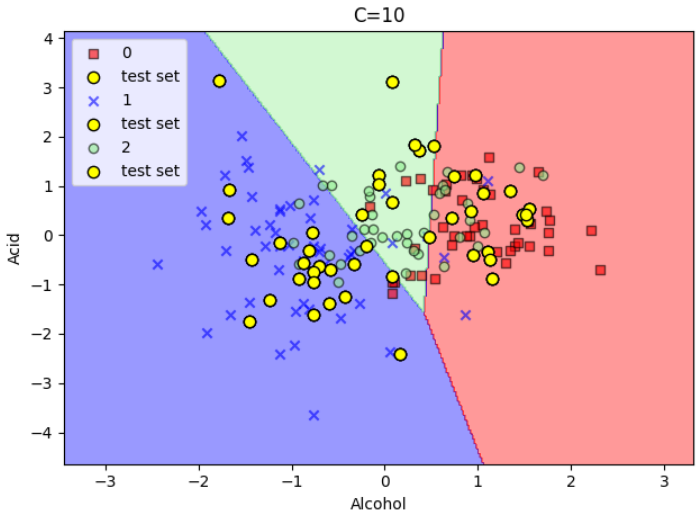
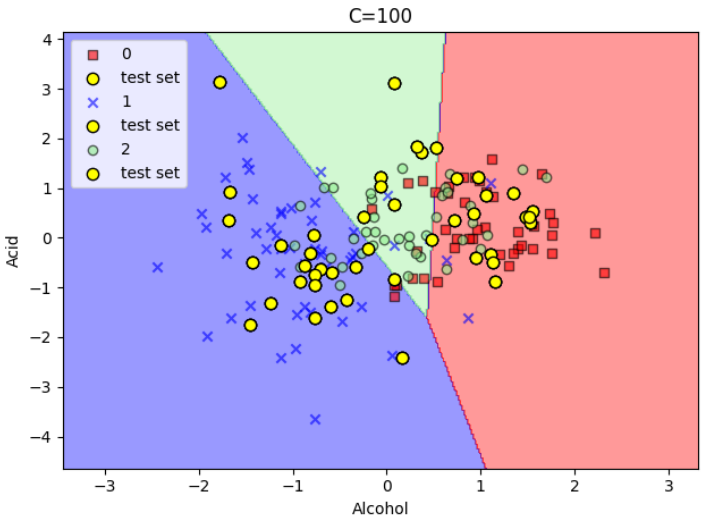
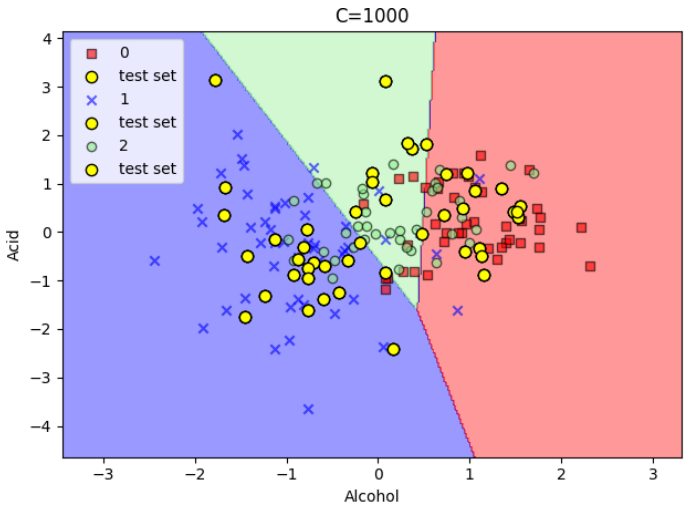
Small Project1 – Report

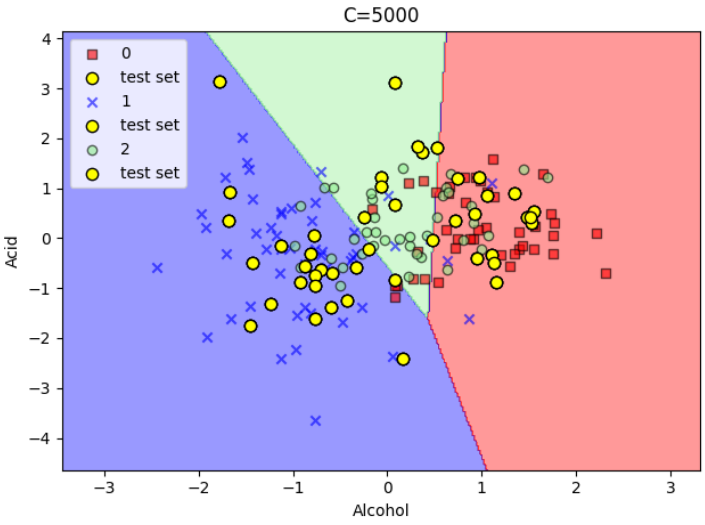
CS 4275: Machine Learning Foundations

Logistic Regression Graphs









Accuracies:

Accuracy: 0.41

Accuracy: 0.41

Accuracy: 0.72

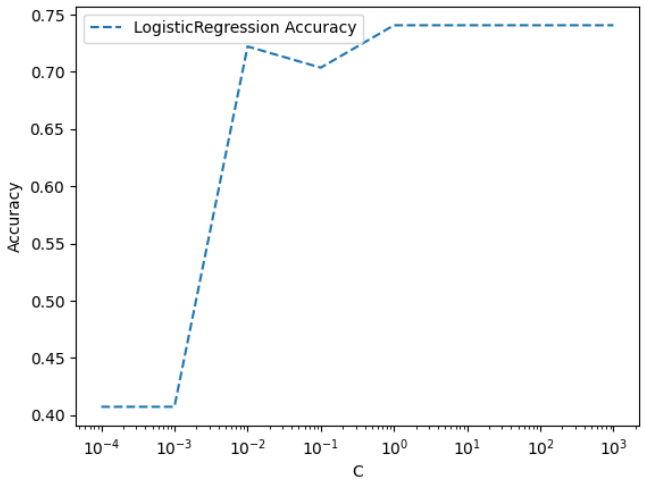
Accuracy: 0.70

Accuracy: 0.74

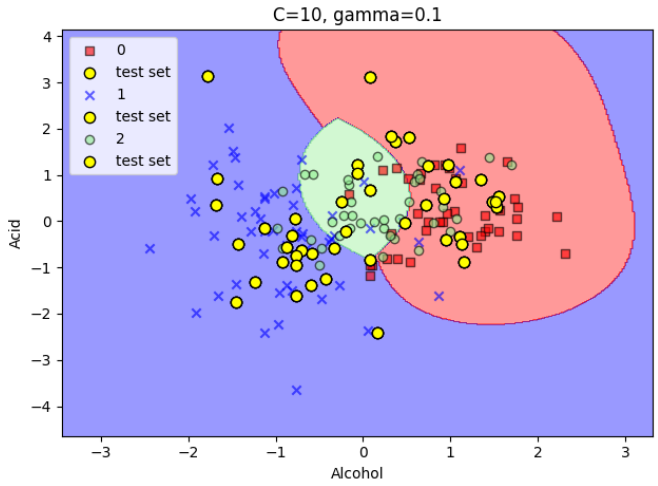
Accuracy: 0.74

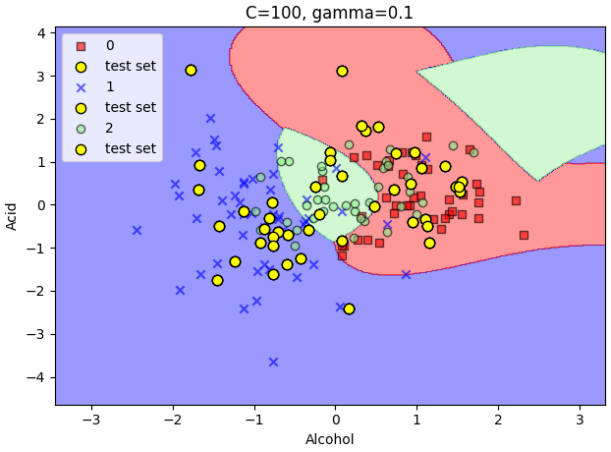
Accuracy: 0.74

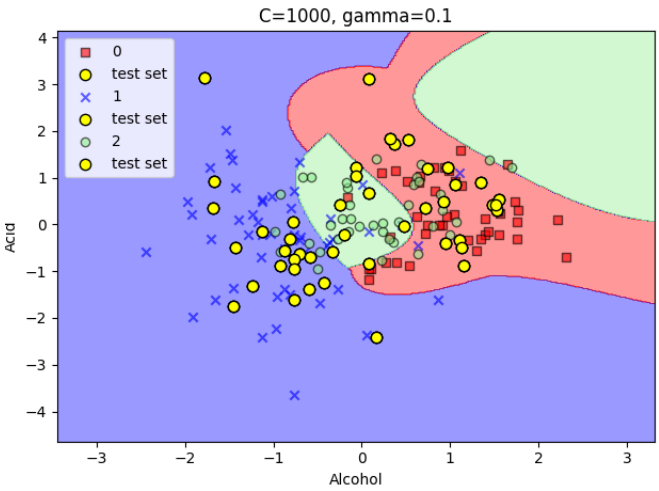
Accuracy: 0.74

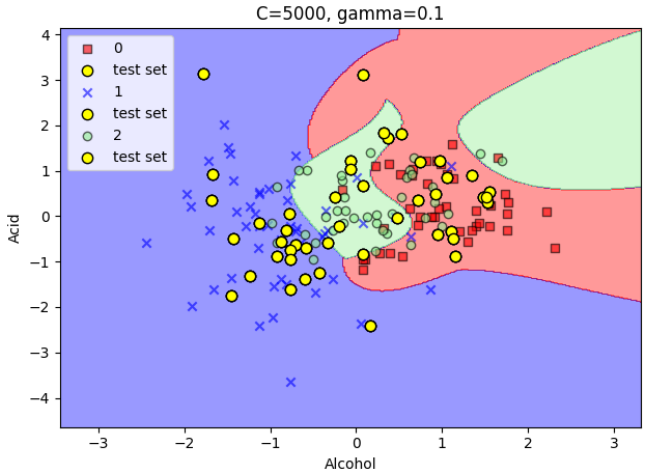


SVM: gamma 0.1:









Accuracies:

Accuracy: 0.41

Accuracy: 0.41

Accuracy: 0.41

Accuracy: 0.74

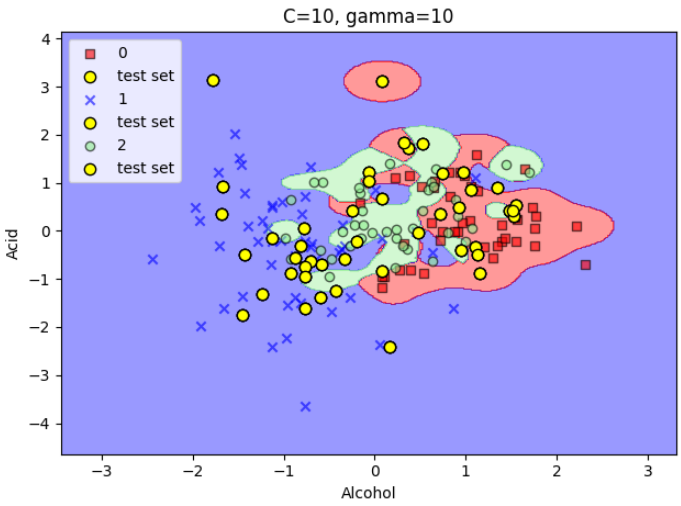
Accuracy: 0.80

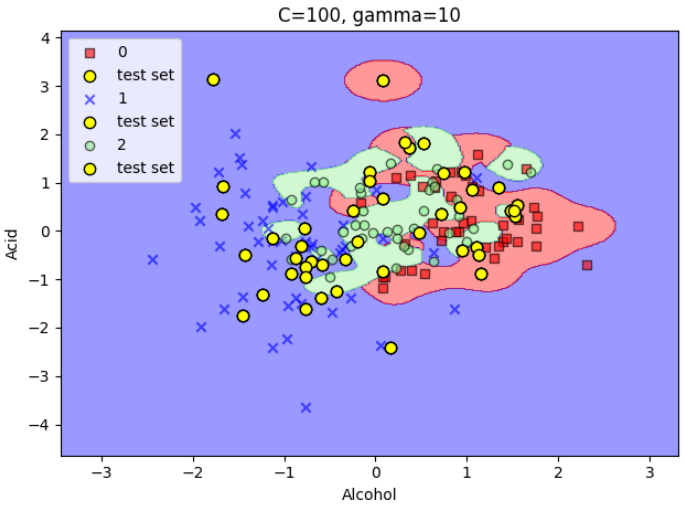
Accuracy: 0.78

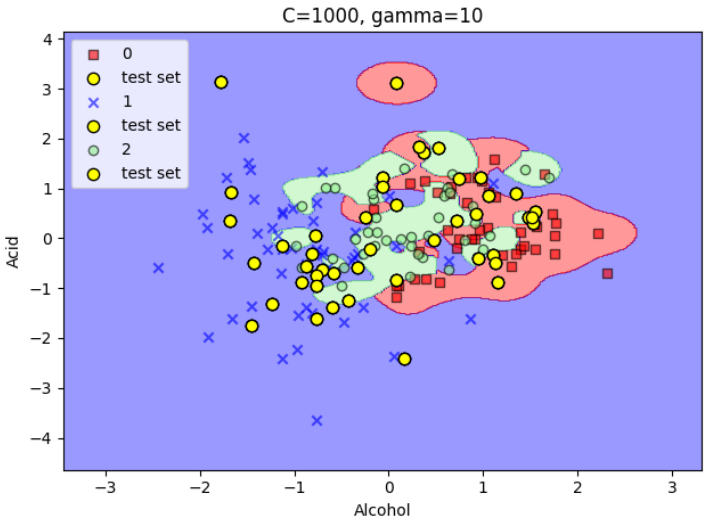
Accuracy: 0.78

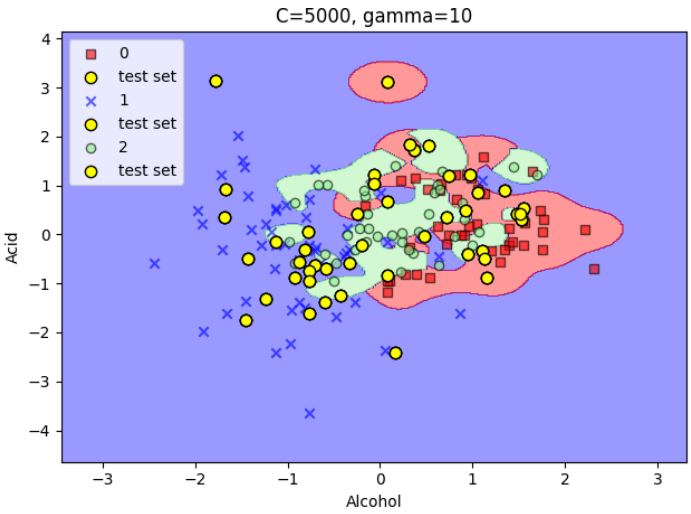
Accuracy: 0.76

SVM: gamma 10









Accuracies:

Accuracy: 0.41

Accuracy: 0.41

Accuracy: 0.41

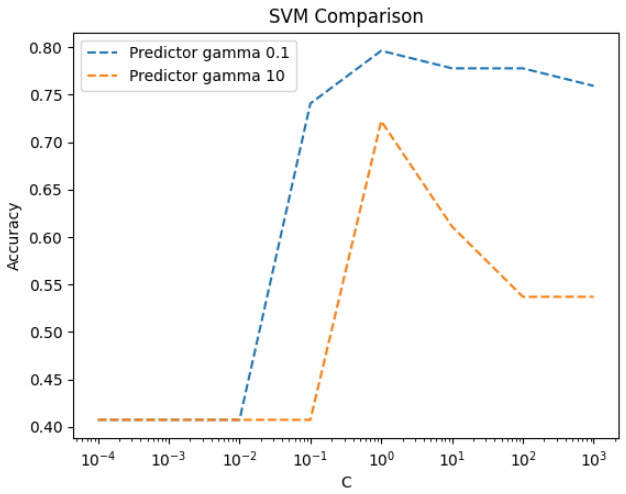
Accuracy: 0.41

Accuracy: 0.72

Accuracy: 0.61

Accuracy: 0.54

Accuracy: 0.54



Insights on Logistic Regression.

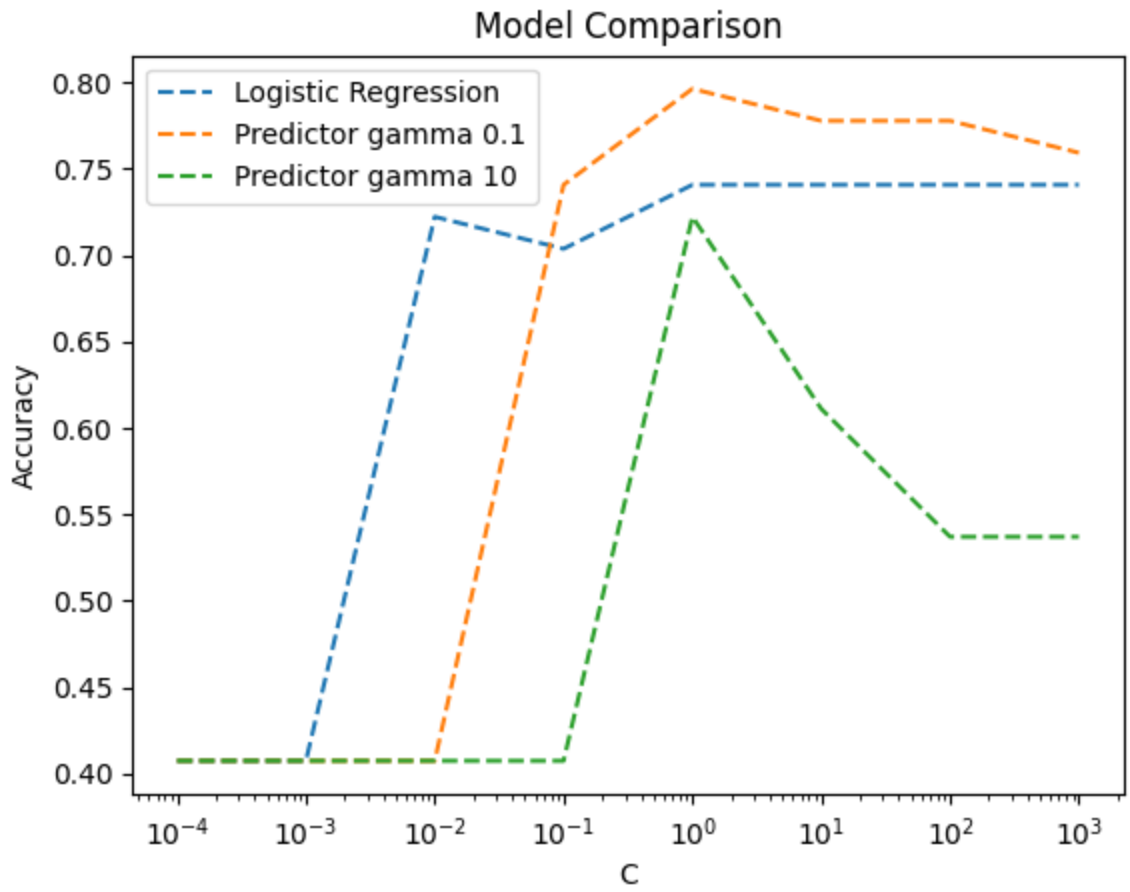
Changing the regularized parameter positively affects the accuracy score in our logistic regression model, as the accuracy score for the highest C value results in an accuracy score of 0.74. After reaching 0.74, accuracy stabilizes and doesn't improve with larger C values, which means there is a lower chance of overfitting. The highest accuracy, 0.74, could mean the dataset might be noisy or difficult to model perfectly. We could try different features to get better performance.

Insights on SVM (gamma 0.1):

In the SVM model with the gamma parameter set at 0.1, which affects the decision boundary's shape, we see almost a similar learning pattern as logistic regression with respect to the regularization parameter. When the regularization is strong, the SVM model also results in lower accuracy; when the C value is increased (i.e., regularization is weakened), the accuracy improves significantly, reaching the max of 0.80, which helps the SVM fit the data better. SVM reaches its maximum accuracy at 0.80 and decreases the accuracy score upon further increasing the C value.

Insights on SVM (gamma 10):

Similar to our previous models, the regularization parameter has the same effect: lower C value, strong regularization, lower accuracy; higher C value, weaker regularization, and higher accuracy, reaching the max of 0.72. However, when the C value is raised further, we see a decline in the accuracy score, which suggests the model is more complex and might be overfitting (could be because of a higher gamma value). The graphs show that the decision boundary follows the training set too closely and fails to generalize well like the last two models.



Comparison:

Comparing the three models, we learn that the SVM model with a gamma value of 0.1 performs best, with the highest accuracy score of 0.8; however, it seems to dip to stabilize. Although logistic regression generalizes well, it is our second-best-performing model, with accuracy stabilizing at 0.74. Lastly, SVM with gamma value ten is outperformed by both the models as it doesn't generalize with the test data compared to the other two models. Below is the graph that compares the three predictor models. From the above graph, we can see that the max accuracy for all the models is near the C value of 10^0.

Comments:

The model accuracy reaches its highest at 0.8, which could be the noise in our data set or the inability to learn trends from the training set. Working on a more extensive dataset than this could also help in learning more trends in the training set for better model performance. For the SVM model with a gamma value of 10, we see a sharp dip after it reaches the highest accuracy score of 0.72, which could be the result of overfitting, so trying different combinations of gamma and regular values might address this issue.