsilon-SVR with RBF Kernel (C=100)** showed an even lower result (-0.0574), indicating that a higher C value alone is not enough to improve performance.
5. **Optimized RBF SVR (C=100, epsilon=0.1)** performed better than the default RBF but still did not outperform Linear SVR.

Recommendations

1. **Use Linear SVR as the primary model** since it achieved the best result. This suggests that the dataset has a strong linear relationship, making non-linear kernels unnecessary.
2. **Avoid Polynomial and Sigmoid SVR** as they yield poor performance, making them unsuitable for this dataset.
3. **Optimize RBF SVR further** by fine-tuning C, gamma, and epsilon. While the optimized RBF SVR performed better than the default, it still did not surpass Linear SVR.
4. **Perform feature scaling and engineering** to improve model performance, especially for non-linear kernels like RBF.
5. **Experiment with different regularization techniques** such as adjusting C and epsilon values for better generalization.

Final Recommendation

Based on the analysis, **Linear SVR is the best choice** for this dataset due to its superior performance. If non-linearity is expected in future datasets, **further tuning of RBF SVR** may be explored. Regularization and feature selection should also be considered to enhance model robustness