

Introduction

In this project, the objective is to analyze capabilities of SVC in classification of noisy moon datasets.

About Dataset

The dataset is created by `make_moons` module of `sklearn`. A specific amount of noise is added to the data and then they are saved as `.csv` file. The data points are labeled in two classes.

Dataset1.csv contains moons data of 500 points and Dataset2.csv has moon data of 1200 points. A preview of dataset1.csv is as following picture.

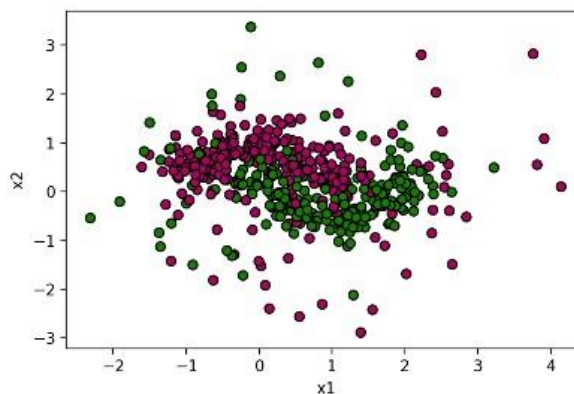


Fig. 1: Moon data points in Dataset 1.

Algorithm Implementation

SVC classifier is applied on the dataset and its ability in classifying is analyzed for different values of hyper parameters. In this project, value of C is chosen from $\{0.001, 0.01, 0.1, 1, 10, 100, 1000, 10e4, 10e5, 10e6\}$ set and Γ is chosen from $\{0.01, 0.1, 1, 10, 100, 1000\}$ set. The results of analysis are plotted in several figures.

Preprocessing

No preprocessing is needed in this project.

Project Outline

The report is outlined in 3 parts:

- ✚ Part 1: Analyzing the effect of C
- ✚ Part 2: Analyzing the effect of Γ
- ✚ Part 3: Conclusion
- ✚ Appendix A: Modules used in coding

Part 1: Analyzing the effect of C

In this part, the effect of different values of C parameter on the capability of SVC classifier is analyzed and the results of its behavior is depicted in the following plots. It should be mentioned that Dataset1.csv is used here. User can choose Dataset2.csv to investigate hyper parameter effects. In addition, plot of Accuracy Vs C values are plot afterwise.

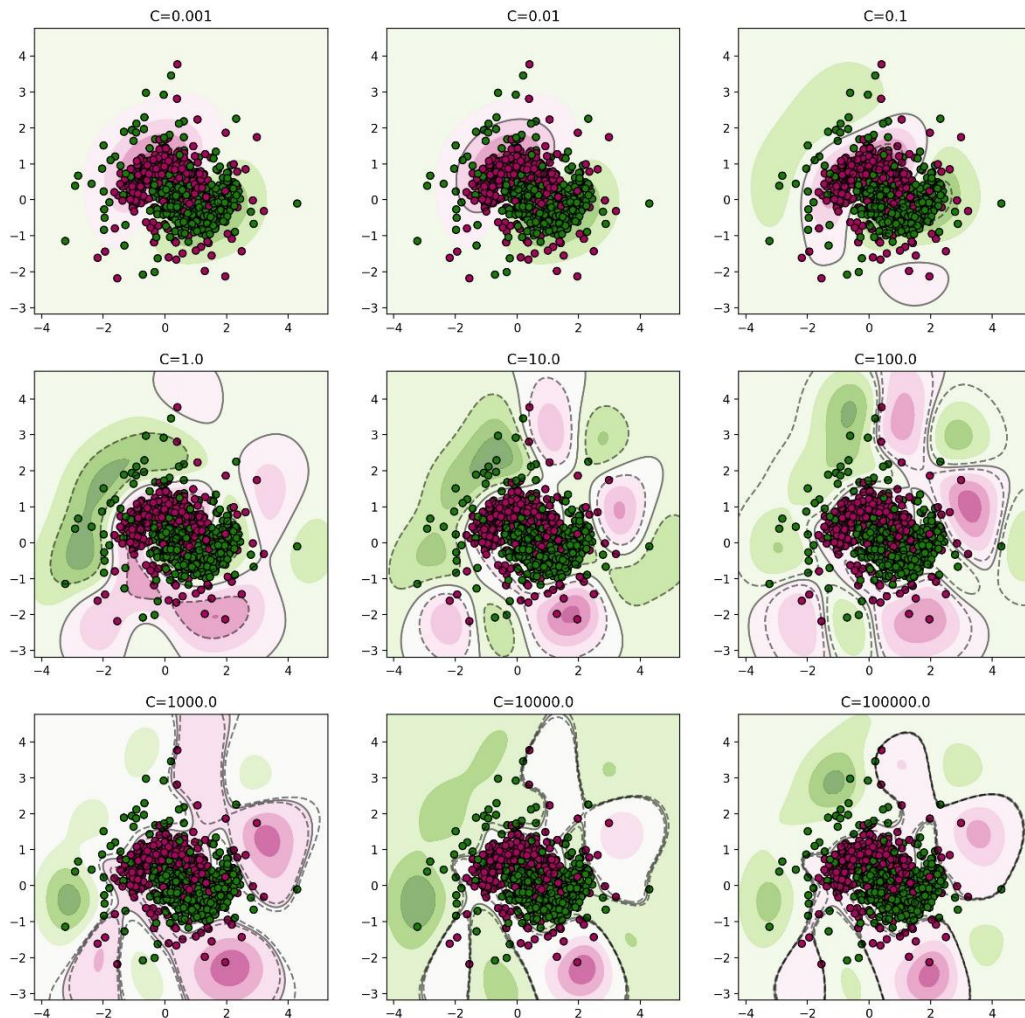


Fig 2: Effect of C on the behavior of SVC.

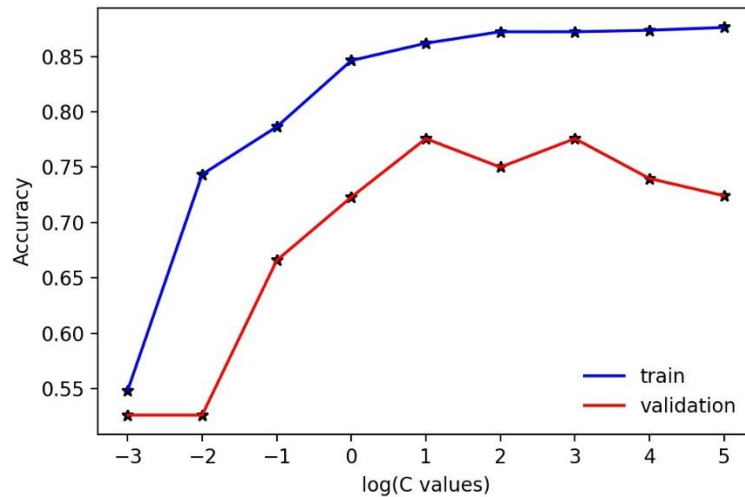


Fig 3: Train and Validation Accuracy Vs C values.

Part 2: Analyzing the effect of Gamma

In this part, the effect of different values of C parameter on the capability of SVC classifier is analyzed and the results of its behavior is depicted in the following plots. It should be mentioned that Dataset1.csv is used here. User can choose Dataset2.csv to investigate hyper parameter effects. In addition, plot of Accuracy Vs Gamma values are plot afterwise.

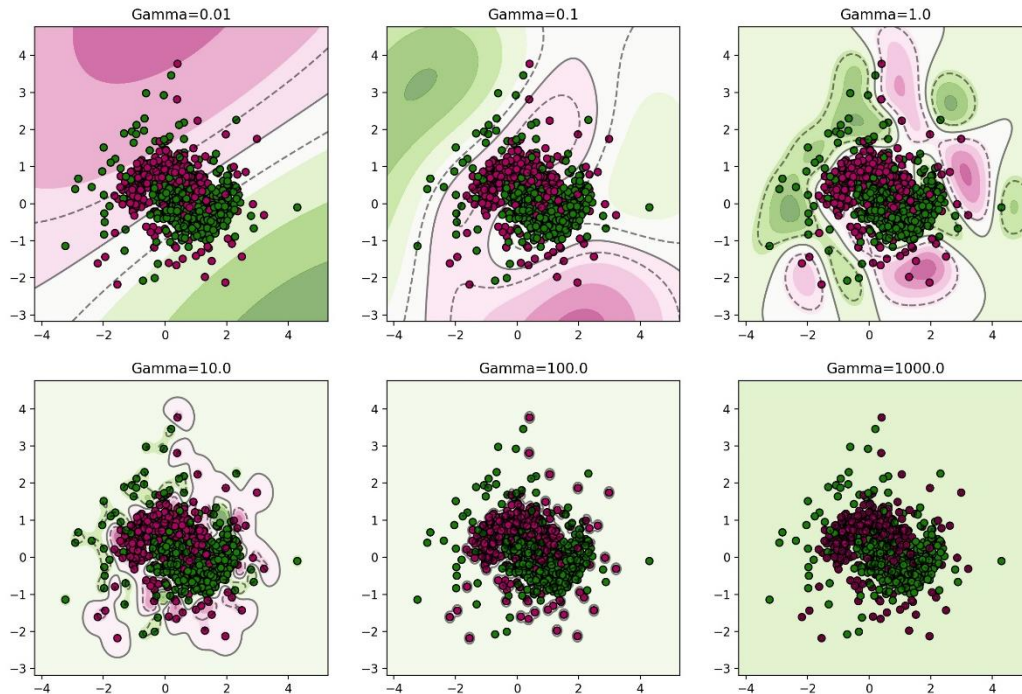


Fig 4: Effect of Gamma on the behavior of SVC.

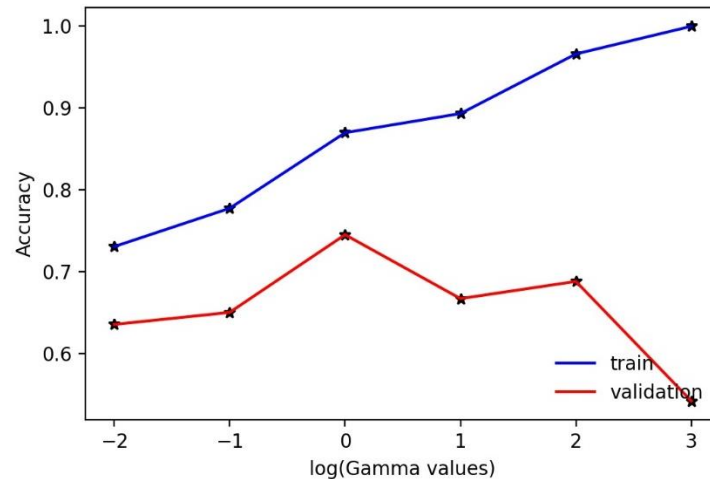


Fig 5: Train and Validation Accuracy Vs Gamma values.

Part 3: Conclusion

By analyzing the results, following points are observed:

1. SVM classifier works pretty nice on small datasets.
2. After a specific value of C or Gamma, train error increases while validation decreases. Therefore, a compromise must be made in implementation of SVC.

Appendix A: Modules used in coding

The following Modules are used in python codes:

- **Threading:** used of parallel processing to decrease run time.
- **Class:** is created and used for better programming quality.
- **Logging:** an enhanced logger file is created to monitor the progression of codes and log any messages emitted from process. The results of run are logged in the file too in a pretty manner.
- **Argparse:** to read the data input by user.
- **Plots** are saved in a separate specific folder for ease of access and analysis.