Name: Minh Pham
Username: mnpham

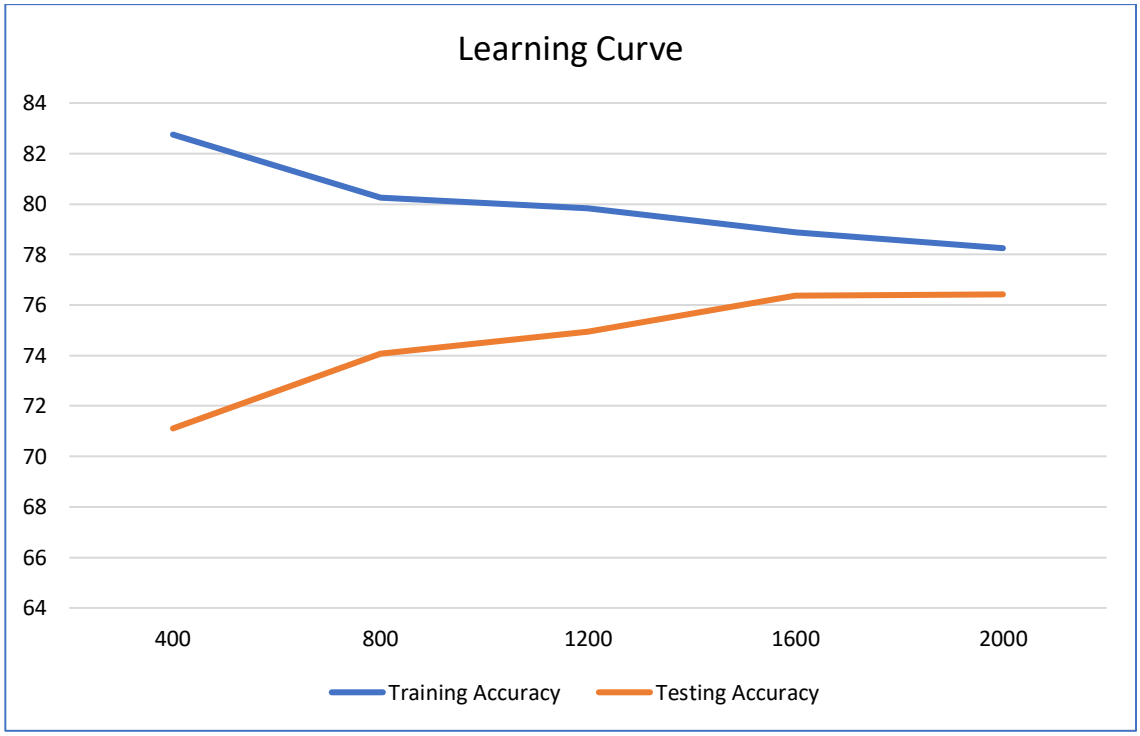
Machine Learning CS 453X
Homework 1

Methodology

The image dataset is split into training set and testing set. The stepwise classifier is trained over different sample sizes of the training data. After each model is trained its accuracy is measured over the entire training set and testing set.

The results are shown in the table below:

Training Size	Training Accuracy	Testing Accuracy	Time (Seconds)
400	82.75	71.11	104.33
800	80.25	74.07	147.41
1200	79.83	74.94	206.74
1600	78.87	76.36	260.09
2000	78.25	76.42	313.51



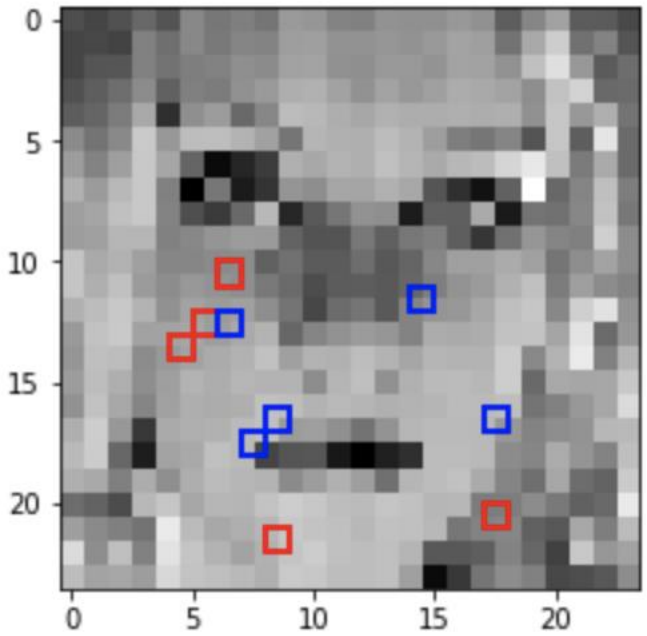
As we can see from the plot, the training accuracy tends to decreases while the testing accuracy tends to increases as the training size grows. This is understandable since as there is more data to train, the model will have to fit to more data, which makes the training accuracy drops. On

the other hand, the model will be better at generalizing on new data, increasing the testing accuracy where the number of data is fixed.

Visualization

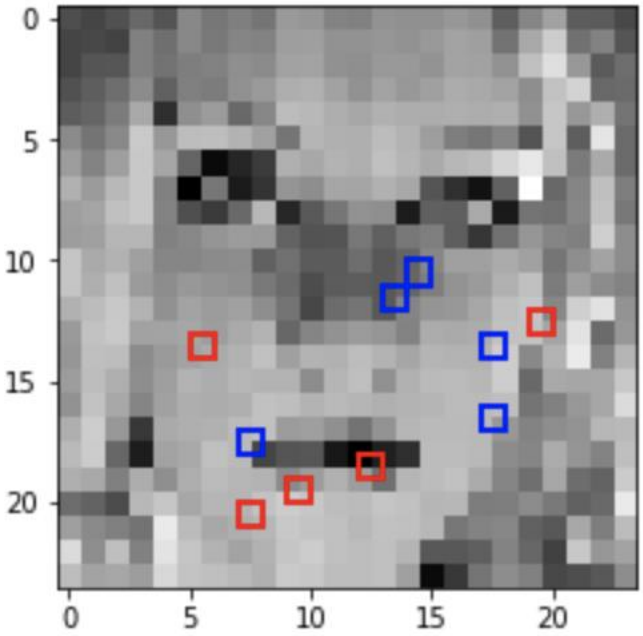
Training size: 400

- Top 5 predictors: [20, 17, 17, 7], [13, 4, 11, 14], [21, 8, 16, 8], [12, 5, 16, 17], [10, 6, 12, 6]



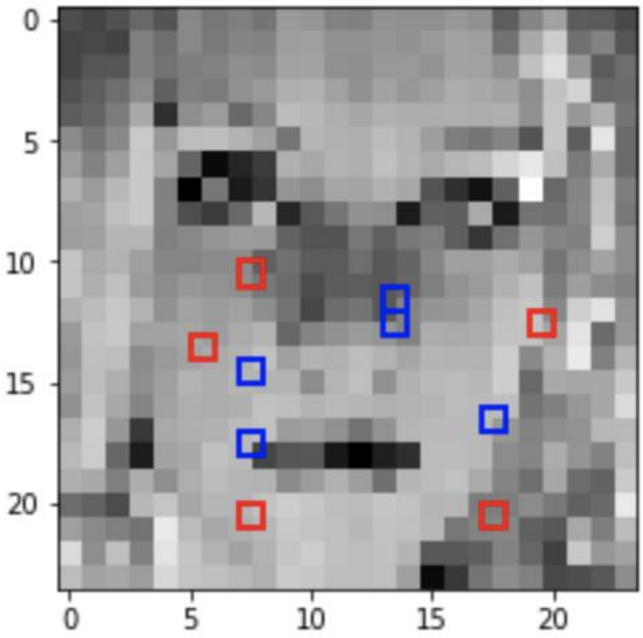
Training size: 800

- Top 5 predictors: [20, 7, 17, 7], [13, 5, 11, 13], [18, 12, 16, 17], [12, 19, 10, 14], [19, 9, 13, 17]



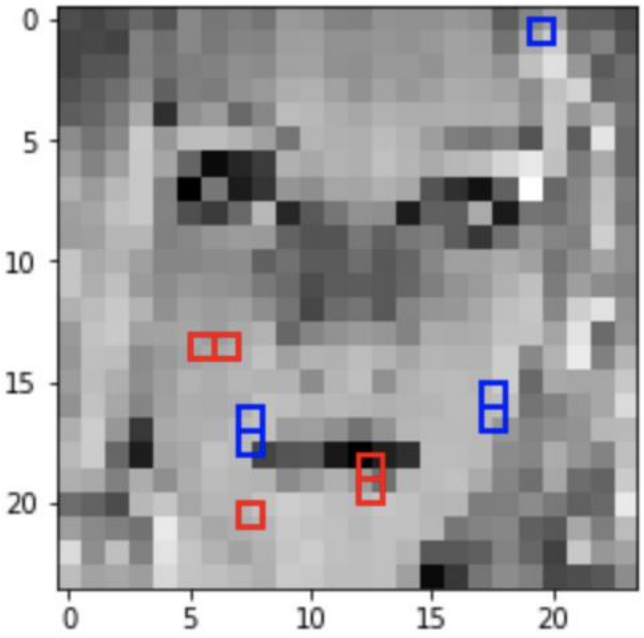
Training size: 1200

- Top 5 predictors: [20, 7, 17, 7], [13, 5, 11, 13], [20, 17, 16, 17], [12, 19, 12, 13], [10, 7, 14, 7]



Training size: 1600

- Top 5 predictors: [20, 7, 17, 7], [13, 6, 16, 17], [18, 12, 16, 7], [13, 5, 0, 19], [19, 12, 15, 17]



Training size: 2000

- Top 5 predictors: [20, 7, 17, 7], [12, 5, 10, 13], [20, 17, 16, 17], [11, 19, 12, 12], [19, 11, 14, 7]