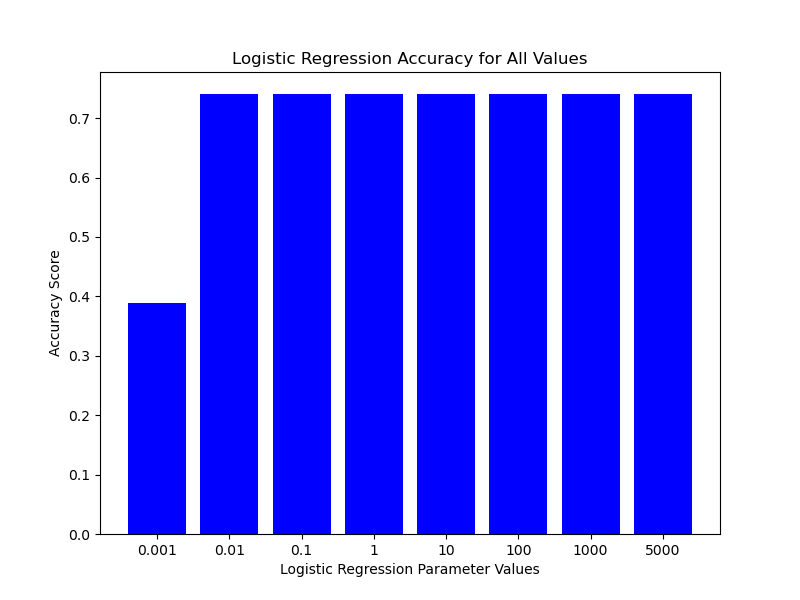
Predictor Comparison Report

# Logistic Regression

Logistic Regression was tested with various regularization strengths (C values). As the C value increases, the model tends to overfit, creating tighter decision boundaries. For smaller C values, the model becomes more regularized and generalizes better.

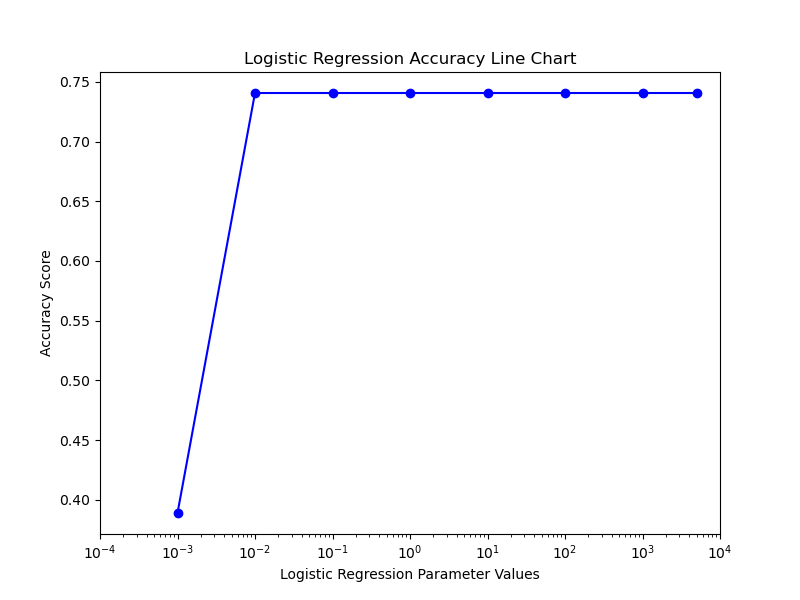
## Logistic Regression Accuracy

The chart below shows the accuracy for all C values in Logistic Regression.



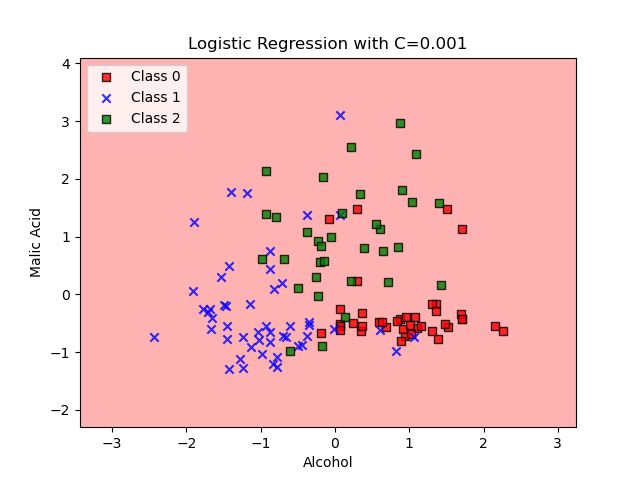
## Logistic Regression Accuracy Line Chart

The line chart below shows the accuracy across different C values in Logistic Regression.



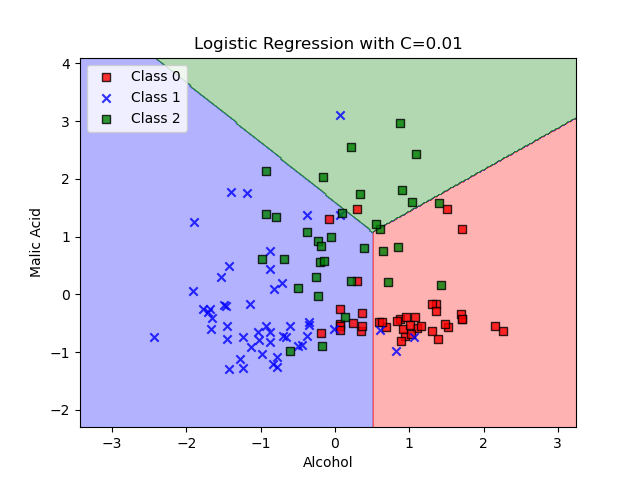
## Logistic Regression with C=0.001

The decision boundary for Logistic Regression with C=0.001 is shown below.



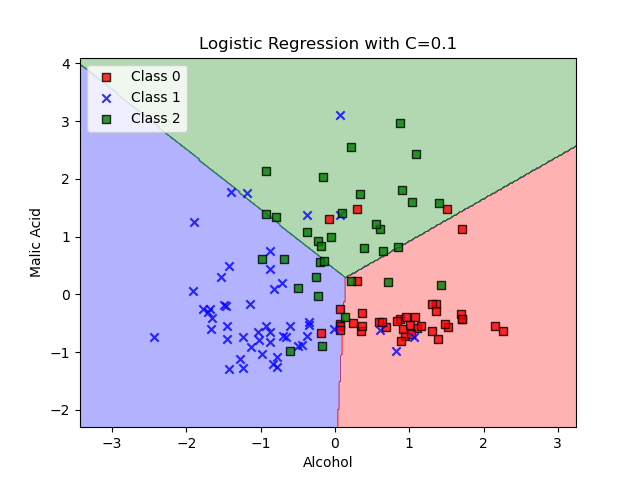
## Logistic Regression with C=0.01

The decision boundary for Logistic Regression with C=0.01 is shown below.



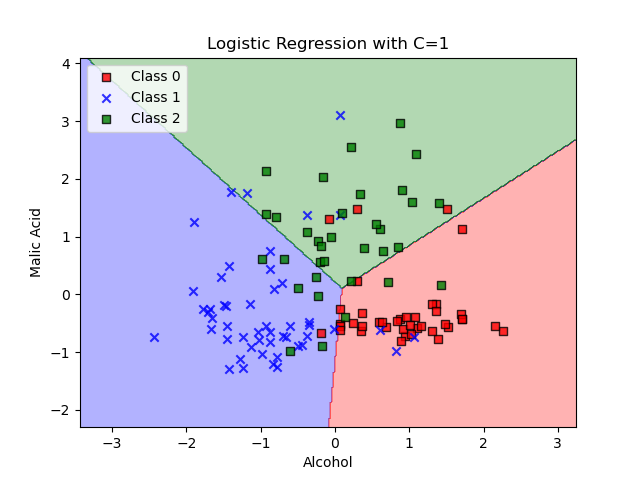
## Logistic Regression with C=0.1

The decision boundary for Logistic Regression with C=0.1 is shown below.



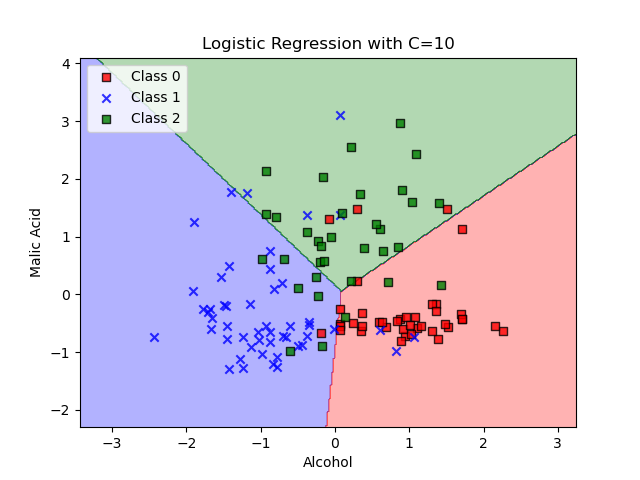
## Logistic Regression with C=1

The decision boundary for Logistic Regression with C=1 is shown below.



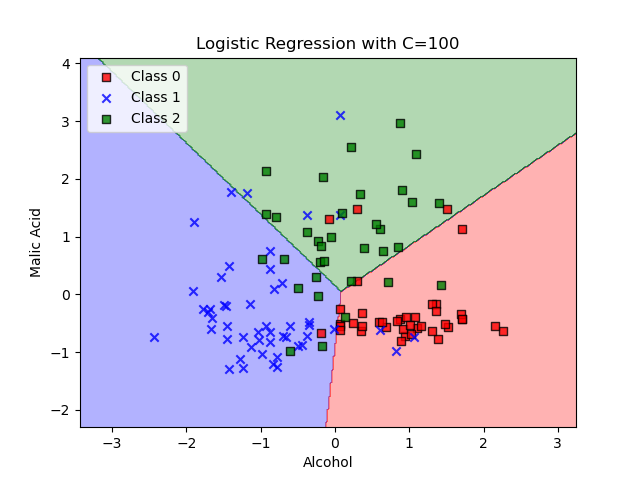
## Logistic Regression with C=10

The decision boundary for Logistic Regression with C=10 is shown below.



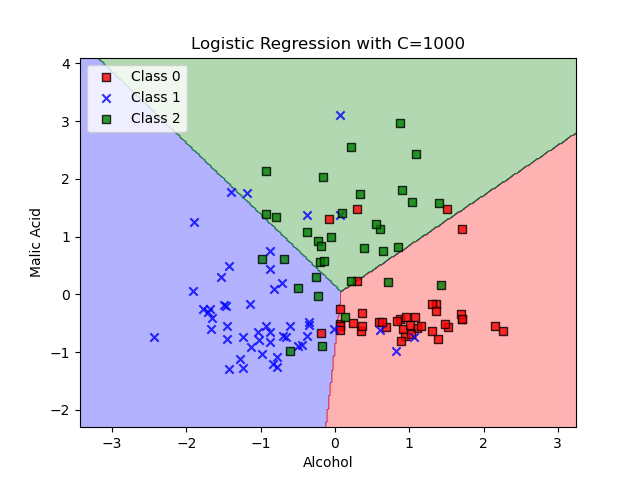
## Logistic Regression with C=100

The decision boundary for Logistic Regression with C=100 is shown below.



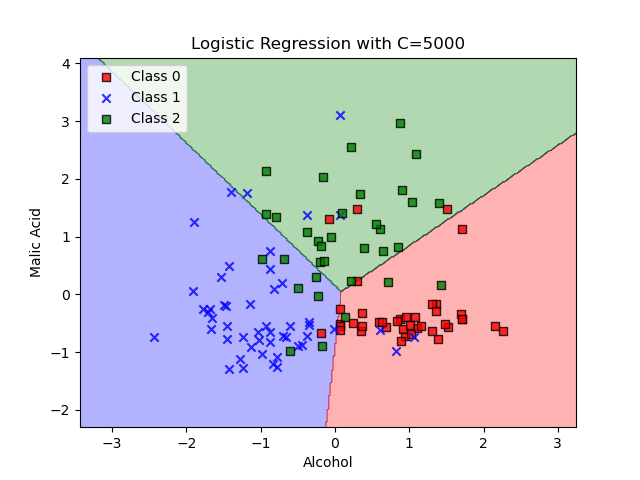
## Logistic Regression with C=1000

The decision boundary for Logistic Regression with C=1000 is shown below.



## Logistic Regression with C=5000

The decision boundary for Logistic Regression with C=5000 is shown below.

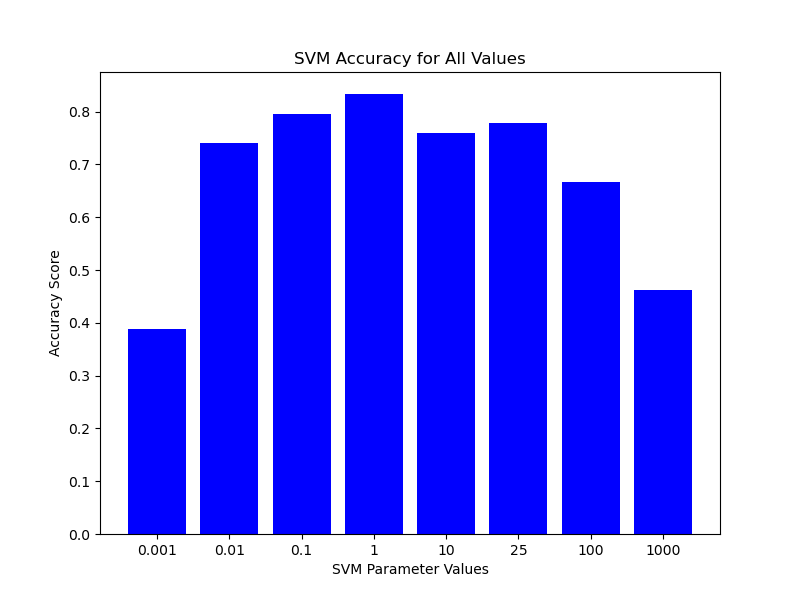


# SVM with RBF Kernel

Support Vector Machines (SVM) with an RBF kernel were tested with different gamma values. The gamma parameter controls the width of the RBF kernel. A lower gamma value results in a smoother decision boundary, while a higher gamma value makes the decision boundary more complex.

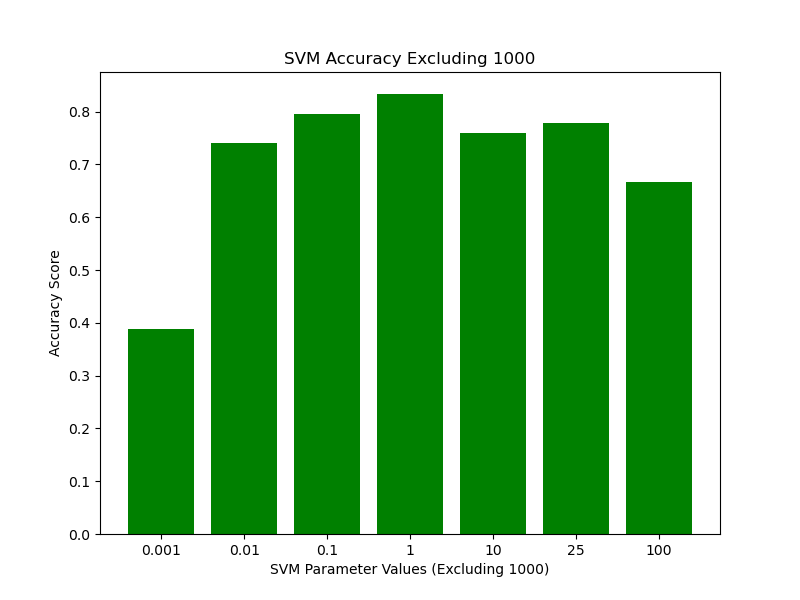
## SVM Accuracy (All Values)

The chart below shows the accuracy for all gamma values in SVM.



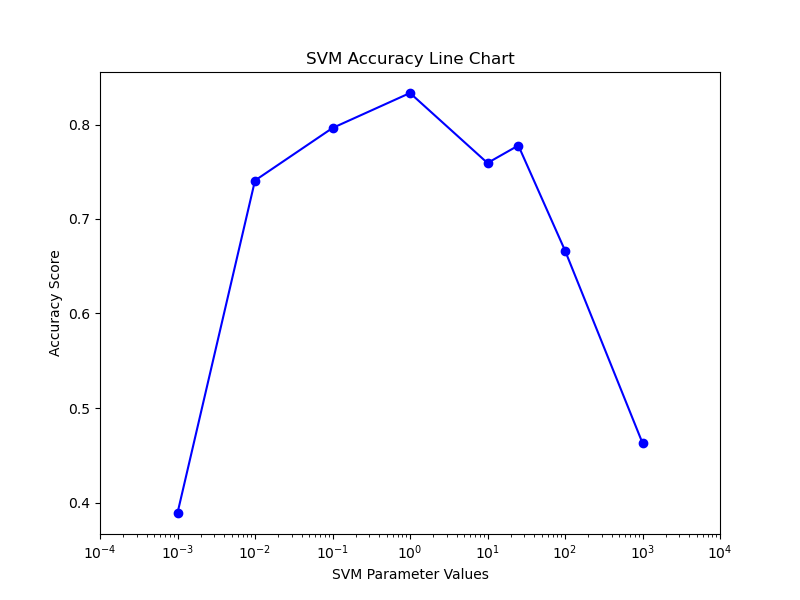
## SVM Accuracy (Excluding Gamma=100)

The chart below shows the accuracy for gamma values in SVM, excluding gamma=100.



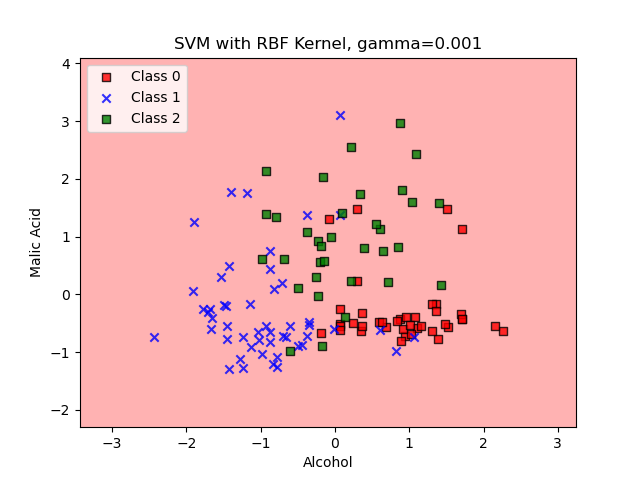
## SVM Accuracy Line Chart

The line chart below shows the accuracy across different gamma values in SVM.



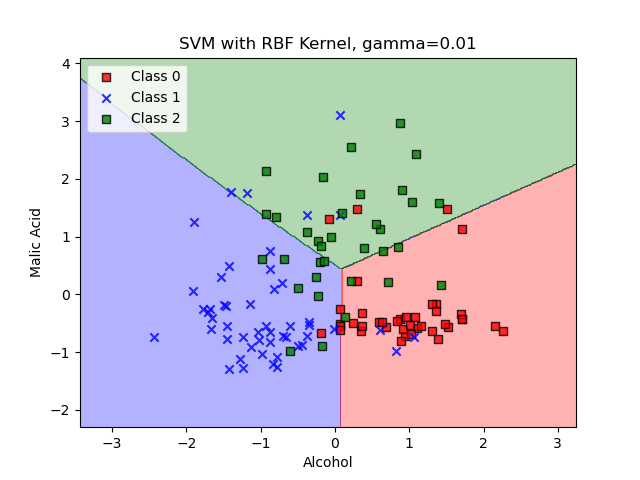
## SVM with RBF Kernel, gamma=0.001

The decision boundary for SVM with gamma=0.001 is shown below.



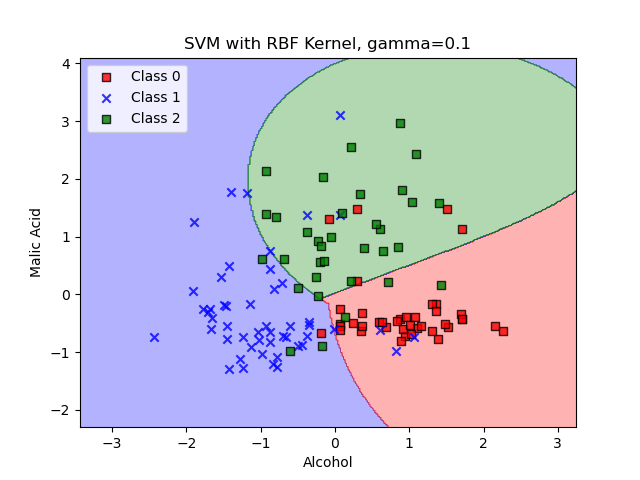
## SVM with RBF Kernel, gamma=0.01

The decision boundary for SVM with gamma=0.01 is shown below.



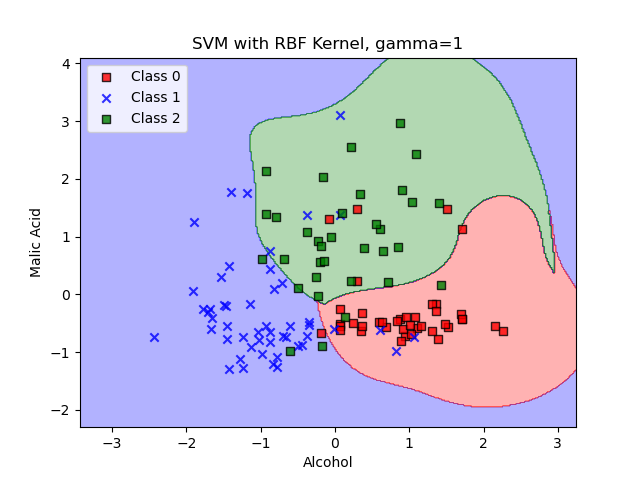
## SVM with RBF Kernel, gamma=0.1

The decision boundary for SVM with gamma=0.1 is shown below.



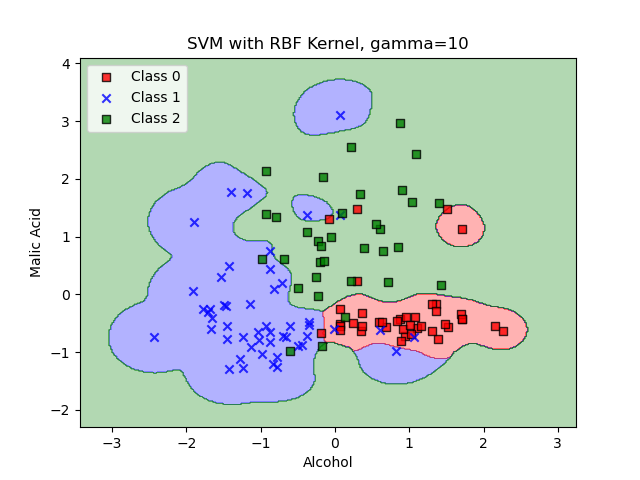
## SVM with RBF Kernel, gamma=1

The decision boundary for SVM with gamma=1 is shown below.



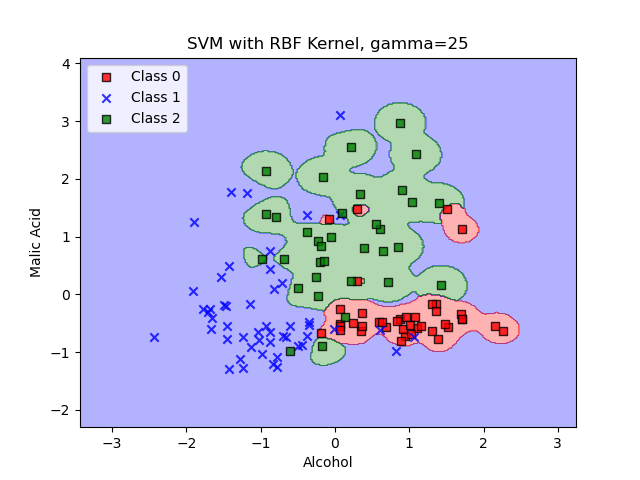
## SVM with RBF Kernel, gamma=10

The decision boundary for SVM with gamma=10 is shown below.



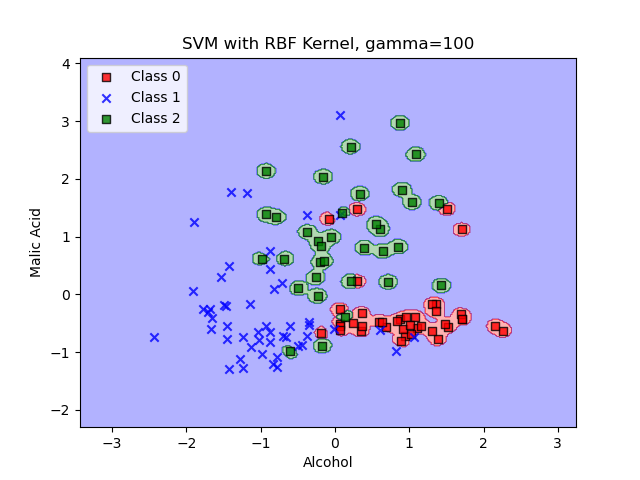
## SVM with RBF Kernel, gamma=25

The decision boundary for SVM with gamma=25 is shown below.



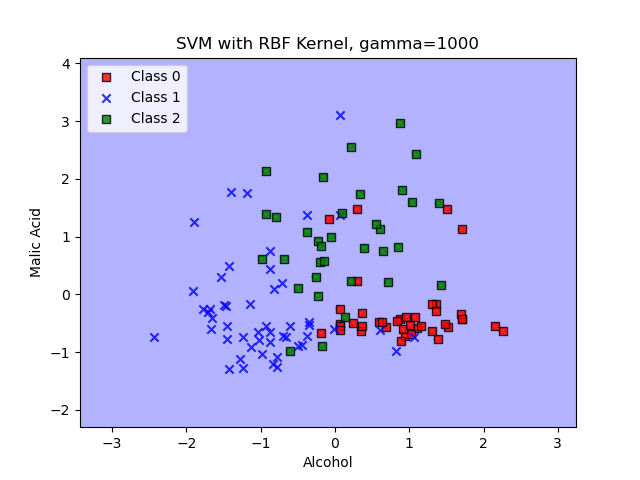
## SVM with RBF Kernel, gamma=100

The decision boundary for SVM with gamma=100 is shown below.



## SVM with RBF Kernel, gamma=1000

The decision boundary for SVM with gamma=1000 is shown below.



# Conclusion

## Logistic Regression:

Logistic Regression is a linear model that works well for simple, linearly separable datasets. It’s easy to understand and explain, as it generates straightforward decision boundaries. However, it struggles with non-linear data. The regularization parameter (C) controls model complexity: small values simplify the model, while large values increase complexity, possibly leading to overfitting. It is a great option when interpretability is important, especially for business or healthcare applications where understanding feature importance is key.

## Support Vector Machines (SVM):

SVM with the RBF kernel excels at capturing non-linear patterns in data, making it powerful for complex datasets. The gamma parameter controls how flexible the decision boundary is: low gamma produces smooth boundaries, while high gamma can result in overfitting by creating highly detailed boundaries around data points. SVM is highly effective for datasets where non-linear relationships are crucial, but it can be computationally expensive and harder to explain to non-technical stakeholders due to its complexity.

## Recommendation for Real-World Use:

For most real-world use cases, Logistic Regression is preferred when simplicity, speed, and interpretability are important, especially for explaining results to non-technical audiences. However, for complex problems with non-linear data, SVM with RBF kernel offers more flexibility, though it may be harder to explain and requires careful tuning to avoid overfitting.