

## CST 3170 Course Work 2

### Machine Learning

Due Date: Week 19 (February 22nd, 2024)

40% of Overall Course Mark

This work is to build a machine learning system to categorise one of the UCI digit tasks. You should develop the system on your own from scratch. You should then run a two-fold test, and report your results.

The data is from the University of California at Irvine's Machine Learning Repository. It's the [Optical Recognition of Handwritten Digits Data Set](#). This gives you two data sets, training set and a test set. I've converted them to two data sets, cw2 data set 1, and cw2 data set 2, which should be used by your system.

You should write all of your code. If you use an existing algorithm, you should reference that algorithm in your code and in your report. The code should be written in Java, and should run in the lab from eclipse.

#### Marking scheme

Points	Area
20	Report
20	Running Code
20	Quality of Code
20	Quality of Algorithm
20	Quality of Results

You should write a brief (1-2 page) report on your system. This should describe the algorithm you used, and why you chose this algorithm. It should also show the results of a two-fold test using the provided data; a brief discussion of data usage would be useful.

Quality of code and algorithm are important for good marks. The code should be well commented and structured. Selection of a good algorithm is also important. Simple algorithms may be effective, but a relatively complex algorithm may get you more points just for effort.

Finally, the quality of the results do matter. To get reasonable marks you need to surpass the baseline reported on the UCI website. This is not a competition between students, and discussing performance with your colleagues will probably be useful.

Note for scraping by: the base line reported on UCI website is nearest neighbour using Euclidean distance. You should be able to implement this quite easily (and will probably want to start with this). This should be enough to pass (10 report, 20 running, 10 code, and 5 results).

Please submit the code, the mark sheet, and analysis to the coursework folder of CST 3170 on myunihub. You are also welcome to email a copy to the tutor.