Greg Matthews
CS 383
February 19th, 2018

Assignment 5:
Support Vector Machines

1 Theory

None.

2 Support Vector Machines

Classification Statistics:

$$Precision = \frac{t_p}{t_p + f_p} = 0.9274 = \mathbf{92.7}\%$$

$$Recall = \frac{t_p}{t_p + f_n} = 0.9097 = \mathbf{91.0}\%$$

$$F1\ Measure = 2 * \frac{precision * recall}{precison + recall} = 0.9185 = \mathbf{91.9}\%$$

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} = 0.9394 = \mathbf{93.9}\%$$

3 Multi-Class Support Vector Machines

Classification Statistics:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N}(Y_i = \hat{Y}_i) = 0.9195 = \mathbf{92.0}\%$$

3 Confusion Matrix

Table 1: Confusion Matrix (True Classification vs Predicted)

	True Class 1	**True Class 2**	**True Class 3**
Predict Class 1	76%	2.4%	1.4%
Predict Class 2	3.0%	9.8%	1.4%
Predict Class 3	1.0%	1.4%	6.4%

Looking at the confusion matrix, we can determine that the class that had the most trouble with classification is class 1 getting labelled as class 2 by 3%. This however is slightly disproportional because the overall number of class 1 predictions was around 80% and the other 2 classes being around 20%, therefore since Class 1 predictions were more abundant, naturally incorrect prediction rates for Class 1 can be higher. Keeping all classes proportionate, then Class 2 would have the higher chance of misprediction with $2.4/(2.4 + 9.8 + 1.4) = 17.6\%$ of the time being misclassified as class 1, in terms of looking at only Class 2 classifications.