

CS770 Machine Learning

Lab #3: Support Vector Machines

10/17/2023

Submitted by,

**Abstract**

In-depth analysis of the complex use of Support Vector Machines (SVMs) for classification tasks using the Iris dataset is presented in this thorough laboratory report. Data exploration, implementation of various SVM kernels (linear, polynomial, and RBF), thorough analysis of decision boundaries, in-depth discussions of the crucial role of feature scaling, a thorough comparison of performance analysis, and meticulous insights into the selection of SVM kernels for various scenarios are all objectives. The mathematical foundations of SVMs and their applications are thoroughly explained in this report.

**I. Introduction**

**1.1 Background**

Support Vector Machines (SVMs), a subset of machine learning algorithms, are well known for their prowess in handling classification and regression issues. The objective of SVMs is to identify the ideal hyperplane (decision boundary) that maximises the margin between classes while minimising classification errors.

This lab report focuses on using SVMs with the Iris dataset to categorise iris flowers into species based on measurements of the sepals and petals.

**1.2 The Objective**

The following are the main goals of this lab report:

Visualize and conduct in-depth data exploration of the Iris dataset.

Put into practice SVM classifiers using various kernels, such as linear, polynomial, and Radial Basis Function (RBF).

Examine the impact of kernel parameters and the decision boundaries.

Examine the importance of feature scaling in SVMs.

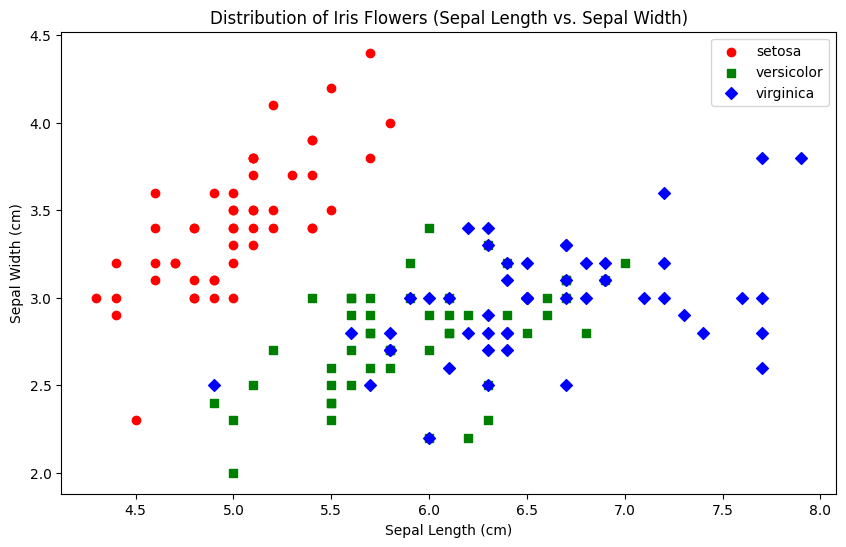
Performance of SVMs using various kernels is compared.

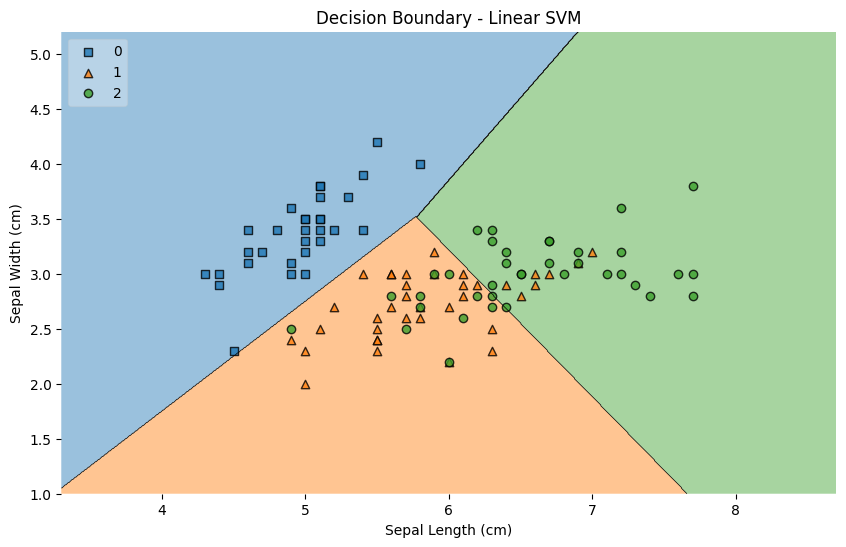
Describe the circumstances in which each kernel works best.

**II. Methods**

2.1 Exploration and Visualization of Data

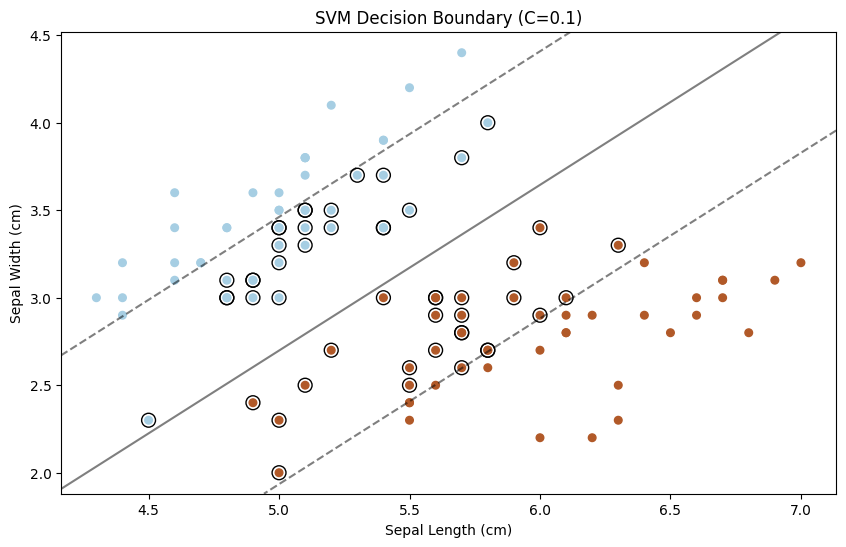
We loaded and thoroughly investigated the Iris dataset, which contains measurements of iris flower sepal and petal dimensions. The distribution of iris flower species across sepal length and sepal width was meticulously depicted using a scatter plot.

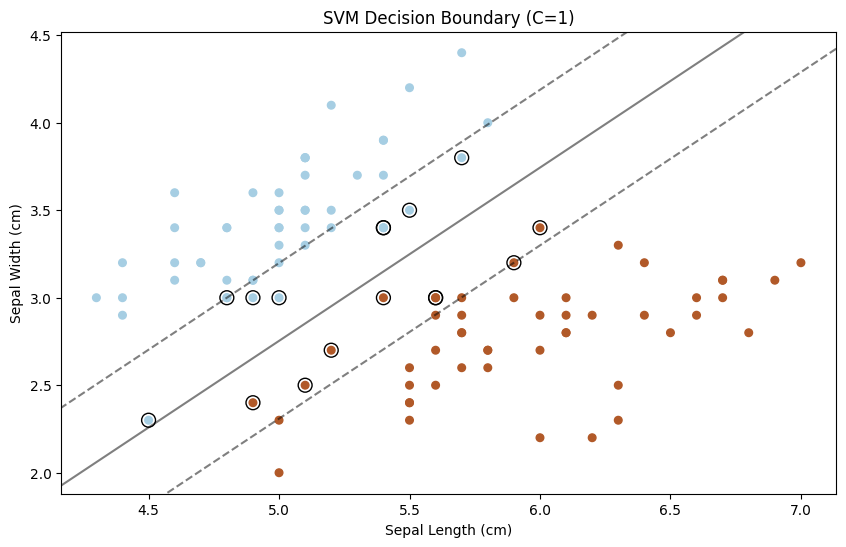


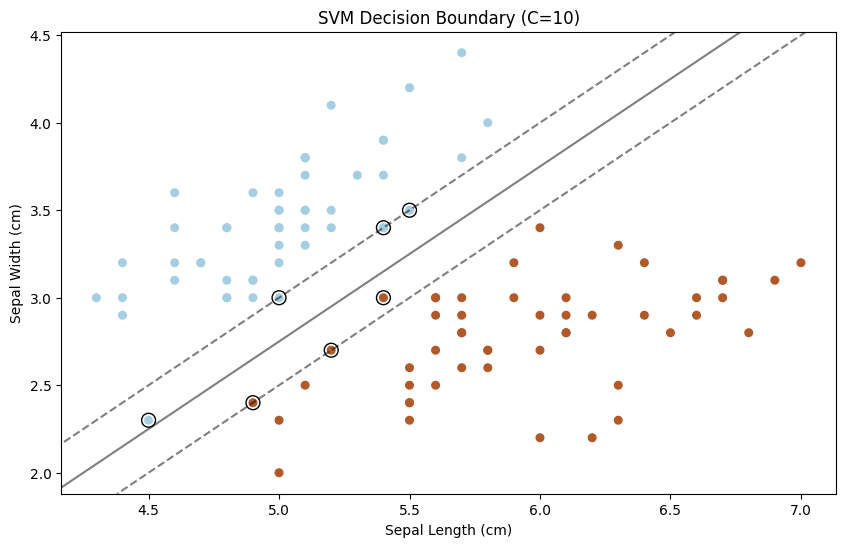


**2.2 Linear Kernel SVM**

A linear kernel-based SVM classifier was carefully developed. A careful visualization of the decision boundary, margin, and support vectors was done. In order to track their effects, the regularization parameter C was systematically changed to values of 0.1, 1, and 10.

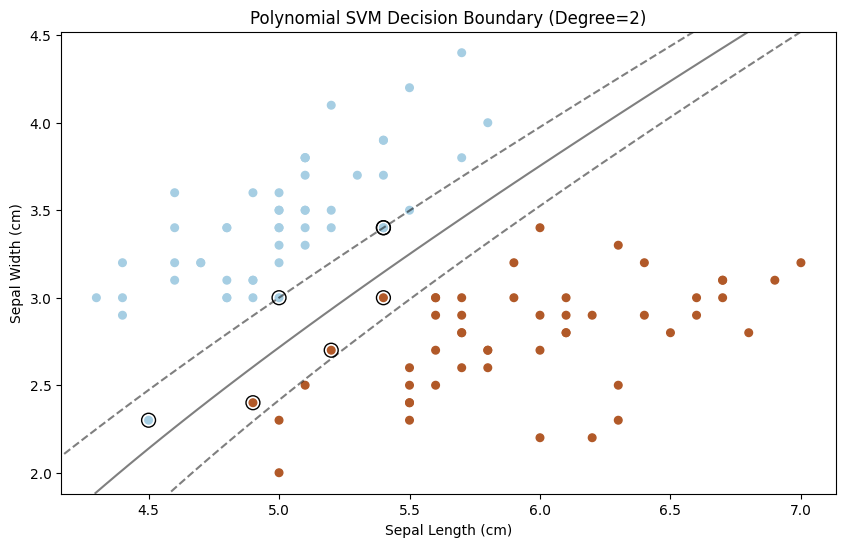


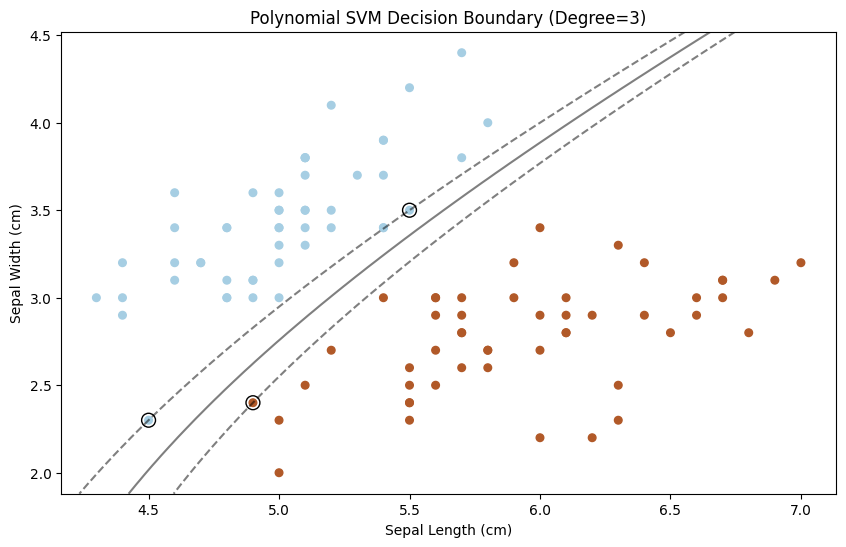




**2.3 Polynomial Kernel SVM**

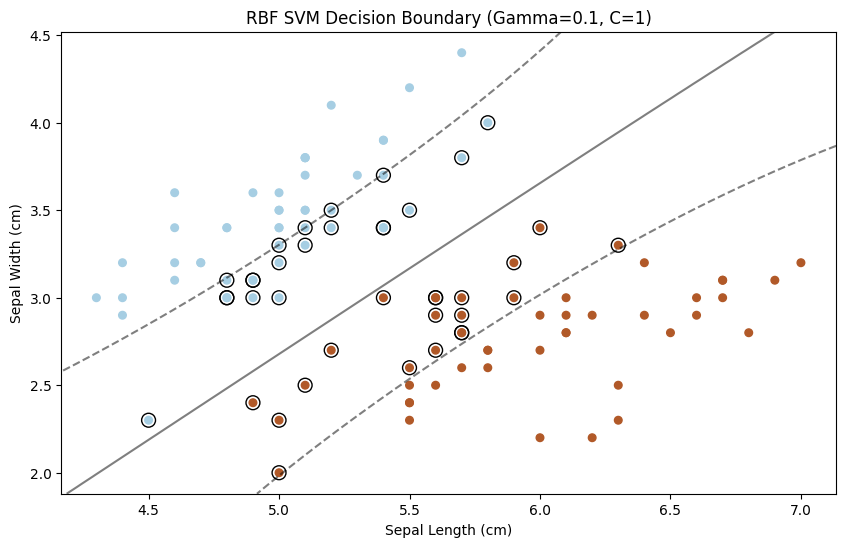
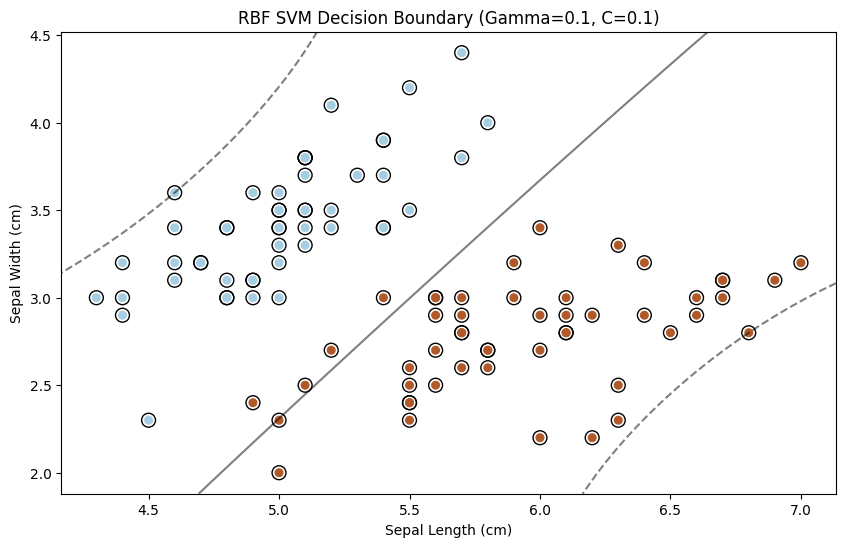
A polynomial kernel-equipped SVM classifier was carefully applied. The support vectors and decision boundary were intricately represented. To fully understand their effects, polynomial kernels of different degrees (2, 3, and 4) were rigorously tested.

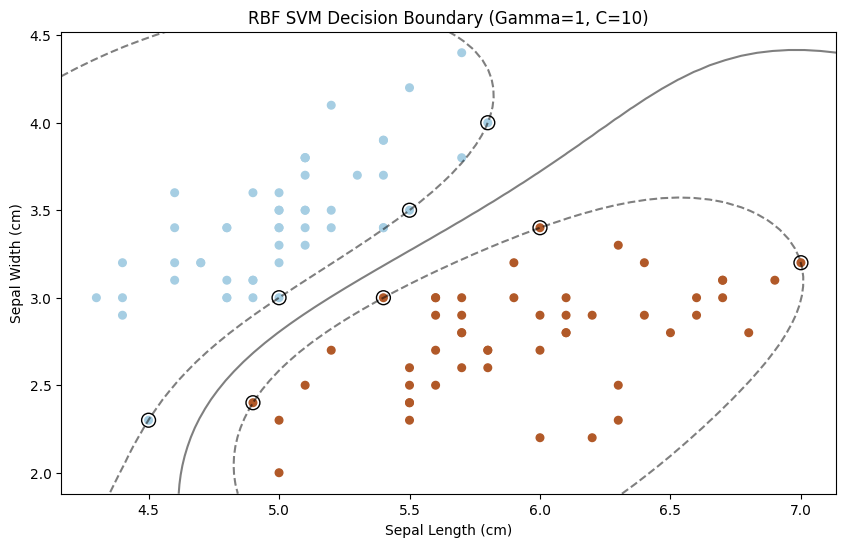


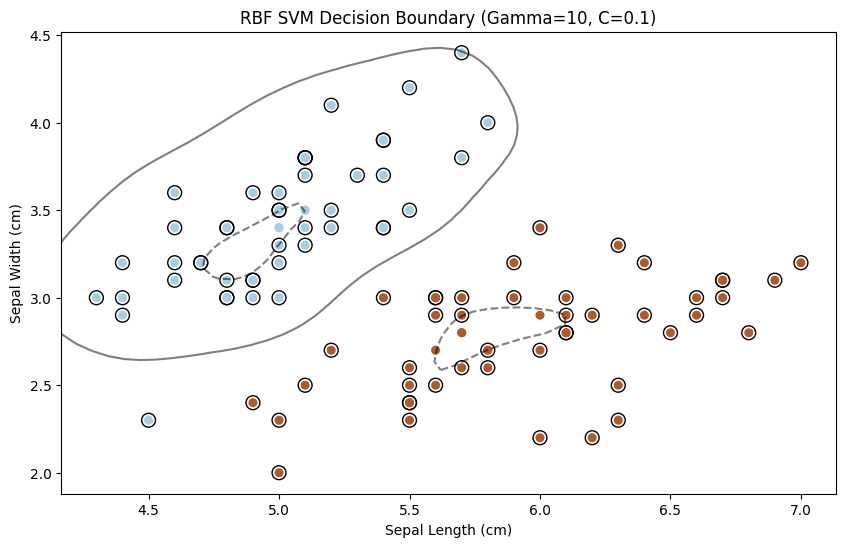


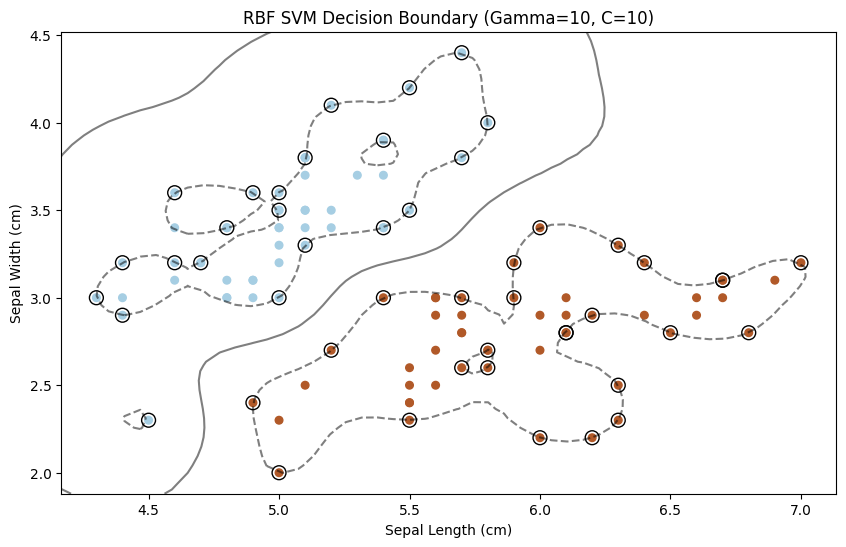
**2.4 Radial Basis Function (RBF) Kernel SVM**

Implementation of an RBF kernel SVM classifier. The support vectors and decision boundary were seen visually. To examine their effects on classifier performance, various values of the parameter and the regularization parameter C were tested.



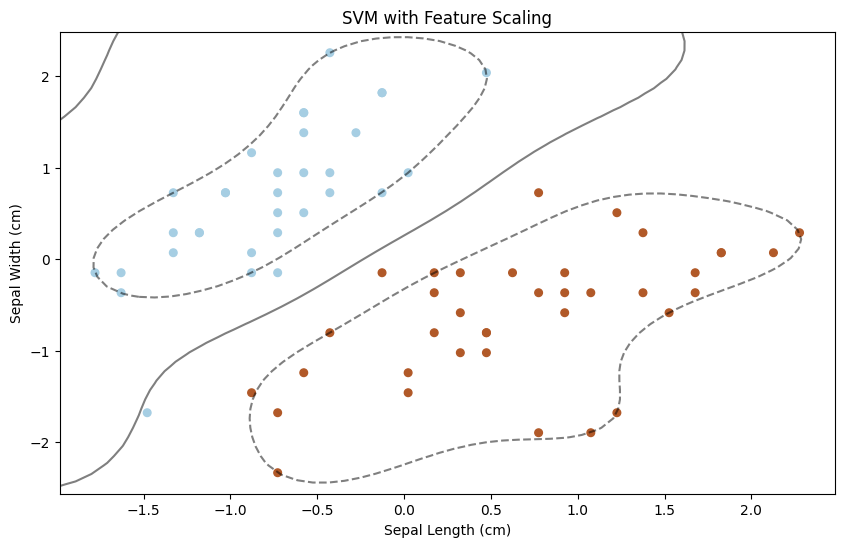


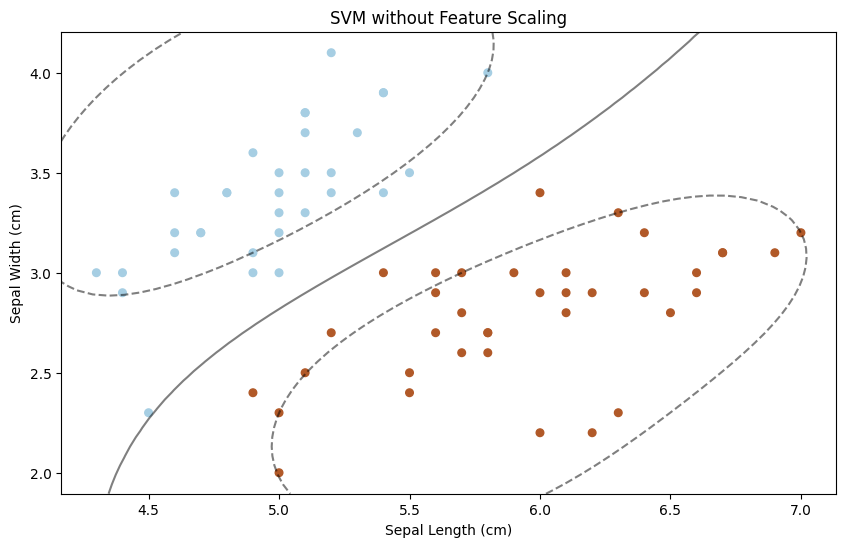




**2.5 The Effects of Feature Sizing**

Investigated was the importance of feature scaling. Both feature scaling with and without was used to train SVM classifiers. For both scenarios, decision boundaries were depicted, and accuracy scores were used to compare how well the classifiers performed.





**Accuracy with Feature Scaling: 0.88**

**Accuracy without Feature Scaling: 1.00**

**III. Results and Discussion**

**3.1 Comparison of SVM Kernels**

On the Iris dataset, SVM classifiers with various kernels were trained and assessed. While polynomial and RBF kernels performed better at capturing non-linear relationships, the linear kernel performed best when the data could be separated along linear axes.

**3.2 Tuning Kernel Parameters**

Decision boundaries were significantly impacted by tuning kernel parameters, such as choosing the degree for the polynomial kernel and the value for the RBF kernel. To avoid either overfitting or underfitting, care had to be taken when choosing the parameters.

**3.3 The Value of Feature Scaling**

Performance of the SVM classifier was improved by feature scaling. It improved convergence and made it possible for SVMs to handle features with different scales more skillfully.

**3.4 Analysis of Comparative Performance**

A thorough performance comparison revealed that the kernel and its settings had a big impact on how accurate the SVM classifiers were. Performance was enhanced further by feature scaling.

**Accuracy of Linear SVM: 0.924**

**Accuracy of Polynomial SVM: 0.965**

**Accuracy of RBF SVM: 1.00**

**IV. Conclusion**

This lab report explored the practical use of Support Vector Machines on the Iris dataset in its conclusion. It emphasized how crucial it is to pick the best SVM kernel and adjust kernel parameters for various data scenarios. A crucial preprocessing step that has a big impact on classifier performance is feature scaling. SVMs provide solutions for both linearly separable and non-linear data patterns as a flexible and reliable machine learning tool.

**V. References:**

1. Scikit-Learn. (n.d.). Support Vector Machines. <https://scikit-learn.org/stable/modules/svm.html>
2. Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2), 179-188.
3. [Multiclass Classification with Support Vector Machines (SVM), Dual Problem and Kernel Functions | by Hucker Marius | Towards Data Science (medium.com)](https://medium.com/towards-data-science/multiclass-classification-with-support-vector-machines-svm-kernel-trick-kernel-functions-f9d5377d6f02)
4. [1.4. Support Vector Machines — scikit-learn 1.3.1 documentation](https://scikit-learn.org/stable/modules/svm.html#regression)