Dear All,

Due to the limitations around closed-book exams, we are going to use the Kaggle competition as a take-home exam.

The Kaggle competition was launched, please subscribe using this link: https://www.kaggle.com/t/97f3deae19e746f68fc7b183e1ec6175

You will find the train and test data in the data section of the competition, with a description of the features. You will also find a sample solution file. You will need to create the same format for your own solution. An example code is attached, which prepares a solution for the TItanic (https://www.kaggle.com/c/titanic) toy competition. Please note a difference between the formats: submitted values for the Titanic problem are 0 or 1, while in our competition real values are expected between 0 and 1 (in our competition, no rounding is needed in the final csv).

The deadline for the Kaggle solutions is 17:10 CET, 25th of March. By this time make sure you submit at least the following 4 types of solutions:

a linear model prediction after parameter tuning

a random forest prediction after parameter tuning

a gradient boosting prediction after parameter tuning

a neural network prediction after parameter tuning.

Your best model will need to have at least 0.65 AUC to consider this part of the exam complete.

You will also need to submit to Moodle the documentation (Rmd and HTML) of the work you did, including exploratory data analysis, data cleaning, parameter tuning and evaluation. For extra points, build a stacked model, explain how it works and evaluate its results.

The deadline for this is the 29th of March, at the end of the day. Please make sure you include the R code as well in the HTML export.

We will use the lecture on Monday and the lab on Wednesday to discuss any questions. In the meantime, feel free to start working on this task, the Kaggle competition is open.

Please let us know if you have any questions or if you encounter any difficulties when submitting solutions.