Introduction to Statistical Learning and Personalized Medicine  
Spring, 2018

• COURSE DESCRIPTION (3 credit hours)  
The first part of the course gives an introduction to statistical learning methods, including a complete review of supervised learning methods (discriminant analysis, kernel methods, nearest neighborhood, tree methods, neural network, support vector machine, random forest, and boosting methods) and unsupervised learning methods (principal component analysis, factor analysis, cluster analysis, multidimensional scaling, self organizing map). R-functions and real data demo are used for illustration. It also includes learning theory for supervised learning methods such as Bayesian error, concentration inequalities, VC-theory, risk bound and etc.  
The second part of the course focuses on recent development of statistical methods for personalized medicine, with particular emphasis on using statistical learning methods. This part starts with potential outcome framework and concepts of dynamic treatment regimes, discusses the use of observational studies and sequentially randomized trials for this context, then introduces the methods based on reinforcement learning, Q-learning, A-learning, G-computation, and O-learning for optimal dynamic treatment regimes in personalized medicine.

• MEETING TIME  TTH 9:30-10:45 @MHRC 003

• CLASS WEBSITE  http://www.bios.unc.edu/~dzeng/Bios740.html

• LECTURE NOTES AND TEXTBOOKS
  – The Elements of Statistical Learning: Data Mining, Inference, and Prediction, by Hastie, Tibshirani and Freedman et al.  
  – Statistical Reinforcement Learning, by Sugiyama.
  – Support Vector Machines, by Steinwart and Christmann.
  – A Probability Theory of Pattern Recognition, by Devroye, Gyorfi and Lugosi.

• INSTRUCTOR
Dr. Donglin Zeng  
Office: 3103B McGavran-Greenberg Building  
Email: dzeng@email.unc.edu  
Phone: (919)966-7273

• GRADING SYSTEM
There will be no homework or exams for this course. A final project, which can be research-oriented work, paper review, or real data application, will be required. Final grades will be based on the evaluation of the final project and a final presentation. For this project, students are encouraged to meet the instructor to discuss choices during the semester.

• LIST OF LECTURES
Lecture 1. Introduction to Statistical Learning
Lecture 2. Bayes Rule and Bayes Error
Lecture 3. Direct Rule Learning: Linear Regression
Lecture 4. Direct Rule Learning: Linear Classification
Lecture 5. Direct Rule Learning: Semi-Nonparametric Methods
Lecture 6. Indirect Rule Learning: Support Vector Machines
Lecture 7. Neural Network and Deep Learning
Lecture 8. Learning Theory: Consistency, Risk Bound and Model Selection
Lecture 9. Unsupervised Learning Donglin
Lecture 10. Reinforcement Learning
Lecture 11. Special Topic: Support Vector Hazard Regression (SVHR) for Survival Outcomes
Lecture 12. Special Topic: Multicategory Learning
Lecture 13. Introduction to Personalized Medicine
Lecture 14. Causal Inference Framework for Constant Treatment Strategies
Lecture 15. Special Topic: Infer Treatment Effect under Crossover in The Panitumumab Study
Lecture 16. Dynamic Treatment Regimes
Lecture 17. Sequential Multiple Assignment Randomized Trials
Lecture 18. Special Topic: Improve SMART with EnRichment
Lecture 19. Q-Learning for Optimal Dynamic Treatment Strategies
Lecture 20. Outcome-Weighted Learning for Static Treatment Strategies
Lecture 21. Learning Optimal Dynamic Treatment Regimes
Lecture 22. Special Topic: Optimal Treatment Strategy for Improving Survival

***The remaining classes are reserved for student presentations.