PROFESSIONAL EDUCATION
Short Programs

Bioprocess Data Analytics and Machine Learning
July 13 – 15, 2020

Day One

Time	Topics	Instructor(s)	Readings/ Assignments
9 to 10:30	Introduction to bioprocess data analytics: opportunities, types of data analytics problems; supervised, unsupervised, and partially supervised learning; data visualization (software tools using data from bioprocess time series, Raman spectroscopy, spectral imaging, LC-MS)	Braatz/ Anthony	Lecture notes
10:30 to 11	Coffee and networking		
11 to 12:30	Regression: least squares, response surface methodology, ridge regression, lasso, elastic net, cross-validation, residual analysis, outlier detection, uncertainty quantification, optimal experimental design, feature engineering, nonlinearities, systematic method selection, monoclonal antibody manufacturing industrial case study	Braatz	Lecture notes
12:30 to 1:30	Lunch and networking		
1:30 to 2:30	Tour of biopharmaceutical manufacturing labs at MIT, which includes fully automated and instrumented end-to-end continuous CHO-based monoclonal antibody production	Braatz	Lecture notes
2:30 to 3:30	Tips and traps: correlation vs. causation, inferences drawn from matching statistics, effects of feedback loops, selection among too many models, studying features for biological/ chemical/physical reasonableness, Google Flu Trends	Braatz	Lecture notes
3:30 to 5 pm	Hands-on activities in visualizing and analyzing real biopharmaceutical process datasets	Anthony/ Braatz	Students use their own laptops
5 to 7	Course reception/networking		

Day Two

Time	Topics	Instructor(s)	Readings/ Assignments
9 to 10:30	Time series analysis: low-, high-, and band-pass filters; oscillatory data; Fourier transforms; ARMA; ARMAX; noise; AD convertors; sampling; aliasing; statistical process control (aka control charts), UV absorption for protein concentration measurement case study	Yoon/ Anthony	Lecture notes
10:30 to 11	Coffee and networking		
11 to 12:30	Latent variable methods I: PCA, multivariable statistical process control, spectral sensor calibration, PCR, PLS, spectral data artifacts, Raman spectroscopy case study	Yoon	Lecture notes
12:30 to 1:30	Lunch and networking		
1:30 to 3	Latent variable methods II: CCA, FDA, missing data, sparse models, Raman and near-infrared case studies	Yoon	Lecture notes
3 to 3:30	Coffee and networking		
3:30 to 4:30	Hands-on activities in visualizing and analyzing real biopharmaceutical process datasets	Yoon/Braatz/ Anthony	Students use their own laptops
4:30 to 5:30	Open discussions with the instructors	Yoon/Braatz/ Anthony	
5:30 to 6	Coffee and networking		

Day Three

Time	Topics	Instructor(s)	Readings/ Assignments
9 to 10:30	Big data analytics: real-time video, thermal imaging, tensor data, hyperspectral imaging, LC-MS, multiway methods, multilinear subspace learning, lyophilization case studies	Anthony/ Braatz	Lecture notes
10:30 to 11	Coffee and networking		
11 to 12:30	Nonlinear analytics: support vector machines; random forests; comparisons to FDA and k-nearest neighbor classification; feature engineering revisited; kernel methods; hybrid models; autoassociative and recursive neural networks	Anthony	Lecture notes

12:30 to 2:30	Course lunch and networking		
2:30 to 4 pm	What can bioprocess analytics learn from other industries?	Anthony	Lecture notes
4 to 4:30	Coffee and networking		
4:30 to 5:30	Hands-on activities in visualizing and analyzing real biopharmaceutical process datasets	Anthony/ Braatz	Students use their own laptops
5:30 to 6:30	Closing reception	Anthony/ Braatz	