Machine Learning 361 – Assignment 1

Worth: 5% of total grade [5 marks]
Due Date: Friday March 13th 2020, 23:59

Signal or No Signal ??? (5 marks):

There are two data files given: obscuredA.arff and obscuredB.arff. One of them is the real dataset and one of them is a randomized dataset where the class attribute has been randomly shuffled. Can you determine which is which? Most of the marks will be for explaining how you made this determination. Please try to determine which is which using at least two methods and discuss why you think one method is more reliable than the other. You can use whatever tools you wish to do this assignment, but Weka will be taught in tutorial.

In addition you will be given two smaller datasets for each case.
obscuredA-50.arff, obscuredB-50.arff which contains ½ the number of rows as the original
obscuredA-25.arff, obscuredB-25.arff which contains ¼ the number of rows as the original
Do your techniques continue to work with the smaller datasets?

What you need to turn in:
You need to turn in MAXIMUM 2 page .pdf which includes:

a) Which dataset you think is the real file
b) 2 paragraphs, one on each of the 2 different ways you found to tell which dataset had a signal in it.
c) A paragraph discussing whether these techniques continue to work as the dataset gets smaller, please indicate the reasons for this behaviour.
d) A final paragraph discussing which method you feel is more reliable and why.

Marking is based on the following material:
1 mark choosing the right data set
1 mark for clearly describing each of the 2 methods used to determine the signal strength
1 mark for discussing how well these methods work as the dataset gets smaller
0.5 marks for good presentation and English
0.5 marks for your final paragraph with a comparison of the two methods

The datasets can be found at
https://canvas.auckland.ac.nz/courses/45892/files/folder/Pat/Assign1

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

Copyright Warning Notice

Copyright © University of Auckland. This material is provided to you for your own use. You may not copy or distribute any part of this material to any other person. Failure to comply with this warning may expose you to legal action for copyright infringement and/or disciplinary action by the University.