DATA SCIENCE FOR CHEMICAL ENGINEERING EDUCATION

S. Joe Qin, Fluor Professor
Dept. of Chemical Engineering and Materials Science
Ming Hsieh Department of Electrical and Computer Engineering
University of Southern California

【Current】 Chair Professor of Data Science
City University of Hong Kong, Hong Kong
joe.qin@cityu.edu.hk

Nov. 10-15, 2019
AIChE Annual Meeting
Orlando, FL

CHE 586 at USC: Process Data Analytics and Machine Learning

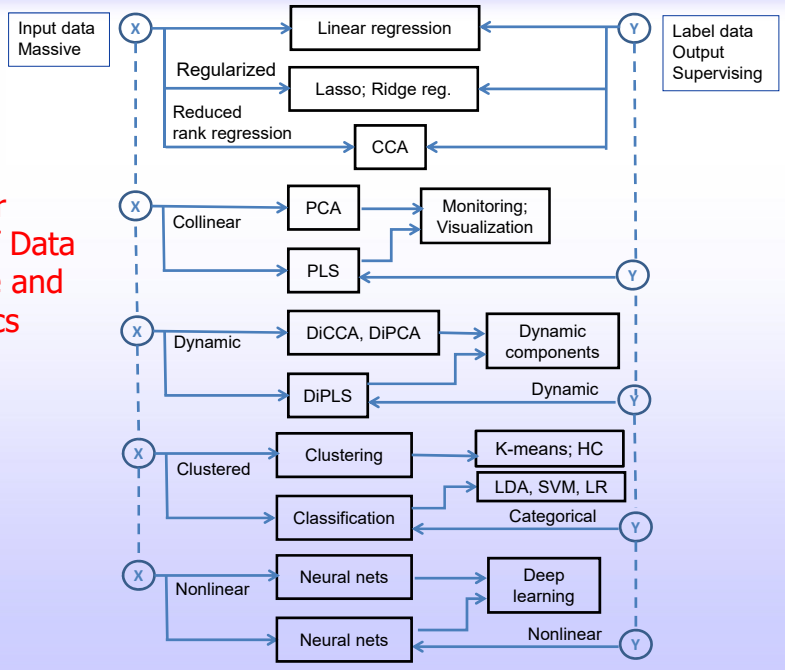
- Offered for the past three years to MS students; could be an elective for BS students
  - Topics include multi-linear regression, supervised learning, unsupervised learning, principal component analysis, partial least squares, canonical correlation analysis, clustering methods, lasso, neural networks, and deep learning.
  - Applications include analysis of chemical process data, quality data, and indirectly measured data.
  - Domain of application: industrial manufacturing data, process operation data, and data from other engineering disciplines.

Focus

- Emphasize on a *data-driven* and *linear algebraic approach* over a probabilistic approach in this course.
- Gain programming experience in R
- Textbook: An Introduction to Statistical Learning with Applications in R. Authors: James, G., Witten, D., Hastie, T., Tibshirani, R. (2013), Springer

	Topics/Daily Activities
Week 1 1/11/2019	L1. Introduction to big data, machine learning, and process data analytics
Week 2 1/18/2019	**L2. Linear regression**
Week 3 1/25/2019	L3. Principal component analysis (PCA)
Week 4 2/1/2019	**PCA**
Week 5 2/8/2019	**L4. Clustering**
Week 6 2/15/2019	**L5.** PCA for process monitoring
Week 7 2/22/2019	**L6. Fault Diagnosis**
Week 8 3/1/2019	**L6. Partial least squares**
Week 9 3/8/2019	**L7. Lasso, Ridge regression**
3/15/2019	**SPRING BREAK**
Week 10 3/22/2019	**Mid-term**
Week 11 3/29/2019	kernel methods
Week 12 4/5/2019	support vector machines
Week 13 4/12/2019	neural networks and soft sensors
Week 14 4/19/2019	deep learning
Week 15 4/26/2019	**Project report due**

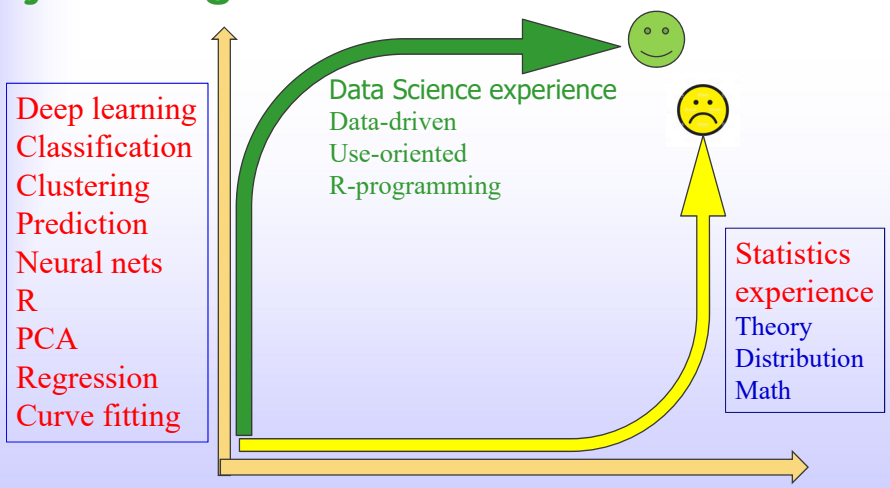
Another View of Data Science and Analytics



Statistics for Chemical Engineers Course: Sequence for Comparison

- Introduction to Probability and Statistics
- Data Collection and Analysis
- Discrete Probability Models
  - Basic probability concepts; Binomial and Poisson distributions
  - Multivariate distributions
- Continuous Probability Models
- Statistical Inference: Sampling Distribution, Confidence Intervals, p-values, and Tests of Hypotheses
- Statistical Quality Control/Statistical Process Control
- Design of Experiments with One Factor
- Design of Experiments with Two or More Factors
- Regression Analysis –
  - simple linear regression; multilinear regression

Data Science vs. Statistics experience: day and night for the student



What is Data Science Editor: Xiao-Li MENG

HDSR
HARVARD DATA SCIENCE REVIEW
A Microscopic, Telescopic, and Kaleidoscopic View of Data Science

- **Indeed, what exactly is *data science* (DS)?**
  "… the answer depends on whom you ask. Some say DS is CS (computer science). Others think DS is simply S (statistics). You may even run into someone who declares DS is just hyped-up BS (and I don't mean "Bayesian statistics")."
- **Joe Qin: Could Data Science emerge as a discipline like control theory in engineering?**
  - **What are the PID's of data science for engineers? (Qin and Chiang, 2019 overview paper)**

Essence of Data Science in ChE

- **Prediction**: Predictive Analytics
  - Predict critical but hard to measure variables from others
- **Interpretation**: Feature or knowledge discovery from data
  - Visualization
  - Diagnostics
- Decision Making

My PID for Data Science

Predict-Interpret-Decide (PID)