Fun with classifiers report

Yang Song

What hyperparameter tuning did you do? Did you find ranges of hyperparameters that worked extremely badly? If so, explain why.

For the svc, I tried out linear, polynomial, and rbf. Among all of those, the rbf worked the best and the linear worked worst. When tuning the hyperparameters, I first focused on setting the value of C, and after rbf turned out to be the best one, I added the parameter gamma, which gave even better results. For all the three methods, I tried out several values of C, and it turned out that larger C's like 1000 worked badly, because lesser value specify stronger regularization.

Comparing the classifiers on the same training set, how do they rank in terms of the time needed to train them?

Time: SVC > Logistic Regression > Decision Tree

If some of the classes are more misclassified than others in your final classifiers, explain why?

Decision tree is more misclassified because there could be effects of imbalanced data over the course of fitting.

Based on the accuracy of the classifiers and the time needed to train and apply them, which would you recommend?

The Decision Tree is the fastest, but the accuracy is not that good. On the contrary, svc is the most accurate but takes a lot of time. A good way of predicting the data can be doing several decision trees and then combining their results in order to give a better accuracy, while avoiding the huge amount of time spent with svc.