CST4050 - Modelling, Regression and Machine-Learning

Formative assessment — Coursework 1

General information

Deadline for submission: Friday 22nd November 2019, 23:59pm.

You are required to submit your work via the dedicated Unihub assignment link by the specified deadline.

Note that this link will '*timeout*' at the submission deadline. *Your work will not be accepted as an email attachment if you miss this deadline*. Therefore, you are strongly advised to allow plenty of time to upload your work prior to the deadline.

Your submission comprises is composed by a *group report* and *an individual report*. Each report needs to be a single PDF file including your step-by-step process.

The challenge

This coursework is about developing a linear regression classifier on some synthetic data composed by 200 observations, 50 independent variables and 1 independent variable. You can download your data on UniHub.

Your goal is to build a predictive model for the dependent variable *y*. You need to make sure to find the best trade-off between model bias-variance.

Your submission

Individual report

Goal of the individual submission is to show how you, individually, are able to train and tune a predictive model.

As individual submission, you need to export your individual Jupyter Notebook file into a PDF file. The submitted PDF file needs to cover the following tasks.

- *Task 1*. Open the data provided on UniHub and get a summary of the data. Standardise your data if needed.
- *Task 2*. Use train-test or, much better, 10-fold cross validation to train and tune your classifier.
- *Task 3*. Compute the accuracy of your classifier including MSE and R^2.
- *Task 4*. Explain the model you have got. In other words, show and comment the β coefficients and residuals of your final model.

Group report

Goal of the group submission is to show how you, as part of a team of 3-5 people, are able to compare different predictive models and decide — as a team — which model has the best fit in term of bias-variance trade-off.

As group submission, you need to export your group Jupyter Notebook file into a PDF file. The submitted PDF file needs to cover the following tasks.

- *Task 1*. Compare the performance of the models of your peers based on any measure (e.g., MSE, R^2).
- *Task 2*. Comment each model, in term of bias-variance trade-off.
- *Task 3*. Choose the model offering the best bias-variance trade-off.

Formative assessement

Note that this is a formative assessment. You will receive feedback in class on your solution so you can learn and improve.

Shortly a new problem will be released and you need to submit a solution by the end of week 11. This will be a summative assessment and it will be worth 50% of your final score.

Marking scheme

Your summative project will be evaluated according to:

Individual report:

- *Task 1*. Up to 10 marks.
- *Task 2*. Up to 20 marks.
- *Task 3*. Up to 20 marks.
- *Task 4*. Up to 20 marks.

Group report:

- *Task 1*. Up to 10 marks.

- *Task 2*. Up to 10 marks.
- *Task 3*. Up to 10 marks.