

Homework 3

SUBMISSION INSTRUCTIONS

- 1) You have to use Jupyter Notebook
- 2) Click the Save button at the top of the Jupyter Notebook.
- 3) Select Cell → All Output → Clear. This will clear all the outputs from all cells (but will keep the content of all cells).
- 4) Select Cell → Run All. This will run all the cells in order, and will take several minutes.
- 5) Once you've rerun everything, select File → Download as → PDF via LaTeX (If you have trouble using "PDF via LaTeX", you can also save the webpage as pdf. *Make sure all your solutions especially the coding parts are displayed in the pdf, it's okay if the provided codes get cut off because lines are not wrapped in code cells*).
- 6) Look at the PDF file and make sure all your solutions are there, displayed correctly. The PDF is the only thing your graders will see!
- 7) Submit your PDF on Latte.

Question 1. Say we are running a binary classification loan model, and rejected applicants must be supplied with a reason why they were rejected. Without digging into the weights of features, how would you supply these reasons?

Question 2. Now, we are going to analyze **Kaggle** project of the Titanic dataset. You can download the data from <https://homl.info/titanic.tgz>. This will give you two CSV files, *train.csv* and *test.csv*. The goal is to train a classifier that can predict the Survived column based on the other columns.

- You will work through this project end to end
- Start with a problem statement
- Explore and visualize the data to gain insights
- Prepare the data for machine learning algorithms
- Select a model and train it
- Fine-tune your model
- Present your final outcome
- Comment on your code if necessary