**Exploring Support Vector Machines (SVM) and its Kernel Methods**

**Name: Mahesh Kumar Konda**

**Student ID: 21059972**

**Module Name: Machine Learning and Neural networks.**

**Word Count: 2400 words**

Table of Contents

[Introduction to Support Vector Machines (SVM) 3](#_Toc183528119)

[Kernels 4](#_Toc183528120)

[Linear Kernel 5](#_Toc183528121)

[Polynomial Kernel 5](#_Toc183528122)

[Radial Basis Function (RBF) Kernel 6](#_Toc183528123)

[Dimensionality Reduction with PCA 6](#_Toc183528124)

[Why is PCA used? 7](#_Toc183528125)

[Dataset: Iris Dataset 8](#_Toc183528126)

[Code Walkthrough 8](#_Toc183528127)

[Step 1: Importing the necessary Libraries 8](#_Toc183528128)

[Step 2: Loading and Preparing the dataset 9](#_Toc183528129)

[Step 3: We Used PCA for Dimensionality Reduction 9](#_Toc183528130)

[Step 4: Training SVM Models 10](#_Toc183528131)

[Step 5: Visualizing Decision Boundaries 11](#_Toc183528132)

[Step 6: Evaluating the Model Performance 11](#_Toc183528133)

[Discussion Results 13](#_Toc183528134)

[Decision Boundaries 13](#_Toc183528135)

[Accuracy Evaluation 16](#_Toc183528136)

[Conclusion 17](#_Toc183528137)

[References 18](#_Toc183528138)

# Introduction to Support Vector Machines (SVM)

Support Vector Machines (SVM) are one of the most important and universally used supervised machine learning techniques or algorithms that is used mainly for classification tasks. The SVM algorithm has the capability of making decision boundaries that can work in high-dimensional spaces; hence, it is very popular among the machine learning engineers. This algorithm is also widely used for image classification, bioinformatics, and text classification.

The main concept that SVM follows is finding a hyperplane, which helps them separate data points from different classes in the feature space. The primary objective of SVM is to find the best hyperplace that is best for maximizing the margin between the classes. This margin is best for ensuring the best separation between classes of the hyperplane (Tariq *et al*., 2023). This tutorial will be focused on some points, which are written below:

1. **SVM with different kernels**

The SVM, collaborating with the kernels, is able to perform brilliant tasks of handling linear and non-linear classification tasks. There are three types of kernels present in the SVM; these kernels are the polynomial, linear, and radial basis.

1. **Dimensionality reduction using principal component**

This is used to analyze decision boundaries in 2D. For this dimensionality reduction, we will use the PCA which is Principal Component Analysis, which is a statistical tools used to reduce the dimension for the sake of convenience for understanding the visualization on the hyperplane

1. **Code implementation**

Python is the best-suited coding language in order to accommodate machine learning algorithms because of its powerful libraries. In this report, we will use Python's scikit-learn library. The SVM code is implemented in Python, in order to use the famous database available in Python, that is the Iris Dataset. The Scikit Learn library is used here in order to implement the Support Vector Machines with the Iris dataset.

By the end of this article, a clear picture of how the SVM works will be presented, and along with that, it will also produce how different kernels affect the decision boundaries and how the dimensionality reduction helps all of us in order to depict the performance of SVM (Mustaqeem and Saqib, 2021 ).

# Kernels

In SVM, kernels play a huge role in determining decision boundaries. Talking about the kernel, this is a mathematically functioning tool. The decision boundary defined by the SVM is heavily influenced by the type of kernel function used over there. The kernel is used for transforming the data into a higher-dimensional space where it is easier to find a hyperplace that separates this from the classes. There are three major types of kernels present in the SVM. The first type is the linear kernel, which is used for linearly separated data and cannot handle nonlinear data and is mostly used in datasets that are simple, then there are polynomial kernels; and the last one is the RBF, which is Radial Basis Function which is used for datasets that are nonlinear and are more complex in nature (Singgalen 2023).

## Linear Kernel

Among the three kernels, the linear kernel is the easiest, and this finds the hyperplane in order to separate the available data points.

**Formula**

**K(X,Y) = X \* Y**

* **Use Case:**

This stage is ideal when the data is already separated linearly. There is another case when this stage is efficient: when the user does not want to complicate the model even more with complex mappings and with some higher-degree polynomials.

## Polynomial Kernel

Unlike the linear kernel, this kernel can handle the data when the data is not linearly separable. With this kernel, it is possible to create more flexible decision boundaries, and this kernel is also capable of handling data with a higher-dimensional severability.

**Formula**

**K(x,y)=(x⋅y+c)d**

**Use case:**

This kernel is to be used when the user is expecting the data to have a polynomial relationship and, in other cases, if the user needs a decision boundary that is higher in degree.

## Radial Basis Function (RBF) Kernel

Among the three, this is the kernel, which is most popular and can handle the data more flexibly. This kernel will be in use when the decision boundary is nonlinear. It is best fitted for handling complex datasets by mapping them into an infinite-dimensional space (Roy and Chakraborty, 2023).

**Formula: K(x,y)=exp(−γ∣∣x−y∣∣2)**

**Use Case**

The RBF kernel is used when the user cannot detect any linear boundary between the classes and at the time when the data points that are provided are quite complex.

# Dimensionality Reduction with PCA

The PCA, which is Principal Component Analysis (PCA), helps in depicting the performance of the SVM models. Before we get into the implementation of the SVM models, it is important to understand the importance of PCA. This PCA technique is used mostly in order to reduce the dimensionality of the large datasets. While this is reducing the dimensionality of the datasets, this PCA maintains the information contained inside the datasets. This technique projects the data into the lower-dimensional space. Thus, by doing this, we will be able to see the decision boundaries created by SVM in two-dimensional space and three-dimensional space.

## Why is PCA used?

* **Simplification:**

The PCA helps us by reducing the complexity of high-dimensional data, thereby allowing us to see the projections in the 2D and 3D models. By using this principal component, it is much easier to identify the decision boundaries in the hyperplane.

* **Performance:**

High dimensions mean the speed of the data is reduced. Hence, by reducing the dimensions of the data, it increases the time of the model so that it could be trained and evaluated. This work is done especially with high-dimensional data. When the dimensions are reduced, this helps in improving the performance and computational efficiency of the model. This helps them in enhancing and sprinting the training and testing process (Ruskanda *et al*., 2023).

* **Visual Insights:**

The PCA helps us to understand how the data is distributed in the hyperplane by depicting the data in a graphical representation. This graphical representation also shows how the decision boundaries separate the data among themselves. By using the PCA, we can reduce the data in two-dimensional and three-dimensional spaces, thereby simplifying the data representation. This helps in a clear understanding of the separated data classes. This graphical representation helps in identifying the data boundaries, and this also helps in making the prediction of the model.

In this tutorial, we will use the principal component analysis so that the dimension of the Iris dataset can be reduced from a 4-dimensional feature space to a two-dimensional feature space. According to our convenience, we will reduce it to a 2-dimensional feature space so that the visual representation is easy (Al-Mejibli *et al*., 2023).

# Dataset: Iris Dataset

There is an excellent data set available, which is the Iris dataset. This iris dataset contains a total of 150 flowers from three different species of Iris. These species include Setosa, Virginica, and Versicolor. The features of this dataset are:

* Sepal Lenght.
* Sepal width.
* Petal length.
* Petal width.

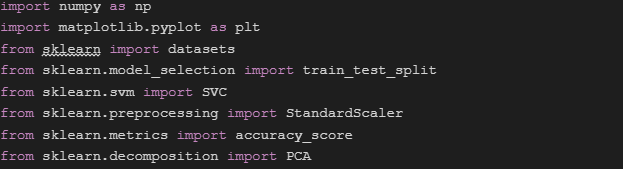
The data points that are there in the Iris dataset is connected with one of the three species defined here. This dataset is a milestone in the algorithm of machine learning and is widely used for the classification tasks. Furthermore, it is to be noted that is SVM also included in this dataset.

The samples depicted below are labeled with one of these different species:

## Code Walkthrough

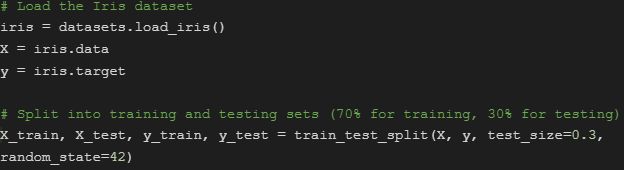
### Step 1: Importing the necessary Libraries

In the first step, we will start by importing the necessary libraries from the Python library; which is mainly done for PCA and SVM visualization.



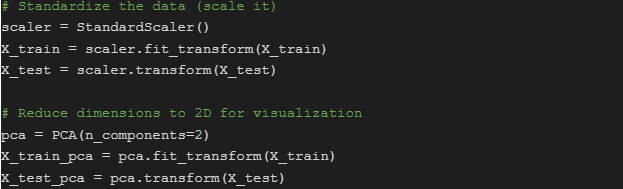
### Step 2: Loading and Preparing the dataset

Now it is time to load the iris dataset into the code, and then we will split this dataset into training sets and testing sets. In this step, we first load the dataset using the scikit-learn’s function, which is datasets.load\_iris(). Then, in the next step, we split the dataset into two parts: one is for testing and another for learning. At the computation, it is assumed that 70% of the total data is used for learning purposes, and the rest 30% data is used for testing purposes.



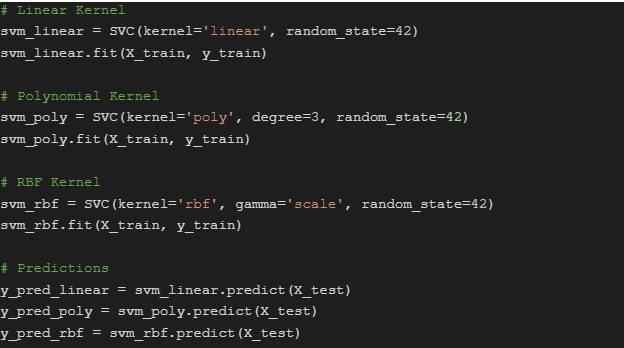
### Step 3: We Used PCA for Dimensionality Reduction

In step 3, we will use PCA in order to reduce the dimensionality, which means reducing it from 4 dimensional to 2-dimensional. This makes graphs for decision boundaries easier (Gomiasti *et al*., 2024).



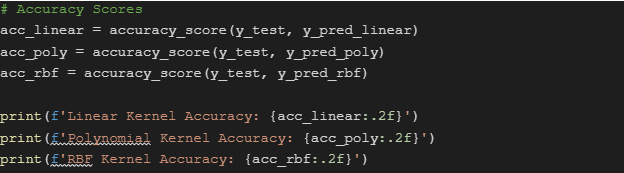
### Step 4: Training SVM Models

This step includes training the SVM model by using three different kernels: polynomial, linear, and RBF. The training of these three different SVM models is done using the training data, which is rather split from the PCA.



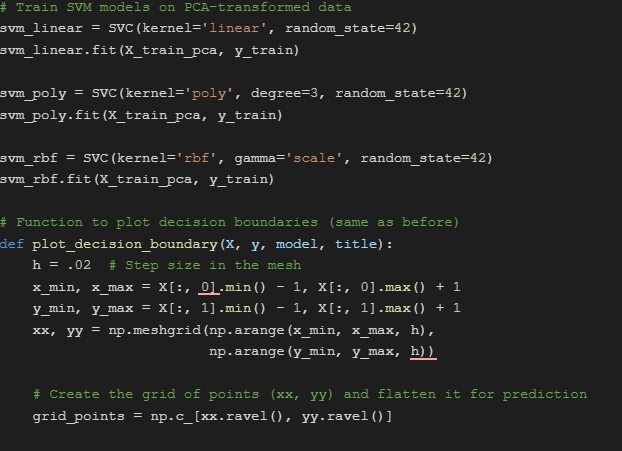
### Step 5: Visualizing Decision Boundaries

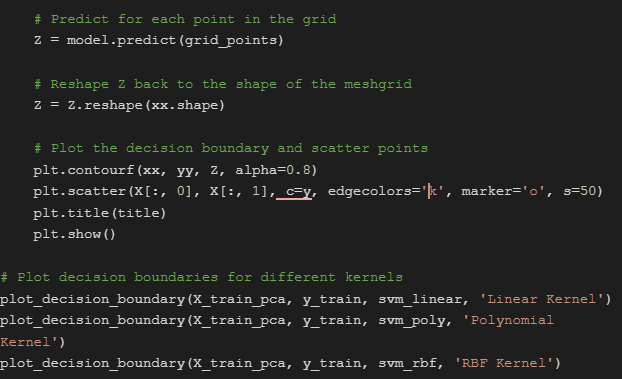
In order to visualize the decision boundaries, we will define a function to plot each of these three SVM models. The decision boundaries we will get will be plotted on the 2-dimensional feature space after PCA reduces the dimension of the datasets from 4 dimensions to 2 dimensions (Schuld 2021).



### Step 6: Evaluating the Model Performance

After seeing the visual representation of the code and training, now it is time to evaluate the performance of each model defined here. In order to calculate model performance, comparing their accuracy with the true labels will be very helpful:





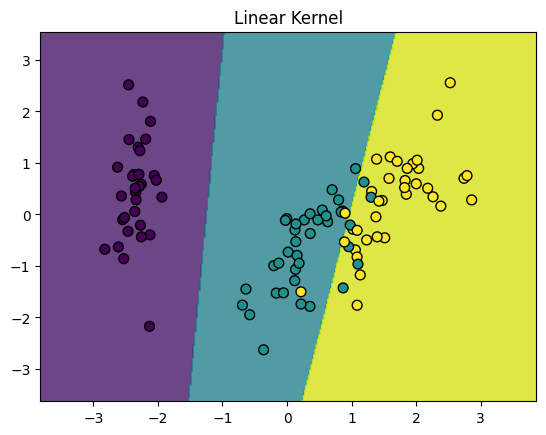
# Discussion Results

## Decision Boundaries

From the decision boundaries we got from comparing the three different kernels, we gained valuable insights from them about how the SVM adapts to the structure of the data (Thurnhofer-Hemsi *et al*., 2020).

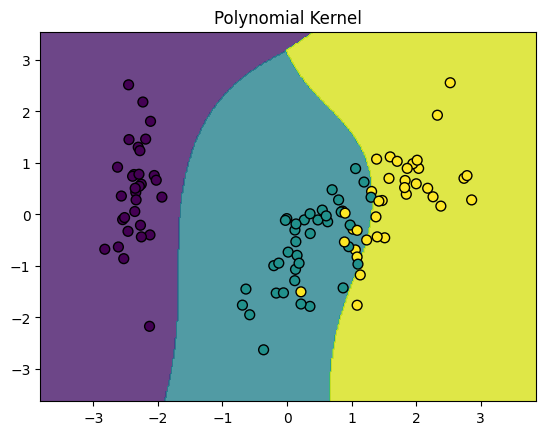
1. **Linear Kernel:**

The linear kernel is mainly used for simpler data sets or when someone does not want to add unnecessary complications to the code. This kernel has some limitations, like this will work only with the datasets that are linearly separable. The linear kernel struggles with some more complex data distributions.



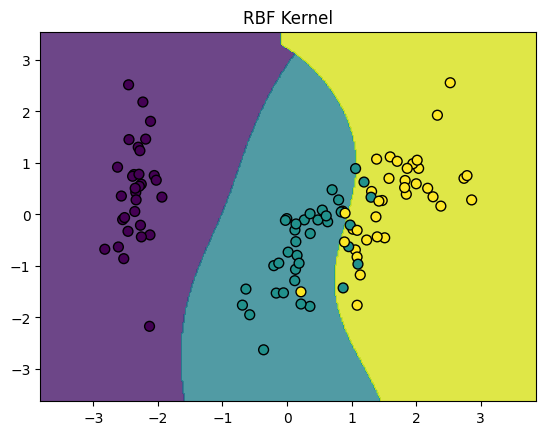
1. **Polynomial Kernel**

This particular kernel trick allows the algorithm to operate in high-dimensional feature spaces without explicitly computing the coordinates of the data in the space. This kernel is used to create decision boundaries that are curved. This curved type of boundary offers more flexibility at the time of handling nonlinear linear relationships between the classes.



1. **RBF Kernel**

Among these three, the most flexible offered by any kernel is this RBF kernel which is effective at the time of generating highly accurate nonlinear decision boundaries. The boundaries made with this RBF kernel are able to effectively separate complex data distributions (Aldino *et al*., 2021).



## Accuracy Evaluation

The accuracy evaluation metrics generally provide the accuracy of each kernel on each test set, which concludes a brief of how each data point depicts the unseen data. Among the three, the RBF kernel works the best when dealing with complex datasets and provides the best visualization of the datasets (Eskandari *et al*., 2020). The polynomial kernel is able to deal with nonlinear data, but when it comes to very complex datasets, the accuracy level drops in the polynomial kernel. On the other hand, the linear kernel only works with data that are linearly separated, and it is the easiest one among the three.

# Conclusion

For the classification tasks, the Support Vector Machine (SVM) powered with three different kernels is a powerful tool. The SVM model with the kernel is able to handle both linear and nonlinear data. From handling very basic datasets to very complex datasets, SVM can handle most of the datasets. In order to handle relatively less complex datasets, we have the linear kernel which handles only the linearly separable datasets. However, for mediocre nonlinear datasets, we have the polynomial kernel; and for the most complex flexible nonlinear datasets, we have the RBF kernel. In order to control the dimensionality reduction, we have the PCA, which, for simplification, can reduce the dimension from 4 dimensional to 2 dimensional feature space. By doing the dimensionality reduction, we can envision the decision boundaries, which shall help us in better understanding the working of the SVM model.

# References

Tariq, A., Jiango, Y., Li, Q., Gao, J., Lu, L., Soufan, W., ... & Habib-ur-Rahman, M. (2023). Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data. *Heliyon*, *9*(2).

Mustaqeem, M., & Saqib, M. (2021). Principal component based support vector machine (PC-SVM): a hybrid technique for software defect detection. *Cluster Computing*, *24*(3), 2581-2595.

Singgalen, Y. A. (2023). Comparative analysis of decision tree and support vector machine algorithm in sentiment classification for birds of paradise content. *International Journal of Basic and Applied Science*, *12*(3), 100-109.

Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, *233*, 109126.

Al-Mejibli, I. S., Alwan, J. K., & Abd, D. H. (2020). The effect of gamma value on support vector machine performance with different kernels. *Int. J. Electr. Comput. Eng*, *10*(5), 5497-5506.

Schuld, M. (2021). Supervised quantum machine learning models are kernel methods. *arXiv preprint arXiv:2101.11020*.

Thurnhofer-Hemsi, K., López-Rubio, E., Molina-Cabello, M. A., & Najarian, K. (2020). Radial basis function kernel optimization for support vector machine classifiers. *arXiv preprint arXiv:2007.08233*.

Aldino, A. A., Saputra, A., Nurkholis, A., & Setiawansyah, S. (2021). Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur. *Building of Informatics, Technology and Science (BITS)*, *3*(3), 325-330.

Eskandari, A., Milimonfared, J., Aghaei, M., & Reinders, A. H. (2020). Autonomous monitoring of line-to-line faults in photovoltaic systems by feature selection and parameter optimization of support vector machine using genetic algorithms. *Applied Sciences*, *10*(16), 5527.

Gomiasti, F. S., Warto, W., Kartikadarma, E., & Gondohanindijo, J. (2024). Enhancing Lung Cancer Classification Effectiveness Through Hyperparameter-Tuned Support Vector Machine. *Journal of Computing Theories and Applications*, *1*(4), 396-406.

Ruskanda, F. Z., Abiwardani, M. R., Mulyawan, R., Syafalni, I., & Larasati, H. T. (2023). Quantum-enhanced support vector machine for sentiment classification. *IEEE Access*, *11*, 87520-87532.