Sadie Lombardi
Machine Learning
Homework 1
Professor Jacob Whitehill

Training vs. Testing Data

n	Training accuracy		Testing accuracy	
400	0.7875	79%	0.7016	70%
600	0.7792	78%	0.7157	72%
800	0.7713	77%	0.7253	73%
1000	0.7709	77%	0.7295	73%
1200	0.7655	76%	0.7303	73%
1400	0.7645	76%	0.7206	72%
1600	0.7583	76%	0.7295	73%
1800	0.7558	76%	0.7232	72%
2000	0.7557	76%	0.7287	73%

Moreover, characterize in words (and write them in the PDF) how the training accuracy and testing accuracy changes as a function of n, and how the two curves relate to each other; what trends do you observe?

According to the data listed above, we observe that the accuracy over the training data is higher than over the testing set for all values of n. When n was 400, we received our highest overall training accuracy, and lowest overall testing accuracy. As we can see in Figure 1 below, as we increase our n value, the training accuracy decreases, while the testing accuracy increases. If we were to increase our n value, we can infer from the current data that both curves will meet somewhere around 0.75 (75%) accuracy.

Training and Testing Accuracy (Figure 1)

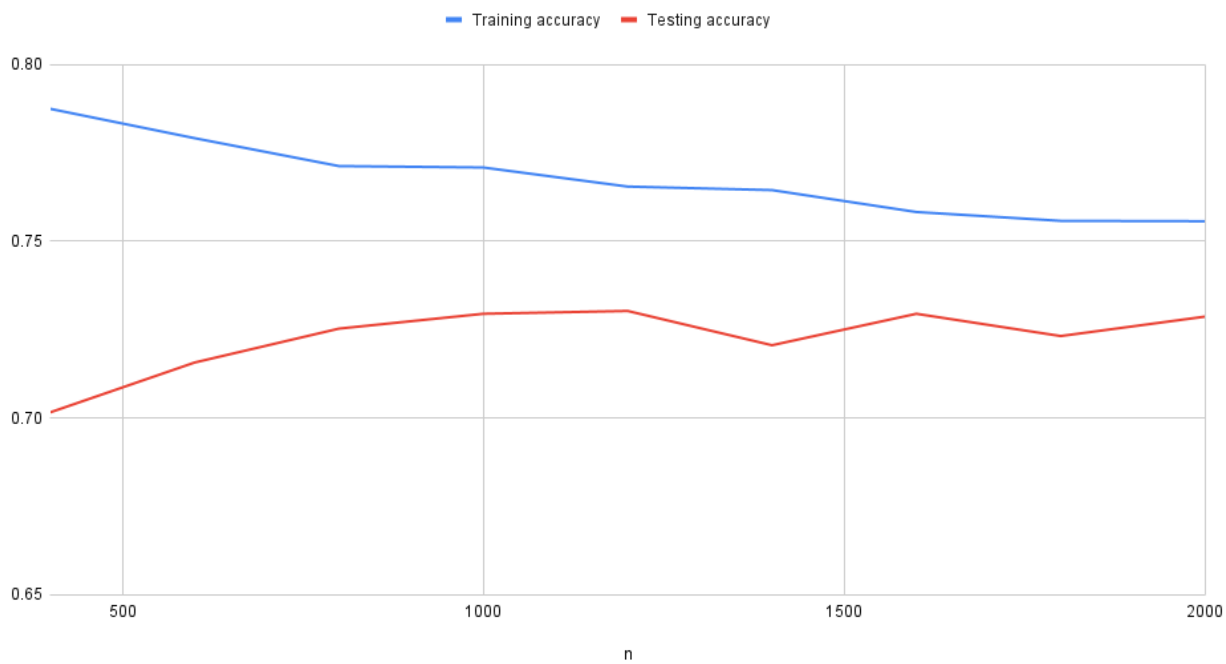

Figure 1: Training accuracy and testing accuracy averages plotted against each other over n.

Features on Singular Image (Figure 2)

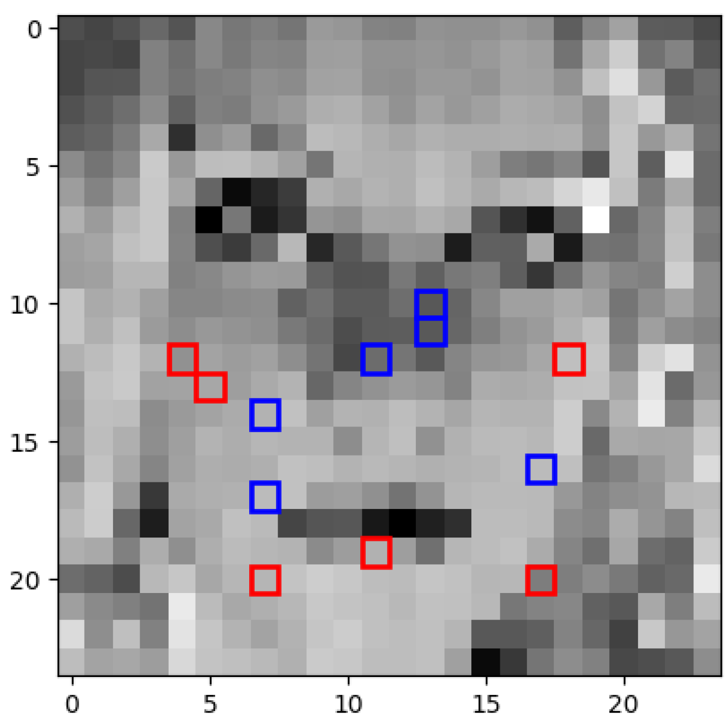

Figure 2: Image with best features (pixel pairs) shown in blue and red.