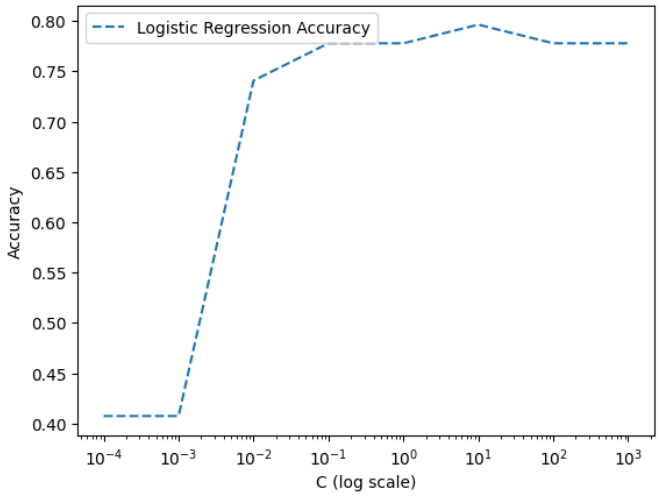
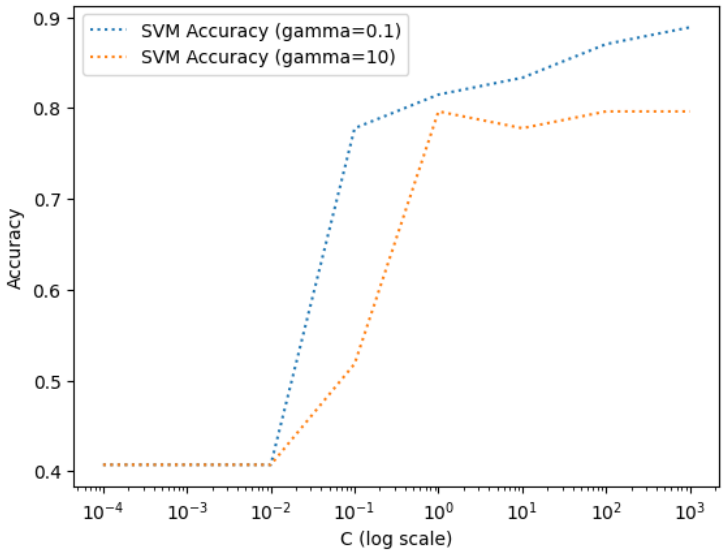
Thomas Lamont

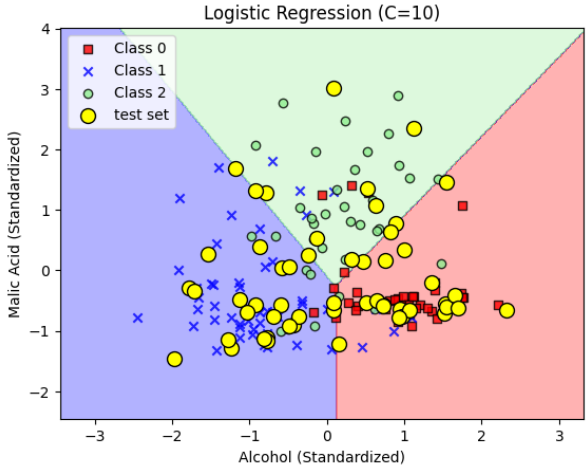
CS4275

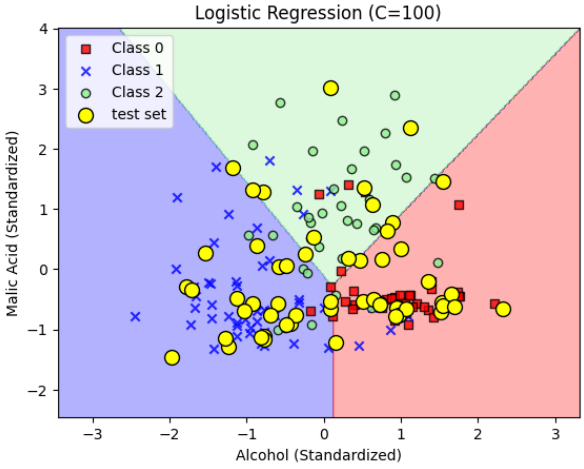
Small Project 1

The accuracies of these models show that the SVM model with a gamma of .1 is the best performing model, with an accuracy of .89 at C=10^3. This model outperforms both the LR and gamma 10 SVM models which both show a plateau in accuracy around 10^0. The low gamma value broadens the influence of each training example and has a smooth, though nonlinear, decision boundary. The LR model takes an oversimplified approach with its linear boundaries while the high gamma model makes more complex decisions that ultimately lead to overfitting and a decrease in performance.

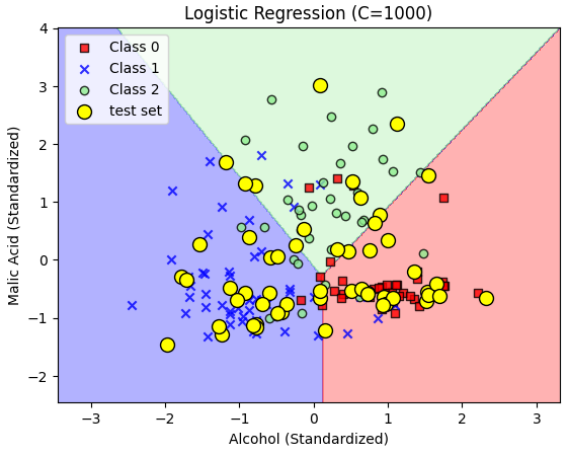


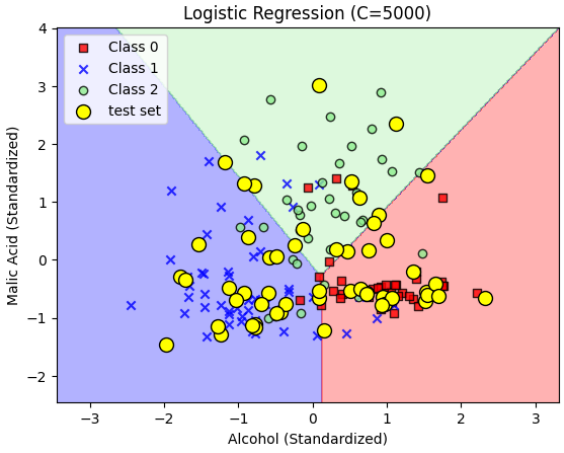


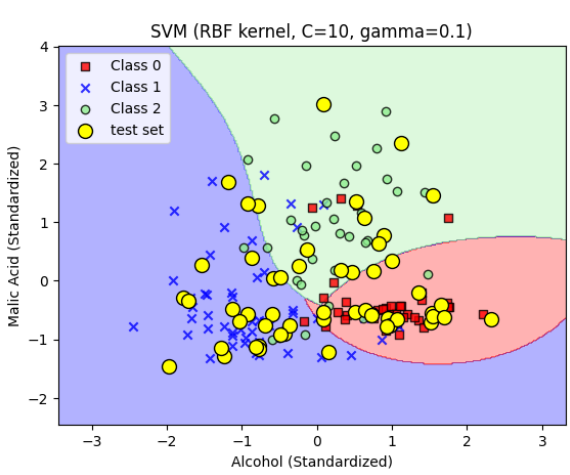


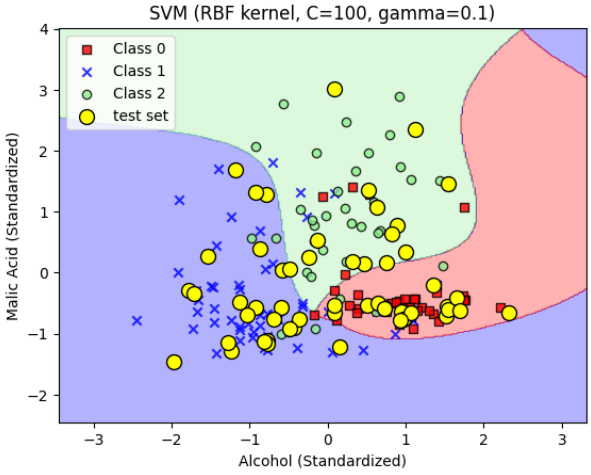


In the logistic regression models, there are only minor differences in the boundaries as C increases. The linear boundaries of logistic regression led to a simpler model that is fairly accurate but is outperformed by the low gamma SVM models. The minor differences between boundaries show the plateau in accuracy, at sufficiently high C, there are severe diminishing returns as C increases.

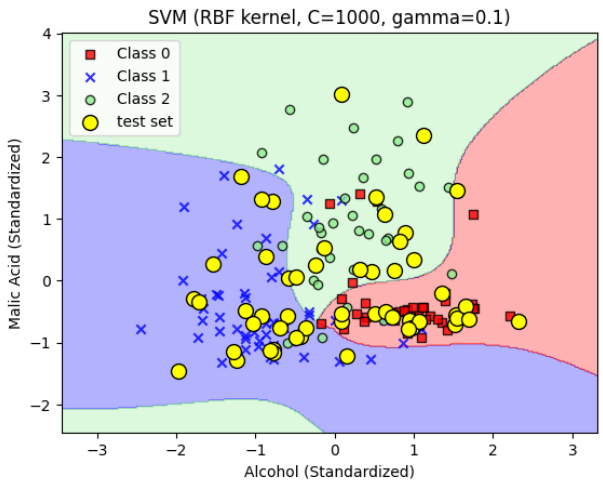


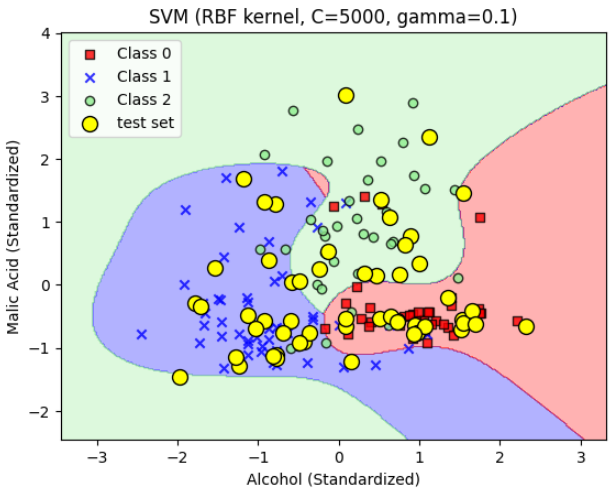


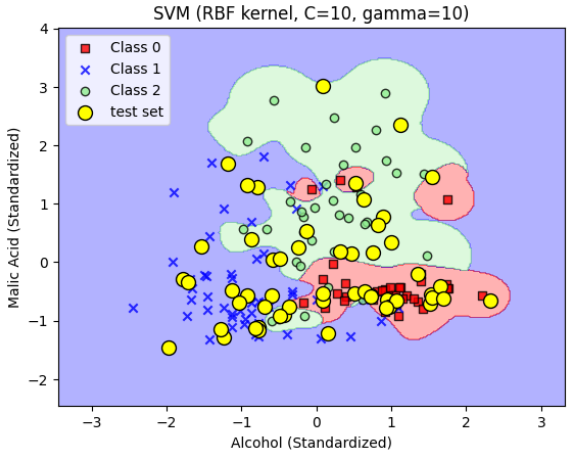


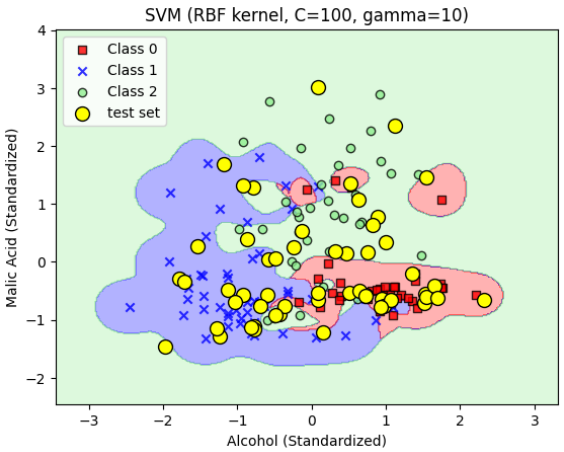


The SVM model outperforms LR as the non-linear boundaries can capture more nuance when predicting classes. The lower gamma also outperforms the high gamma SVM models, possibly due to better generalization and less overfitting than the higher gamma SVM. This model and gamma seem to be a good balance between simplicity and capturing the nuances of the nonlinear relationship.









With the higher gamma values, the shape of the boundaries becomes more complex. This complexity is leading to poor generalization and overfitting on the training data, which hurts the accuracy of the model as the complexity increases. Like the LR model, this also plateaus and there are only small changes to the boundaries after C=100.