Prof. Pierantonio Facco
May 5th, 2026

MACHINE LEARNING FOR PROCESS ENGINEERING

HOMEWORK #1

This is **individual homework**: students **MUST complete the homework in a totally independent manner**.

Objective, case study and available dataset

A multinational in the biopharmaceutical sector would like to build a soft sensor for a small-scale *batch* process which produces monoclonal antibodies. Historical data are delivered in the file `dataset.mat` where you can find:

- `X3Dc` [34×11×10] three-dimensional calibration dataset of 11 process variables collected once a day and recorded in 10 time instants (namely, 10 days) for 34 reference batches which define the normal operating conditions (NOC);
- `Yc` [34×1] end-point titer of the monoclonal antibody for the calibration batches;
- `X3Dv` [2×11×10] three-dimensional validation dataset of the process variables for 2 validation batches;
- `Yv` [2×1] end-point titer for the 2 validation batches.

The names and measurement units of both process and product quality variables measured in the cell culture are reported in Table 1.

Questions:

1. data visualization for both calibration datasets **X** and **Y** and discussion;
2. build a PLS model for the prediction of the end-point titer from the time trajectories of the process variables (provide the PLS model in the structure `PLSm`):
   a. discuss the scaling and unfolding strategy;
   b. discuss the model structure: selected number of LVs and explained variances for both **X** and **Y** (provide the PLS model table in the matrix `PLStable`);
3. plot and discuss critically the **X** score plot of LV1 vs. LV2 (provide the scores for all the selected LVs in the matrix `T`);
4. plot and discuss critically the weights for the first LV (provide the weights for all the selected LVs in the matrix `W`) from the engineering point of view;
5. plot and discuss critically the plot of the regression coefficients (provide the regression coefficients in the matrix `B`) from the engineering point of view;
6. verify (and comment) if the linear structure of the PLS model is appropriate through the plot of the scores **T** and **U** of **X** and **Y**, respectively, for LV1;
7. build for the calibration dataset a Q vs. T^2 monitoring chart with the respective 95% confidence limits, and discuss it critically (provide Q and T^2 in vectors `SPE` and `T2`, respectively);

Dr. Pierantonio Facco Department of Industrial Engineering University of Padova Email: pierantonio.facco@unipd.it

8. build the matrices of the residuals **E** and **F**, and discuss the matrices critically (provide the residuals **E** and **F** in the matrices `E` and `F`);
9. compute the mean relative error MRE $\frac{|y-\hat{y}|}{y}$ for the calibration **Y** matrix and discuss it critically (provide the MRE in the variable `MREc`);
10. plot and discuss the parity plot in calibration;
11. project the validation batches into the PLS model, estimate the quality variables $\hat{y}$, calculate the errors of estimation, compute the MRE in validation and discuss them with respect to the variability of the real measurements (provide the estimations $\hat{y}$, the errors $e = y - \hat{y}$, and the MRE in the matrices `ypredv`, `ev` and in `MREv`);
12. plot and discuss the parity plot in validation;
13. discuss the projection of the validation batches in the Q vs. T^2 monitoring chart built in point 7;
14. discuss critically the validation batches for both prediction performances and their position in the Q vs. T^2 monitoring chart; if either Q or T^2 or both are out of the confidence limits, build the contribution plots to understand what variables time trajectories and what instants deviate from the NOCs;
15. if `LV` is the number of latent variables selected in the PLS model, what happens to the prediction performance if a total number of latent variables `LV+3` is selected?

Table 1. *List of the* (a) *process variables measured online and* (b) *quality variables laboratory measurements during the process for the production of the immunoglobulins.*

(a)

column	process variable	units
1	ammonium	mmol
2	viable cell concentration	$\cdot 10^6$ cells/mL
3	cell culture viability	%
4	dissolved oxygen	%
5	sparge oxygen flow rate	mL/min
6	glucose	g/L
7	glutamate	mmol
8	glutamine	mmol
9	lactate	mg/L
10	pH	-
11	volume normalised sparge oxygen flow rate	vvm

(b)

quality variable	units
product titre	mg/L

Deadline:

- May 20th 2026, h. 17.00.

Deliverable:

- send by **email** to:
  pierantonio.facco@unipd.it

  and to:

  edoardo.tamiazzo@phd.unipd.it

  - email subject: "MLfPE homework 1 – surname and family name of the student"

Dr. Pierantonio Facco Department of Industrial Engineering University of Padova Email: pierantonio.facco@unipd.it

- a `.pdf` file `surname_familyname_homework1_MLfPE.pdf` of **maximum 10 pages** (written in Times New Roman, 12 pt with line spacing 1.5) with the responses to all the questions including all the necessary figures and the tables;
- a `surname_name.m` file with the Matlab® code of the provided solution;
- a `surname_name.mat` file with the required numeric solutions.

Homework evaluation:

- correctness and completeness of the provided solution;
- conciseness and clearness of the presentation.

Dr. Pierantonio Facco Department of Industrial Engineering University of Padova Email: pierantonio.facco@unipd.it