

A Phd Course

Let's start with an observation:

Machine Learning is Optimization

Let's start with an observation:

Machine Learning is Optimization

- Well, maybe a few exceptions apply (looking at you, non-parametric models...)
- ...But broadly speaking, the statement is true

From an optimization point of view:

- Classical ML has focused on large-scale, unconstrained problems
- ...And for good reason!
 - Large input spaces
 - Large parameter spaces
 - Expensive cost functions

Let's start with an observation:

Machine Learning is Optimization

However, that's a bit reductive:

- What if our ML problem has a non-trivial structure?
- What if we have external knowledge?
- What if there are physical laws, or regulations?
- What if we want to use data to help with a decision problem?

Then, we may want to take a broader view...

...Since other optimization techniques may be of help!

This course is about seeing ML and CO as a whole:

- Focus: integration of Machine Learning and Constrained Optimization
 - Emphasis on modeling aspects and knowledge integration
 - No mention of acceleration techniques
- Three parts:
 - Constrained Optimization for Data Mining
 - Handling constraints in Machine Learning
 - Hanlding Machine Learning models in optimization
- It will be far from a complete overview
 - Bias on my own research
 - I'll do my best to provide pointers
 - ...And starting points for related areas

Course material

- Jupyter notebooks for everything; version 6, for compatibility with...
- RISE plugin for the presentation mode
- Poetry for dependency management and reproducibility
- Each part in different github repository

Lectures are meant to be executed

For doing it locally, you need to:

- Install Poetry using pipx or the native installer
- Clone the git repository
- Open a terminal on the cloned repository
- Run poetry shell then jupyter notebook

Let's check our directory structure:

```
In [1]: !ls ..

LICENSE.md data pdfs pyproject.toml
README.md notebooks poetry.lock
```

- The data folder is meant for datasets & co.
- The pdf folder contains PDF exports for the notebooks
- pyproject.toml contains configuration parameters, and lists all dependencies
- The poetry.yml files is used by poetry and reports spefic version numbers
- The **README.md** file contains instructions on how to run the lectures

Let's check our directory structure:

```
In [2]: !ls ../notebooks
        00. Introduction to the Course.ipynb
        01. Constraint Optimization for Data Mining.ipynb
        02. How to Deal with Quadratic Programs.ipynb
        03. Solving the Path Formulation.ipynb
        04. Consolidation as Regularization.ipynb
        05. Mixed Integer Linear Programming.ipynb
        06. Solving the Consolidation Problem.ipynb
        07. From Pricing...ipynb
        08. ... To Column Generation.ipynb
        09. Constraints in the Master.ipynb
        10. Constraints in the Subproblem.ipvnb
        11. Maximum Wait Pricing Model.ipynb
        12. The Alternating Method of Multipliers.ipvnb
        assets
        print-pdf.sh
        rise.css
        util
```

- In the notebooks folder we have the notebooks themselves
- A util folder with a python module for utility functions