

Foundation Masterclasses 2020-2021 – Set 1

Artificial Intelligence and Machine Learning for Bio-scientists

Mandatory for first-year EASTBIO students (optional for second-years)

Workshop leader: Dr Andreas Kapourani

Institute of Genetics and Molecular Medicine (IGMM) and School of Informatics, University of Edinburgh

Date: 25 March 2021, 10:15-15:40

The EASTBIO Administrator will send the online link (Zoom) to registered participants directly.

Workshop description

Thanks to advances in sequencing technology and vast reductions computational biology has led to an exponential growth of biomedical data. Due to the complexity and stochasticity of biological systems, bespoke computational strategies are required to transform these high dimensional data into scientific knowledge and generate new hypotheses. This masterclass will provide the necessary background on (statistical) machine learning and how the general field of AI can help us tackle pressing problems in biomedicine.

We will start off by providing a high-level introduction to machine learning models, which fall broadly in two categories, supervised (regression, classification) and unsupervised learning (clustering and dimensionality reduction). Due to the inherent stochasticity and noisiness of biological experiments, computational models should take into account the uncertainty associated with the data. We will see how the Bayesian paradigm provides an excellent basis for quantifying uncertainty and incorporating biological prior knowledge in a principled way (by using the rules of probability).

As motivating examples, we will focus on modelling/analysing single-cell RNA-seq and spatial transcriptomics datasets. The masterclass will consist of both theoretical and practical examples by using the R programming language. Basic knowledge of R is assumed.

Material for this masterclass will be made available on this Github page (<https://github.com/andreakapou/eastbio>) a few days before the course.

Learning outcomes

At the end of the course students will:

- Gain a deeper understanding of the field of machine learning.
- Understand how and which machine learning methods can be used to analyse and interpret biological data.
- Be able to perform scRNA-seq and spatial transcriptomics analysis.
- Become familiar with R and software frameworks for analysing scRNA-seq data.

Training Schedule

Timings	Session details
10:15-10:50	Introduction to machine learning
11:00-11:50	Focus on specific (statistical) machine learning methods
11:50-13:15	Lunch break
13:15-13:50	Introduction to scRNA-seq and spatial transcriptomics
14:00-15:30	Programming activities in R
15:30-15:40	Closing remarks

Training will be recorded.

Participation requirements:

All participants are requested to have access to a computer with R installed in advance. A textbook resource for learning R is the **R for Data Science** book (<https://r4ds.had.co.nz/index.html>). This is only recommended for students who want to gain a deeper understanding of coding/performing computational analysis using R.

Training web-page: <http://www.eastscotbiodtp.ac.uk/foundation-masterclasses>

For **further info**, please email enquiries@eastscotbiodtp.ac.uk