Machine Learning	**Homework 1**
CSCI 5622 Fall 2017	**Due Time Nov 10, 2017**
Name: Andrew Kramer	**CU identitykey: ankr1041**

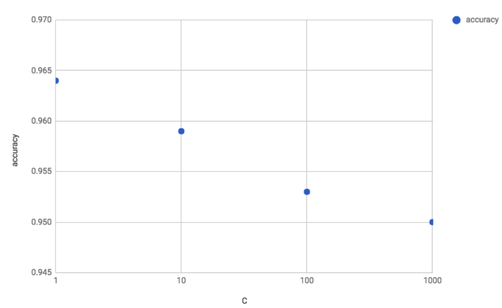
1 Support Vector Machines (40pts)

1.1 Comment on the experiments you ran with GridSearch and the optimal hyperparameters you found

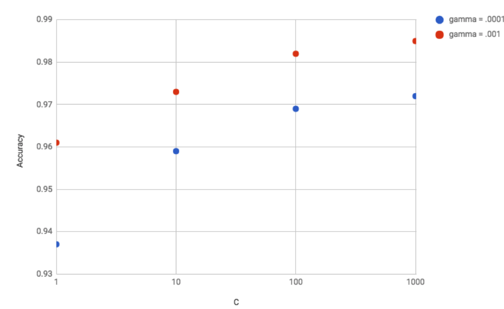
A linear kernel was tested with penalty parameters of 1, 10, 100, and 1000. The highest accuracy for the linear kernel was obtained with a penalty parameter of 1, but overall the maximum accuracy of 0.964 obtained by the linear kernel was not competitive with the RBF and polynomial kernels.

An RBF kernel was tested with penalty parameters of 1, 10, 100, and 1000 and gamma values of 0.0001 and 0.001. The maximum accuracy of 0.985 was obtained with a penalty parameter of 1000 and gamma of 0.001.

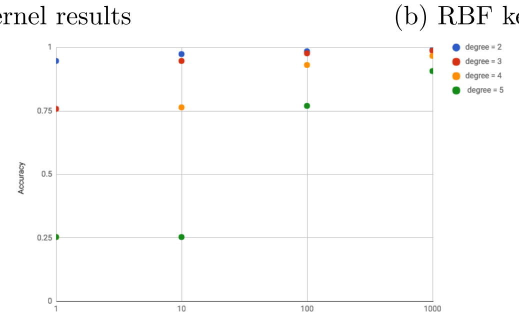
A polynomial kernel was tested with degree of 2, 3, 4, and 5 with penalty parameters of 1, 10, 100, and 1000. The maximum accuracy of 0.991 was obtained with a degree of 2 and penalty parameter of 1000. See the graphs below for details.



(a) linear kernel results



(b) RBF kernel results



(c) polynomial kernel results

Figure 1: results for all tested hyperparameter combinations

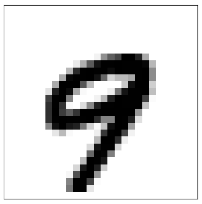
1.2 Comment on classification performance for each model for optimal parameters by either testing on a hold-out set or performing cross-validation.

The highest performing configurations found in the previous section were trained and tested on separate datasets. The RBF and polynomial kernels achieved 0.99 testing accuracy for both the

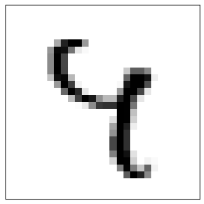
4 and 9 classes. The linear kernel achieved a testing accuracy of 0.97 for 4s, 0.98 for 9s, and 0.97 average.

1.3 Give examples in picture form of support vectors from each class when using a polynomial kernel.

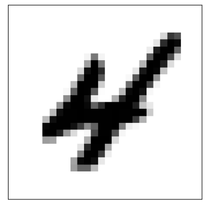
The following examples were obtained by randomly sampling from the set of support vectors found when training a classifier with a 2nd degree polynomial kernel and penalty parameter of 1000.



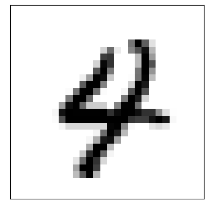
(a) A nine



(b) A second nine



(c) A four



(d) Another four

2 Learnability (25pts)

2.1 Give a bound on the number of randomly drawn training examples sufficient to assure that for any target class c in C, any consistent learner will output a hypothesis with error at most 0.15 with probability 0.95

$$
\begin{aligned}
m &\geq \frac{1}{\epsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right) \\
|H| &= \text{number of possible triangles in the space} \\
|H| &= \binom{\text{possible vertices}}{3} \\
\text{possible vertices} &= 100^2 = 10000 \\
|H| &= \binom{10000}{3} = 166616670000 \\
\delta &= 1 - 0.95 = 0.05 \\
\epsilon &= 0.15 \\
m &\geq \frac{1}{.15}\left(\ln(166616670000) + ln\frac{1}{0.05}\right) \\
m &\geq 192
\end{aligned}
$$

3 VC Dimension (20pts)

Solution.