

## MLCE Coursework 2025: Data-driven Optimisation of a Bioprocess

### **Problem Definition**

This coursework involves the optimisation of a simulated bioprocess at process scale utilising CHO cells to produce a desired protein. Experimentally, this would involve a resource-intensive screening campaign involving the growth and feeding of cells under precise conditions (temperature, pH, feed amount, cell type, etc.) to maximize the production of a desired product. This coursework offers a simulated method of mapping bioprocess input parameters to a final predicted titre concentration: a measure of cell productivity. Your goal is to create an algorithm to obtain input parameters which maximizes titre concentration of the simulated bioprocess. Since the takeaway of this coursework is on optimising chemical engineering systems, we will operate on the following assumptions:

- Evaluations are expensive - there is a ceiling for resources invested in a realistic optimisation campaign. For the coursework, the resource budget is emulated by restricting the maximum iterations to 15 with a maximum batch size of 5. Only evaluations up to a runtime of 60 seconds will be examined.

Your group's submission will be a single .py file containing i) group details ii) any helper functions or classes, iii) a class containing your optimizer algorithm, and iv) an execution block where your search space is initialised and your algorithm is executed. It is advised to read this handout and MLCE\_Coursework2025\_BatchBO.ipynb thoroughly to understand the submission format and criteria.

### **Materials Provided**

MLCE\_Coursework2025\_BatchBO.ipynb

- This jupyter notebook contains further information on the coursework which includes the objectives and constraints of your algorithm, running the virtual bioprocess experiment, information on imports and packages, and an example of a submission .py file.

MLCE\_CWBO2025 package

- This package contains the scripts for the simulated bioprocess experiment. You are welcome to study the implementation and theory behind the simulated process. An example of use can be observed in MLCE\_Coursework2025\_BatchBO.ipynb.

### **Grading (per group)**

Your group's total score will be based on a report (20%) and BO implementation (80%).

The 2-page report should have the following sections:

- Big picture explanation and intuition behind the algorithm
  - Methodology
  - Pseudocode
  - You are allowed a figure for your algorithm which is not considered in the report length.
  - References are not considered for the length of the report.
- The report should include a pseudocode following the format specified [here](#).
  - The report should also explain the rationale behind the algorithm and in paragraph form the main steps in the algorithm.
  - Please do not use a letter smaller than size 11 (with a decent font: e.g., Arial, Times New Roman, Calibri, Latex font), and margins no smaller than 2cm Top, Bottom, Left, Right.
  - You will be graded based on clarity of communication, creativity of your algorithm and scientific explanation of your methods.

Your implementation will make up the remainder of your grade (80%). Please see MLCE\_Coursework2025\_BatchBO.ipynb to be familiar with how your implementation will be graded.

- Please submit your group's .py file with a name according to the convention: MLCE\_{group\_name}\_BO.py

### **Further Guidance**

- Stochasticity: Do not set random seeds to your algorithm.
- Packages: please restrict to use numpy, scipy, sobol-seq, random, time, and datetime. **Make absolutely sure** that your code runs in [an environment](#) where **only** numpy, scipy, sobol-seq, random, time, and datetime packages are available.