Syllabus

Overview

The Software Engineering for Data Scientists course is meant to help data scientists write production ready code as well as gain familiarity with the tools used to make models available to their users. The core idea we will be exploring is making code robust and re-usable across a team. This course can also serve as an introduction toward ideas used in ML Ops and Data Engineering.

In addition to the videos, a git repository stored under the Erdos Institute github organization will be available. Videos will be coordinated with specific tags.

Material

Setup

- The command line
- Text editors and IDEs
- Dependency management

Code quality + reusability

- Code style
- Writing packages
- DRY code + unit testing
- Type hints
- Trunk based development + Pull Requests
- CI/CD

Software engineering concepts

- Object Oriented Programming
- Functional Programming
- Data Structures
- Intro to Algorithms
- Intro to distributed computing

Productionization

- Serialization

- REST Endpoints + HTTP
- Web frameworks
- Docker