

# Machine Learning

## Coursework 1 Specification

### General Information

Please read the following information carefully before starting your coursework.

- Submission date: **19 Oct 2022 by 23:59 GMT** online via Moodle.
- Contribution: 25% of the total mark.
- Academic misconduct: All submissions will be processed through plagiarism tool. If signs of misconduct are found, all students involved will be contacted to discuss further steps. Please see here for information on academic integrity at the university.

<https://portal.roehampton.ac.uk/information/Pages/Academic-Integrity.aspx>.

Our guiding principle is that academic integrity and honesty are fundamental to the academic work you produce at the University of Roehampton. You are expected to complete coursework which is your own and which is referenced appropriately. The university has in place measures to detect academic dishonesty in all its forms. If you are found to be cheating or attempting to gain an unfair advantage over other students in any way, this is considered academic misconduct and you will be penalised accordingly. Please don't do it. It is not worth it.

- Grading advice: While fully functioning programs and its documentation that meet all the requirements are the goal of the test, the programs and the rationale for their design, implementation, testing and discussion will be assessed for efficiency, simplicity, creativity, and good style.

## Coursework Task

To complete this coursework, you will develop a Python program to solve a supervised learning problem. The goals of the program are to import the data, split the data into training and testing set, build a model for prediction and evaluate the model.

- Import the data.

You can access the dataset in section 'Assessment' of the Moodle page, named as 'cw1data.csv'. Please download the data and import it into your program.

- Visualise the data

Develop some codes to visualise the data. You should at least produce one figure of the data.

- Split the data into training and testing set.

The column 'y' will be used as output variable and the remaining columns will be used as input variables. Define the input and output variables and then split them into training and testing datasets.

- Build the model and generate prediction

Develop some codes to build regression models and then use the models to generate predictions on testing dataset. You should at least choose two types of regression models, build the models and use the models to generate predictions.

- Evaluate the model

Evaluate the performances of the models using at least one method. Compare the performances of the models and identify the model with the best performance.

- Develop an additional feature (optional)

Develop codes to implement one additional feature. For example, you could try to identify the importance of the variables, visualise your predictions results, or perform a cross validation, etc.

## Submission Requirement

You should submit the codes with your name and ID number. The file should contain documentation explaining your implementation (comment your code).

You will also need to present your work in either the following two ways:

- Present to a tutor during lab session within two weeks after the due date.
- Record your screen while running the codes and include the video link in your submission. The video should be less than 3 minutes and highlight the key features you implemented.