Machine Learning Homework 1 Step-Wise Classification

Michael Panicci and Kyle Richards

March 18 2019

Results

n	Training Accuracy	Testing Accuracy
400	0.8225	0.7112
800	0.8025	0.7407
1200	0.7983	0.7495
1600	0.7887	0.7647
2000	0.7825	0.7642

Table 1: Accuracies for Training Set Sizes

The results of our Step-Wise Classification algorithm are shown in Table 1. Note that for each training set size, the accuracies using five comparisons are shown. Overall, as n increases, the training accuracy goes down and the testing accuracy goes up. At first, the discovered features were better to classify the smaller amount of training data than the testing data. With less data, patterns were found that applied well to the small amount of training data, but could not be generalized to apply to the testing data. Therefore, the features were actually overfitting the training data. As n increased, our algorithm found features that apply to a larger sample size. It begins to find features that could be generalized and applied to the testing data nearly as well as they did to the training data. Figure 1 shows how the training and testing accuracies begin far apart, but begin to converge as the set size increases. The lines shown are lines of best fit generated as a quadratic polynomial.

The comparison points that our algorithm selected are shown in Figure 2. The points were chosen in the following order: red, blue, green, purple, and orange. These points were selected using a training data set of size 2000.
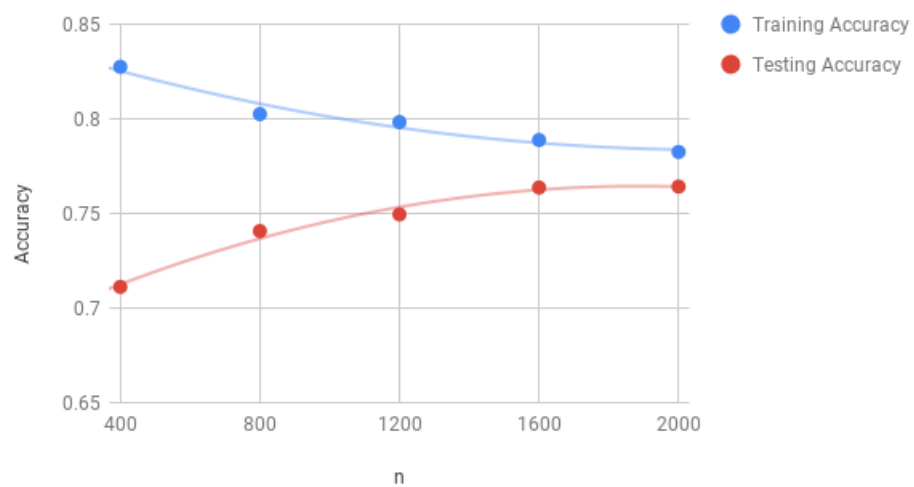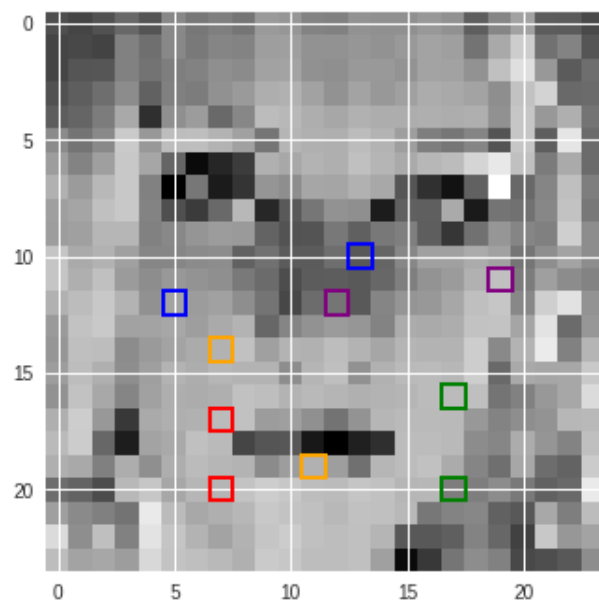
Figure 1: Training and Testing Accuracy for Different Set Sizes

Figure 2: Chosen Comparison Points for $n = 2000$