Iris Setosa SVM Assignment

Dexter Corley

Argyros School of Business & Economics

&

Fowler School of Engineering

CPSC 393 - 01

Dr. Nicholas LaHaye

February 21, 2023

Abstract

The goal of this assignment is to use the given Iris-Setosa dataset containing information on classifying Iris-setosa plants to train a support vector machine. A successful outcome involves the model being able to properly classify the testing dataset upon whether each given point is an Iris-setosa or not. This report details my solution and highlights the path I took to get there including written and visual aids.

Introduction

A support vector machine is a unique classification algorithm that can be used to differentiate different categories by using a decision boundary. The SVM classification algorithm needs to satisfy three things to be successful. Optimizing the hyperplane, seeking a large margin, and using the correct kernel for the features. Essentially, to optimize the hyperplane, there needs to be minimal errors and you get there through choosing the correct kernel. Errors refer to points that pass through the margin lines and get close to the hyperplane. In the case of this dataset, the two categories needing classification are Iris Setosa plants and Non-Iris Setosa plants. Apart from the field containing Iris-Setosa & Non-Iris Setosa, the dataset contains: Id, Petal Length and Width, and Sepal Length and width. These are the columns to be used as predictors for the model. The approach taken for cleaning the data in model prep was simple. The first step was to change the iris-setosa field from a string containing “Iris-setosa” and “Not-iris-setosa” to a binary 1, 0. The other step needing to take place was to drop the ID column as it’s not an important predictor.

Body

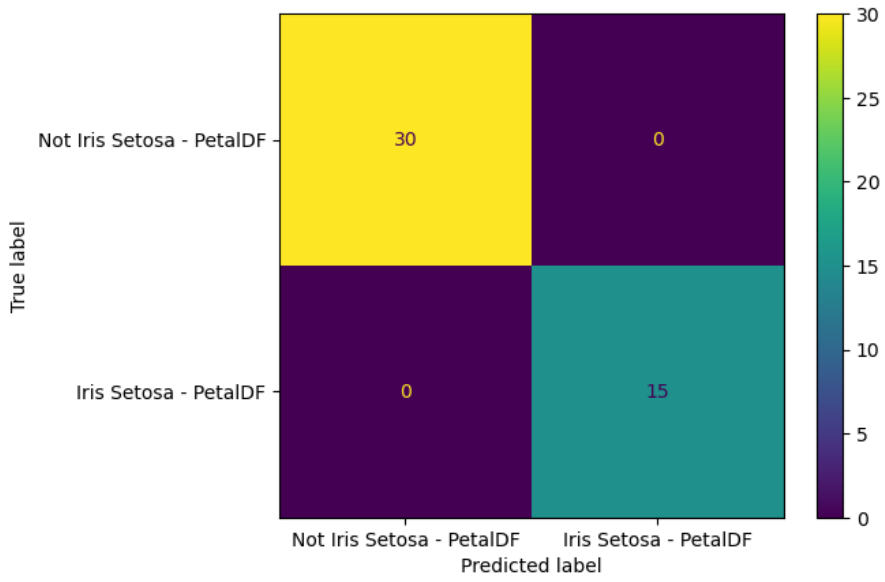
Running the model for this dataset took a couple of iterations to get it right. Originally, the model was run without separating the predictors and was getting near perfect results. After putting together a graph to see the hyperplane, it turned out things weren’t as good as they seemed. The hyperplane wasn’t fit to the data well and the model needed to be re-done. Going through for a second time, two separate data frames were copied out from the original dataset. After dropping the ID column and changing the Iris-setosa column to a binary variable, the two

datasets contained the Iris-setosa (1/0) column as well as petal data for one and sepal data for the other. This allowed the model to have access to data that it can properly classify.

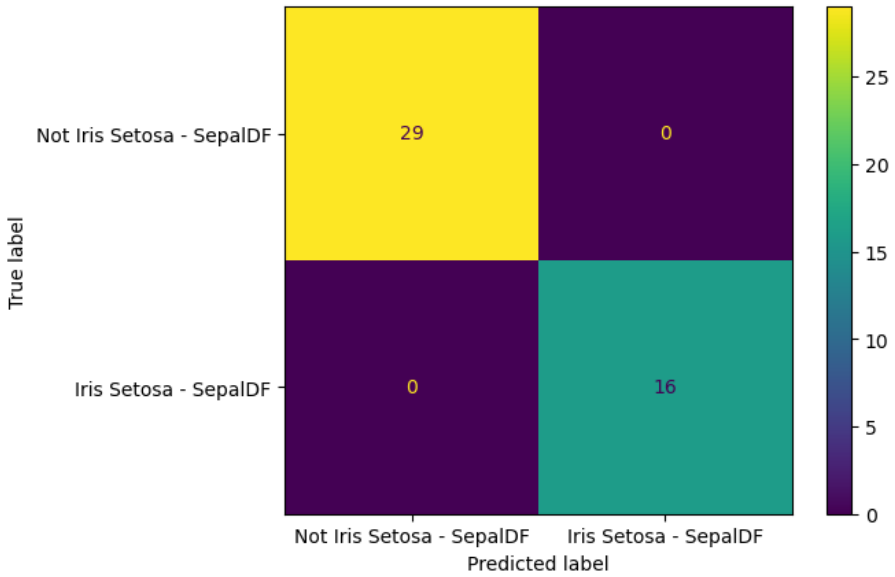
From these two sets resulted in two models which both performed excellent. The confusion matrix for predicted test results correctly classified all datapoints. The hyperplane for each model maximized the distance between each margin and only contained a few slack points. This resulted in two successful SVM models that were very accurately classifying unseen data points.

Performance Experiments

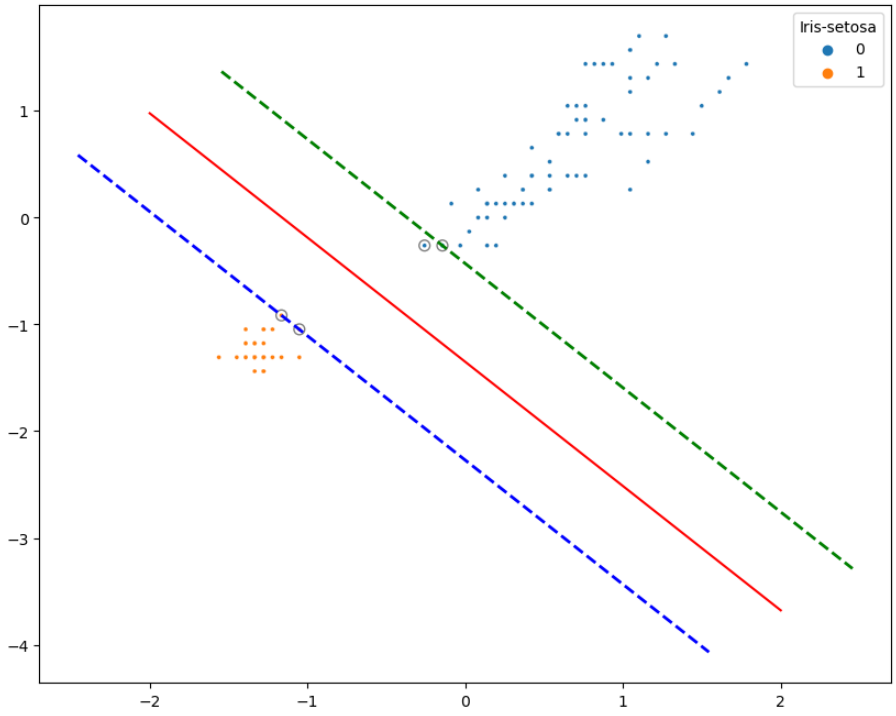
Confusion Matrix for Iris-Setosa results using the Petal Size data-frame



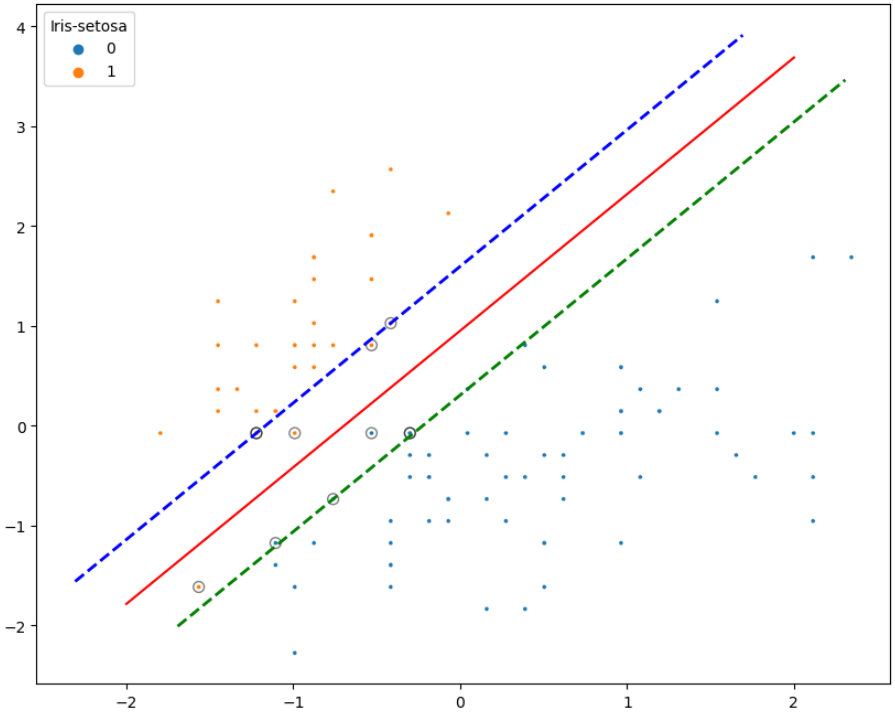
Confusion Matrix for Iris-Setosa results using the Sepal Size data-frame



SVM Model with Hyperplane and Margins for Iris-setosa results using the Petal Size data-frame



SVM Model with Hyperplane and Margins for Iris-setosa results using the Sepal Size data-frame



Conclusions

In conclusion, to result with a successful SVM implementation from using the given dataset, we had to run two separate models. The Petal Size dataset as well as Sepal Size datasets became good predictor variables for classifying the difference in Iris Setosa vs non-Iris Setosa plants. Additionally, by using hyperplane and margin visualization the model's success is clearly visualized to show any slack and where the decision boundary lies.

Citations

Kataev, Alex. “SVM Classification with Sklearn.svm.SVC: How to Plot a Decision Boundary with Margins in 2d Space.” Medium, Geek Culture, 19 Aug. 2021, https://medium.com/geekculture/svm-classification-with-sklearn-svm-svc-how-to-plot-a-decision-boundary-with-margins-in-2d-space-7232cb3962c0.

Starmer, Josh. “Support Vector Machines in Python from Start to Finish.” YouTube, YouTube, 30 June 2020, https://www.youtube.com/watch?v=8A7L0GsBiLQ.