

Machine Learning 361 – Assignment 1

Worth: 5% of total grade [5 marks]

Due Date: Friday March 15th 2019, 23:59

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

There are two data files given: obscured1.arff and obscured2.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 easier or better than the other.

What you need to turn in:

You need to turn in MAXIMUM 1 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 final paragraph discussing which method your feel is more reliable

Marking is based on the following material:

1 mark choosing the right data set

1.5 marks for each of 2 methods used to determine the signal strength

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/40682/files/folder/Assignment1>

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