A Comparative Study of LS-SVM and SVM Algorithms for Regression and Classification Problems

Sai Likhith Yadav Seera
Computer Science
School of Computing and Augmented Intelligence
Arizona State University
Tempe, United States
sseera@asu.edu

***Abstract*—SVM and LS-SVM are widely used algorithms for classification and nonlinear function estimation. However, SVM's main drawback is its high computational burden due to constrained optimization programming. In contrast, LS-SVM solves linear equations instead of quadratic programming problems [1], addressing this issue. SVM and LS-SVM are popular supervised learning algorithms with diverse applications in ML, including classification, regression, and outlier detection. LS-SVM is a modified version of SVM designed to enhance computational efficiency and generalization performance. These algorithms have demonstrated impressive results in image and speech recognition, text classification, and bioinformatics. In our work, we focus on implementing SVM and LS-SVM as classifiers and regressors, comparing their performance using benchmark datasets.**

***Index Terms*—Support Vector Machine, Least Square Support Vector Machine, Kernel, Classifier, Regressor.**

I. Introduction

A. Support Vector Machine (SVM)

Support Vector Machine, is a widely used machine learning algorithm for classification and regression tasks. Its primary objective is to find a hyperplane that effectively separates different classes in a dataset. In binary classification, SVM aims to discover a decision boundary with maximum margin, representing the distance between the boundary and the nearest data points from each class, known as support vectors. This hyperplane achieves optimal classification. For multi-class classification, SVM employs one-vs-one or one-vs-all approaches. SVMs can perform nonlinear classification using the kernel trick, mapping data to higher-dimensional feature spaces without explicit calculation. This capability allows SVMs to handle complex decision boundaries effectively. Additionally, SVM can be used for regression by identifying a hyperplane that minimizes the distance between predicted and actual values. [2]

B. Least Square Support Vector Machine (LS-SVM)

Suykens introduced the Least Square Support Vector Machine (LSSVM) as a variation of the traditional SVM, with the aim of reducing computational complexity. In LSSVM, inequality constraints are replaced with equality constraints during the quadratic programming solution, resulting in a faster training process compared to SVM. However, the lack of sparseness and robustness in the computational solution are two primary drawbacks of LSSVM. These limitations may increase the training time and reduce the accuracy of predicted models when dealing with real industrial data sets with complex characteristics such as imbalanced distribution, heteroscedasticity, and the explosion of data. [3] [4]

In this project, we conduct a comparative analysis of two algorithms used for regression and classification tasks. We have implemented both LS-SVM and SVM algorithms as classifiers and regressors. These methods were applied to benchmark datasets [5] [6], including the Breast Cancer Wisconsin (Diagnostic) Data Set and Extreme-Weather Temperature Prediction. We assessed their performance using various metrics, including accuracy, precision, recall, explained variance score (EV), and root mean square error (RMSE).

II. Description of Solution

We have used utility libraries such as sklearn, scipy, numpy, and pandas in Python to perform our project. Here is an overview of our solution:

- **Data Import and Preprocessing:** We have imported the datasets into a python notebook using pandas, thereby organizing them into data frames. Preprocessing techniques were applied on the data frames which include handling missing values by dropping them and encoding labels.

- **Feature Analysis:** We have conducted an analysis of the dataset's features which involves determining the correlation between them.

- **Data Split:** To train and validate our models, we have split the dataset into two parts: one with the features (X) and the other with the labels (Y). This split was achieved using the train_test_split() function from sklearn in python, with a ratio of 70% for training data and 30% for validation data.

- **Data Normalization:** We have applied Min-Max normalization to both the training and validation data, transforming them to a range of [0,1].

- **Implementation of Model:** We have implemented LSSVC and LSSVR methods for various kernel functions and also we have utilized SVC and SVR models from the sklearn.svm library in python for our work. The models were fitted using the X_train and Y_train data after applying the normalization.

- **Prediction and Evaluation:** With these models fitted using the the X_train and Y_train data, we have predicted the binary labels (in case of classification) and output values (in case of regression) (Y_pred) using the X_test validation data. For assessing the model's performance, we have calculated various performance metrics and confusion matrices for all the kernels in each of the model. In the case of classification, metrics like Accuracy, Misclassification Error, Precision, Recall, F1 score, Specificity, Sensitivity, etc. have been calculated. Whereas, in the case of regression, metrics like Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, R^2 Score, Explained Variance Score, etc. have been calculated.

- **Performance Evaluation:** Based on the calculated metrics and the nature of the data considered in our work, we have evaluated the performance of the classifiers and regressors for both SVC, SVR, LSSVC and LSSVR methods.

III. Results

The performance metrics of the Extreme-Weather Temperature Prediction Data Set have been computed for both the Support Vector Regressor and the Least Square Support Vector Regressor as shown in table I and table II.

TABLE I
METRIC COMPARISONS OF SVR FOR FOUR DIFFERENT KERNELS

Performance Metrics	**Kernels**			
	Linear	***RBF***	***Poly***	***Sigmoid***
MAE	1.01	0.88	0.92	28.95
MSE	1.85	1.46	1.53	1543.29
RMSE	1.36	1.21	1.24	39.28
R^2 Score	0.80	0.85	0.84	-162.64
EV score	0.80	0.85	0.84	-155.37

TABLE II
METRIC COMPARISONS OF LSSVR FOR FOUR DIFFERENT KERNELS

Performance Metrics	**Kernels**			
	Linear	***RBF***	***Poly***	***Sigmoid***
MAE	1.01	0.89	0.90	20.10
MSE	1.84	1.40	1.39	1036.95
RMSE	1.36	1.18	1.18	32.20
R^2 Score	0.81	0.85	0.85	-108.95
EV score	0.81	0.85	0.85	-108.93

The performance metrics of the Breast Cancer Wisconsin (Diagnostic) Data Set have been computed for both the Support Vector Classifier and the Least Square Support Vector Classifier as shown in table III and table IV.

TABLE III
METRIC COMPARISONS OF SVC FOR FOUR DIFFERENT KERNELS

Performance Metrics	**Kernels**			
	Linear	***RBF***	***Poly***	***Sigmoid***
Accuracy	0.99	0.99	0.99	0.37
Misclassification Error	0.01	0.01	0.01	0.63
Precision	1.00	0.98	0.98	0.00
Recall	0.98	0.98	0.98	0.00
F1 score	0.99	0.98	0.98	0.00
Specificity	0.98	0.98	0.98	0.00
Sensitivity	1.00	0.99	0.99	0.57
AUC score	0.99	0.99	0.99	0.28

TABLE IV
METRIC COMPARISONS OF LSSVC FOR FOUR DIFFERENT KERNELS

Performance Metrics	**Kernels**			
	Linear	***RBF***	***Poly***	***Sigmoid***
Accuracy	0.97	0.99	0.98	0.80
Misclassification Error	0.03	0.01	0.02	0.20
Precision	1.00	1.00	0.98	0.72
Recall	0.98	0.97	0.95	0.70
F1 score	0.99	0.98	0.97	0.71
Specificity	0.98	0.97	0.95	0.70
Sensitivity	1.00	1.00	0.99	0.86
AUC score	0.99	0.98	0.97	0.78

The confusion matrix for Support Vector Classifier when run using different kernel functions is shown in the tables V,VI,VII and VIII and the confusion matrix for Least Square Support Vector Classifier when run using different kernel functions is shown in the tables IX,X,XI and XII.

TABLE V
CONFUSION MATRIX OF SVC FOR LINEAR KERNEL

Actual Values	**Predicted Values**	
	Benign	***Malignant***
Benign	59	1
Malignant	0	111

TABLE VI
CONFUSION MATRIX OF SVC FOR RBF KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	59	1
Malignant	1	110

TABLE VII
CONFUSION MATRIX OF SVC FOR POLYNOMIAL KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	59	1
Malignant	1	110

TABLE VIII
CONFUSION MATRIX OF SVC FOR SIGMOIDAL KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	0	60
Malignant	48	63

TABLE IX
CONFUSION MATRIX OF LSSVC FOR LINEAR KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	55	5
Malignant	0	111

TABLE X
CONFUSION MATRIX OF LSSVC FOR RBF KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	58	2
Malignant	0	111

TABLE XI
CONFUSION MATRIX OF LSSVC FOR POLYNOMIAL KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	57	3
Malignant	1	110

TABLE XII
CONFUSION MATRIX OF LSSVC FOR SIGMOIDAL KERNEL

Actual Values	Predicted Values	
	Benign	***Malignant***
Benign	42	18
Malignant	16	95

IV. TEAM MEMBERS AND CONTRIBUTIONS

We are a group of two people, and my teammate's name is ***Riya Salian***. She played an integral part in this project and together made it successful.

The contributions that I have made to this project are as follows:

- Implemented the Classifier and Regressor methods used in Least Square Support Vector Machine.
- Analyzed the evaluation metrics for the models used.
- Documenting our project's report.
- **Extra Work:** I have applied feature selection technique named *Recursive Feature Elimination(RFE)* in order to find a subset of highly important features. In our work, it turned out to be that all the features were relevant in its higher performance of the model. This indicated that it is dependent on the nature of the data.

V. KEY LEARNINGS FROM THE PROJECT

I have learned the following lessons during the course of my project:

- I have learnt the insights of Support Vector Machine and also how Least Square Support Vector Machine is a derived variant from it.
- Also, I have learnt what exactly are classifier and regressor and how SVM and LSSVM could be used in both ways and how kernel functions and hyperplane equation determine the nature of the model.
- Moreover, I have learnt how the performance of these models can be influenced by several factors such as data preprocessing, tuning of hyperparameters, and data analysis to understand the nature of the data.

REFERENCES

[1] Wang, Haifeng & Hu, Dejin. (2005). Comparison of SVM and LS-SVM for regression. Proceedings of 2005 International Conference on Neural Networks and Brain Proceedings, ICNNB'05. 1. 279 - 283. 10.1109/ICNNB.2005.1614615 https://www.researchgate.net/publication/4233965_Comparison_of_SVM_and_LS-SVM_for_regression.

[2] Bernhard Schölkopf, Alexander J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. https://books.google.com/books?hl=en&lr=&id=y8ORL3DWt4sC&oi=fnd&pg=PR13&dq=mathematics+behind+support+vector+classifier .

[3] Lov Kumar, Sai Krishna Sripada, Ashish Sureka, Santanu Ku. Rath, Effective fault prediction model developed using Least Square Support Vector Machine (LSSVM), Journal of Systems and Software, Volume 137, 2018, Pages 686-712, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2017.04.016. https://www.sciencedirect.com/science/article/pii/S0164121217300717.

[4] Suykens, J., Vandewalle, J. Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9, 293–300 (1999). https://doi.org/10.1023/A:1018628609742 https://link.springer.com/article/10.1023/A:1018628609742.

[5] Breast Cancer Wisconsin (Diagnostic) Data Set. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.

[6] Extreme-Weather Temperature Prediction Data Set. https://www.kaggle.com/datasets/gauravduttakiit/extreme-weather-temperature-prediction.