Bioprocess Data Analytics and Machine Learning
June 28–30, 2021

Day One

Time	Topics	Instructor(s)	Readings/Assignments
9 to 10:30	Introduction to bioprocess data analytics: opportunities, bioprocess data types including Raman, hyperspectral imaging, and LC-MS, types of data applications; taxonomy of data analytics and machine learning methods, course overview	Braatz	Lecture notes
10:30 to 11	Break		
11 to 12:30	Regression: least squares, response surface methodology, ridge regression, lasso, elastic net, cross-validation, model accuracy quantification, feature engineering, nonlinearities, monoclonal antibody manufacturing industrial case study	Braatz	Lecture notes
12:30 to 1:30	Networking		
1:30 to 3:00	Tips and traps: Google Flu Trends, correlation vs. causation, misuse of matching statistics, effects of feedback loops, false correlations in microarray data, artifact removal during data preprocessing, assessing features for reasonableness	Braatz	Lecture notes
3:00 to 3:30	Break		
3:30 to 4:30	Open discussions with the instructors	Anthony/Braatz	
4:30 to 6	Networking		

Day Two

Time	Topics	Instructor(s)	Readings/ Assignments
9 to 10:30	Data analytics/machine learning method selection, automated machine learning, the pitfall of selecting from many models/methods by cross-validation, data interrogation, algebraic learning, industrial application to monoclonal antibody manufacturing	Braatz	Lecture notes
10:30 to 11	Break		
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	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	Break		
3:30 to 4:30	Visualization and analysis of real biopharmaceutical process data	Yoon	
4:30 to 5:30	Open discussions with the instructors	Yoon/Braatz/ Anthony	
5:30 to 6:30	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	Braatz	Lecture notes
10:30 to 11	Break		
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 1:30	Networking		
1:30 to 3 pm	What can bioprocess analytics learn from other industries?	Anthony	Lecture notes
3 to 3:30	Break		
3:30 to 4:30	Open discussions with the instructors	Anthony/ Braatz	
4:30 to 5:30	Closing	Anthony/ Braatz	