Machine Learning and Data Science

All information here is provisional. This course is being set up to be delivered for the first time in Spring 2024. It is at a lower level than a course already given on an existing masters programme.

Provisional Lecturers.

- Patrick Dunne (p.dunne12@imperial.ac.uk) Blacket 531a, HEP
- Tim Evans (t.evans@imperial.ac.uk) Huxley 609, Theory

Overall Philosophy of the course. This course will teach you how to use basic data science and machine learning tools to obtain scientific results from data. We will rely on existing implementations of these methods in commonly used python packages, rather than programming them ourselves from scratch. However, using these packages will require strong programming skills and a complete understanding of the 1st and 2nd year python courses.

Course level. FHEQ 6, level of a typical third year physics option. For this year only we will also offer the course as FHEQ 7 (typical of a fourth year physics option) with extra assessment required.

When the course is taught. Term 2, Spring 2024

How the course is taught.

- 10 lectures, one per week.
- 10 weeks of 2 hour peer collaboration work sessions (no demonstrators). Typically these are based around the use of python notebooks.
- 10 one hour rapid feedback sessions, one per week, on that week's notebooks
- 1 mid-term assignment for formative assessment and feedback (mandatory to progress, but no marks)
- Other short question assignments, formative assessment and feedback (no marks)

How the course is assessed. One day practical test (two 3-hour sessions) in lab during summer exam season (may-June 2024). Tasks will involve solving practical problems using python notebooks which are handed in and assessed for marks.

Prerequisites

- Full understanding of 1st and 2nd year python lab or equivalent,
- Computational Physics is **not needed**.

Limit on numbers. The course will be limited in the first development year to 60 students but open as an option to both 3rd and 4th year Physics students.

Provisional Course Outline

Week 1 Data structure, handling and visualisation - pandas, matplotlib, data sanitisation
Week 2 Probability and random variables - Gaussian, Poisson, Binomial
Week 3 How to formulate your problem - Statistics, likelihoods/target functions
Week 4 Fitting - parameter estimation with iminuit, using the language of ML
Week 5 Your first ML model e.g. Classification with kNN (k nearest neighbours) using IRIS dataset Mid Course Mini project for formative assessment and feedback only (no marks)
Week 6 Your second ML model e.g. Regression with SVR (support vector regression)
Week 7 Performance metrics - Types of errors, ROC curves, goodness of fit
Week 8 Decision trees - Basics, why basic is not so good and how to improve them.
Week 9 Introduction to neural networks - Fully connected neural networks
Week 10 Landscape of neural network algorithms - details of convolutional neural networks (CNN)
Week 11 Last week of term is busy with many other tasks so aim to avoid any new material here, Computer lab time, office hours etc