

# Machine Learning

Worth: 5% of total grade [5 marks]

Due Date: Friday March 29<sup>th</sup> 2019, 23:59

## How do Ensembles Work? (5 marks):

You will have to run the Experimenter in Weka. You will need to run Randomforest, AdaBoostM1, and Bagging. You will have to run all three algorithms on 3 datasets: Iris, Car, and Balance-scale. Make sure you run 10-fold cross validation so you can tell which algorithm is better on which dataset. Run the default settings on the 3 algorithms. It is important that you input the datasets in this order:

1. iris.arff
2. car.arff
3. balance-scale.arff

It is important you enter the algorithms in this order:

1. RandomForest
2. AdaBoostM1
3. Bagging

This is to make the marking of your assignments more automated. If you do them in another order you might not get full marks. In addition it is important to use the default settings for all the algorithms, please do not change them.

What you need to turn in:

You need to turn in a .pdf file of 1 page MAXIMUM which should include:

- 1) The two-tailed t-test (corrected) output from Weka "Analyse" in the Experimenter. Include a screen shot of this in your .pdf
- 2) A paragraph about each of the 3 datasets saying why you think a particular algorithm did the best of each dataset.
- 3) A final paragraph saying which algorithm you think is usually better and why.

Marking is based on the following material:

½ mark for having the correct Weka output

3 marks (1 for each description of the 3 datasets)

1 mark for the final paragraph saying which algorithm is more reliable

½ mark for good presentation and English

The datasets can be found at

<https://canvas.auckland.ac.nz/courses/40682/files/folder/Assignment2>

The assignment must be submitted to Canvas. It will be run through Turnitin.