Incubator Summer Academy - Next Level Data Science

Monday 18 September 2023 - Friday 29 September 2023

Course Packages

Please find an overview of the
- level,
- content,
- prerequisites,
- elements,

- and timeline
of each course package.

The complete program can be found here.

Also beware of our opening event as well as the opportunity to meet the Helmholtz Incubator platforms!
Our lecture on "Fairness in Machine Learning" is also open to individuals who are not taking part in the respective course package (→ Register here ←)

Beginner

Course Package 3 (HIFIS): Python

In this course package, you will develop a fundamental understaning of programming with Python, combined with an introduction to the poopular data processing framework pandas and the popular plotting framework matplotlib.

Prerequisites:
No prior knowledge is required.

Elements:
Introduction to Python (Part 1)
Introduction to Python (Part 2)
Introduction to Pandas
Introduction to matplotlib

→ Register here ←

Course Package 4 (HIFIS): Version Control & Project Management

The workshop provides a solid introduction into the practical usage of the version control system Git in combination with the collaboration platform GitLab.

The following topics will be covered in this workshop:
- Introduction to version control
- Initial Git Setup
- Basic Git workflow
- Feature branch workflow
- Working with the remote repository in GitLab

- Collaboration using GitLab Issues and GitLab Merge Requests

Prerequisites:
This is an introductory workshop. No special knowledge is required.

Elements:
Introduction to Git and GitLab (Part 1)
Introduction to Git and GitLab (Part 2)

→ Register here ←

Course Package 6 (HMC): Introduction to Scientific Metadata

In this course package we will look at the intricate relationship between (digital) research data, metadata and knowledge, discuss why metadata is critical in today's research, as well as explain some of the technologies and concepts related to structured machine-readable metadata.

Have you ever struggled to make sense of scientific data provided by a collaborator - or even understanding your own data 5 months after publication? Do you see difficulties in meeting the data description requirements of your funding agency? Do you want your data to have lasting value, but don't know how to ensure that?

Precise and structured description of research data is key for scientific exchange and progress - and also for the recognition of your effort in data collection. The solution: make your data findable, accessible, interoperable and reusable by describing them with metadata.

Prerequisites:
No prior knowledge is required.

organized by HMC Hub Information & HMC Office

Elements:
Fundamentals of Scientific Metadata (Part 1/2)
Fundamentals of Scientific Metadata (Part 2/2)
Metadata Consulting

→ Register here ←

Course package 8 (Helmholtz AI): Introduction to Machine Learning

This course will introduce participants to the concepts of AI and Machine Learning, covering clustering and clasifications fundamentals as well as practical experience with standard methods for both techniques. Lastly, participants will gain an insight on best practises for evaluating a machine learning model's performance (ROC curve, FPR etc.)

Prerequisites:
Our courses require basic python knowledge as e.g. covered by this course: https://swcarpentry.github.io/python-novice-inflammation/

Elements:
Introduction to Machine Learning Part 1
Introduction to Machine Learning Part 2
Introduction to Machine Learning Part 3
Introduction to Machine Learning Part 4
Lecture: Fairness in machine learning

→ Register here ←

Course Package 10 (Helmholtz AI): Introduction to Deep Learning

This is an hands-on introduction to the first steps in Deep Learning, intended for researchers who are familiar with (non-deep) Machine Learning.

The use of Deep Learning has seen a sharp increase of popularity and applicability over the last decade. While Deep Learning can be a useful tool for researchers from a wide range of domains, taking the first steps in the world of Deep Learning can be somewhat intimidating.

We start with explaining the basic concepts of neural networks, and then go through the different steps of a Deep Learning workflow. Learners will learn how to prepare data for deep learning, how to implement a basic Deep Learning model in Python with Keras, how to monitor and troubleshoot the training process and how to implement different layer types such as convolutional layers.

Prerequisites:
Our courses require basic python knowledge as e.g. covered by this course: https://swcarpentry.github.io/python-novice-inflammation/

Elements:
Introduction to Deep Learning Part 1
Introduction to Deep Learning Part 2
Introduction to Deep Learning Part 3

→ Register here ←

Course Package 7 (Helmholtz Imaging): Machine Learning-based Image Analysis

The success of Machine Learning has revolutionized the field of medical image analysis in the past 5 years. This talk will give an introduction to relevant concepts in machine learning with a focus on computer vision. Subsequently, several example applications in the biomedical domain will be discussed to study the current state of research and the associated challenges and opportunities. We will end the package with a hands on tutorial in which we go step by step through using

nnU-Net from training to visualization of the resulting 3D-segementation maps.

Prerequisites:
No prior knowledge is required.

Elements:
1. Introduction to Machine Learning-based Image Analysis
2. Common Pitfalls in ML-based Image Analysis
2. Applications of AI in Medical Imaging
3. Introduction to nnU-Net and Hands-on Tutorial on how to train and apply nnU-Net (using google colab)

→ Register here ←

Course Package 14 (Helmholtz Imaging): Rendering 3d Datasets

This seminar will cover the basics of volumetric data rendering. We will discuss existing tools for displaying datasets in 3D. To simplify the installation and execution of these applications, we will provide a catalog for launching volumetric rendering applications. This catalog is based on Album, a framework developed by Helmholtz Imaging for running existing applications in specific setups. In the last part of the seminar, we will convert 3D segmentations into meshes, import them into Blender, and go through some basic techniques on how to use Blender to achieve effective volumetric data renderings. Participants are welcome to bring their own 3D datasets, ideally in TIFF format.

Prerequisites:
None

Elements:
Lecture: Rendering 3d Datasets

→ Register here ←

Intermediate

Course Package 11 (Helmholtz Imaging): Regularization in Image Reconstruction: From Model to Data Driven Methods

In this course, we are going to provide the participants with knowledge about the typical mathematical tasks and caveats of image reconstruction problems. This covers advanced forward models and uncertainty, regularizing the reconstruction in order to prevent artifacts caused by noisy data and model errors, and eventually computational tasks. The participants will get the chance to test different image reconstruction and regularization schemes in the hands-on tutorial session.

Prerequisites:

Basic Knowledge in Coding, Basics of Image Reconstruction

Elements:
Lecture: Regularization in Image Reconstruction: From Model to Data Driven Methods
Tutorial: Regularization in Image Reconstruction: From Model to Data Driven Methods

→ Register here ←

Course Package 12 (Helmholtz AI); Introduction to Statistical Learning

The course package covers foundations and recent advances of machine learning techniques, including:
● Basic concepts: Linear regression, nearest neighbour, parametric vs. non-parametric methods, Bayesian classifiers, the curse of dimensionality, model accuracy, bias-variance trade-off
● Linear classifiers: linear regression for classification (discriminative model), linear discriminant analysis (generative model)
● Nonlinear classifiers with Ensemble learning: Decision trees, random forests, boosting
● Unsupervised learning: Gaussian mixture models, k-means

Our course aims to provide participants with not only a theoretical foundation, but also the practical skills needed to use and develop effective machine learning solutions to a wide variety of problems. We illustrate the use of the models in the tutorial throughout the course with methods implemented in Python.

Prerequisites:
The participants are expected to know linear algebra and multivariate calculus, basic concepts from linear functional analysis and basic concepts in probability theory. We ask all participants to be able to run Python (>3.5) within a Jupyter notebook on their computer or use google colab notebook.

Elements:
Lecture: Introduction to Statistical Learning (Part 1)
Tutorial: Statistical Learning (Part 1)
Lecture: Introduction to Statistical Learning (Part 2)
Tutorial: Statistical Learning (Part 2)

→ Register here ←

Course Package 13 (Helmholtz Imaging): Machine Learning for Instance Segmentation

The workshop will comprise a 1h lecture on instance segmentation and tracking, followed by a 1h hands-on instance segmentation challenge in colab.

Prerequisites:
Course Package 7 (Helmholtz Imaging): Machine Learning- based Image Analysis
or alternatively
Python programming & basic understanding of deep learning for images

Elements:
Lecture: Instance segmentation and tracking
Tutorial: hands-on instance segmentation challenge in colab

→ Register here ←

Expert

Course Package 5 (HIFIS): Continuous Integration

What is continuous integration and why do we need it? Figure the following scenario:

A team of scientists is working on a little project that takes astronaut data from Wikidata to analyse the time humans spent in space as well as the age distribution of the astronauts. The project quickly gained attraction and a lot of users as well as contributors joined the project. After some time it became hard for the maintainers to ensure new functionality is properly tested. It also frequently happened that contributors followed a different code style or forgot to add license information.

Verifying those criteria manually is tedious and not promising in the long run. This is why the team aims at automating as much as possible to save their valuable time. Luckily, they found a tool called GitLab CI which they can use to automate those tasks.This course provides an introduction to tghis tool.

Prerequisites:
Participants should have basic programming skills, be familiar with the basic operations of Git, and GitLab and ideally already have made some initial experience with a Unix-shell.

Elements:
Introduction to GitLab CI
Using containers in science part 1/2
Using containers in science part 2/2

→ Register here ←

Course Package 9 (Helmholtz AI): Introduction to eXplainable AI and ML Fairness

During this course package participants will get an introduction to the topic of eXplainable AI (XAI). The goal of the course is to help participants understand how XAI methods can help uncover biases in the data or provide interesting insights. After a general introduction to XAI, the course goes deeper into state-of-the-art model agnostic interpretation techniques as well as a practical session covering these techniques. Finally, we will focus on two model specific post-hoc interpretation methods, with hands-on training covering interpretation of random forests and neural networks with imaging data to learn about strengths and weaknesses of these standard methods used in the field.

This technically oriented component of the course package will be complemented by a lecture on fairness in machine learning - a topic that is highly important and relevant when working with any kind of machine learning approaches.

Prerequisites:
XAI:
Participants are expected to possess basic knowledge of machine learning, deep learning, and Python.

Fairness in ML:
No prerequisites needed

Elements:
Introduction to eXplainable AI
Lecture: Fairness in machine learning

→ Register here ←

All levels

Course Package 1 (HIDA/ HIFIS): Reproducibility in Science

This course package introduces you to the topic of Reproducibility in Science. Coming from a metaperspective on the topic to more concrete information on how to make your research more reproducible, you can then roll up your sleeves and participate in our repro hack- the heart of this course package.

Prerequisites:
No prior knowledge with regard to reproducibility needed.
Participants should have basic knowledge of R, Python, or containerized solutions using dockers.
Please note participants need adminstrator rights on their laptop to be able to install software.

Elements:
Introduction to Open Science and Reproducibility
Lecture: Reproducibility in Science
Repro hack

→ Register here ←

Course Package 2 (Helmholtz AI/ HIDA): Entrepreneurship: From Idea to I did

Followed by an introductiory lecture on Lean Methodology for Startups, this course package offers a deep dive into the lean methodology by "Playing Lean", an engaging hands-on "flight simulator" for Lean Startup and innovation. The creators of the game have partnered up with Alexander Osterwalder, inventor of the Business Model Canvas and one of the great minds of Lean Startup. The Playing Lean board game teaches highly valuable lessons on Lean Startup, creates interest in Business Model Canvas, goes deep on the Value Proposition Canvas or running lean with the Lean Canvas.

Prerequisites:
No prior knowledge or skills needed.

Elements:
Lecture: Introduction to Lean Methodology for Startups
Workshop: Playing Lean

→ Register here ←