Examination MAT003 - Introduction to machine learning

Prof. Dr. Ashkan Nikeghbali

17.12.2019

Rules
Deadline: 20.01.2020 at 23:59
Please send the project report via email to gabriele.visentin@math.uzh.ch. The project report must consist of a single compilable Jupyter notebook. You can work alone or in pairs.
Please specify your names at the beginning of the notebook and in the filename of the notebook ("surname1-surname2.ipynb", or "surname.ipynb" if you are alone).
Copying is not allowed, neither in the text nor in the code sections of the project report.
Cooperation between members of different groups is not allowed.

Structure of the report
The project report should consist of detailed explanations together with compilable code, describing your approach to solving the assigned problem from data preprocessing to final model selection.
Make sure to structure your report in sections, including in particular the following topics:

- Problem description
- Data preprocessing
- Model implementations
- Validation
- Testing
- Conclusion

Please add comments to your code and aim for readability.

Goal
The goal of the project is to describe and document in detail how to find and implement the best machine learning algorithm for the dataset assigned to you.
You have been assigned a multi-class classification problem. You must solve it only by reducing it to a sequence of binary classification problems. In particular you are NOT allowed to use the multi-class implementation of sklearn.
One of the models you implement must be the random forest model. This model must be programmed by you. Implementation details can be found in the attached pdf.

List of materials

- Random Forest implementation details (pdf)
- Dataset description (txt)
- Dataset (csv)