# CMPT 404
# Homework 2

Daniel Clark

October 11, 2016

**1** Compare two algorithms on a classifications task: the Pocket algorithm (designed for classification), and linear regression (not designed for classification).

I ran the Pocket, Linear Regression, and hybrid Linear-Pocket algorithms for 200 trials with the pocket algorithms cutting off after 100 unimproved iterations. On average, the Linear Regression algorithm had the largest amount of classification error $E_{out} = 0.0783$. The standard Pocket algorithm came next with an error of $E_{out} = 0.05785$. The hybrid Linear-Pocket algorithm had the best error $E_{out} = 0.05255$. This makes sense. Linear Regression is not designed for classification, and so should perform worse than algorithms tailored to that task. The slight increase in output by the hybrid algorithm is due to data sets with easily separable data. In these cases, the hybrid algorithm has a great solution on the first iteration, while occasionally the standard pocket algorithm finds solutions that just barely separate the training data but fail to perfectly separate the test data.

On average, the standard Pocket algorithm took 76.925 iterations to complete while the Linear-Pocket algorithm took 57.01 iterations. Note that the Linear-Pocket algorithm took substantially fewer iterations on average than the standard pocket algorithm. This comes about from two effects. Firstly, in cases where the data are easily separable, the hybrid algorithm is given a solution instantly by the starting guess, while the standard

pocket algorithm must iterate to find a solution. In other cases the hybrid algorithm still has an advantage because it starts with a very good approximation based on the data, saving on average a dozen iterations that the standard pocket algorithm takes to get to that level of error.

I then ran the same algorithms under the same conditions except that I added four data points at $(\pm 8, \pm 8)$. These were to serve as outliers and ensured that the data were almost never separable. On average, the Linear Regression algorithm had the largest amount of classification error $E_{out} = 0.29865$. The standard Pocket algorithm came next with an error of $E_{out} = 0.0554$. The hybrid Linear-Pocket algorithm had the best error $E_{out} = 0.04905$. The error results for the pocket algorithms are similar to the errors without the outliers. The Linear Regression algorithm however, suffered greatly. Since the Linear Regression algorithm is mathematically more affected by outliers than classification algorithms, these outliers affect the solution given by Linear Regression, and thus affect the true error.

The average number of iterations after the addition of outliers was significantly affected. On average, the standard Pocket algorithm took 145.67 iterations to complete while the Linear-Pocket algorithm took 121.225 iterations. The data are no longer perfectly separable, so the pocket algorithms must both at least reach the cutoff point. Surprisingly, the difference between the average number of iterations of the two pocket algorithms remains approximately equal. The hybrid solution still has the same advantages as before. Now, when it would have terminated immediately with a perfect solution, it must first reach the cutoff, never changing its best solution. This is supported by a more detailed analysis of the number of iterations. In over half of all cases, the hybrid algorithm terminated after exactly 100 iterations.