

OLA 2- Supervised Learning

The following OLA gives you an opportunity to work on applying supervised machine learning algorithms. It gives you an opportunity to take your exam project idea or another problem and work through all the stages of the machine learning process, from business case to deployment of a model.

DUE DATE: Week 11 (Fri 15 March) hand in on Moodle or by email to me if the upload folder won't work.

Task

Select and prepare a data set to explore a problem you are interested in understanding. This can be your exam project problem or in the same problem domain. Using at least three different models explain your problem, how machine learning can be used to explore the problem and demonstrate your solution.

You should demonstrate that you have split the data in a sensible way for testing, training and validation. You should explain how you have measured quality using appropriate measures for accuracy, precision and recall (sensitivity).

You should describe how this solution could be wrapped into an application and deployed - how do you move from the laboratory to the real world?

Deliverables

- A Jupyter notebook containing all the code used with comments/Markdown to explain what you are doing: This includes:
 - Choice of models
 - Selection of the best fit model
 - Explain what 'overfitting' means and how you can spot it
 - Explain what hyperparameters you have adjusted and why
 - Explain how you have measured quality using F_1 score and explain the terms accuracy, precision and recall (sensitivity).
- Prepare a short presentation (save it as a PDF file) on measuring distance between data points
 - Manhattan
 - Euclidean
 - Hamming distance

The presentation should provide examples of calculating these distance and explanation of what we do with that information - why are there different measures?

- Show how you would move code out of the Notebook into an application. You should build a prototype that shows how you would integrate the solution into an application - this can be a simple user interface (console or HTML based) to the Python script or another solution.
- Submit work to the hand-in folder via the Moodle course or by email to me - it is a group project.