Machine Learning & Advanced Analytics for Biomedicine

Winter 2020

Course Registration Info

- **Subject/Course #:** CCTS40500
- **Section Number:** 01
- **Course Name:** Machine Learning & Advanced Analytics for Biomedicine
- **Units:** 100 units
- **Instructor/s:** Ishanu Chattopadhyay (ishanu@uchicago.edu)
- **Day/Time:** M 1730-2030
- **Prerequisites/Remarks:** Prior basic familiarity to coding in python or R is required
- **Is the course undergrad, graduate, or mixed level:** Mixed Level
- **Cross list:** BIOS 29208

Course Description

Easy availability of data is rapidly transforming scientific research, and advanced analytics powered by sophisticated learning algorithms is uncovering new insights in complex open problems in biology and biomedicine. The goal of this course is to provide an introductory overview of the key concepts in machine learning, outlining the potential applications in biomedicine. Beginning from basic statistical concepts, we will discuss concepts and implementations of standard and state of the art classification and prediction algorithms, and go on to discuss more advanced topics in unsupervised learning, deep learning architectures, and stochastic time series analysis. We will also cover emerging ideas in data-driven causal inference, and demonstrate applications in uncovering etiological insights from large scale clinical databases of electronic health records, and publicly available sequence and omics datasets.

Learning Goals

The acquisition of hands-on skills will be emphasized over machine learning theory. On successfully completing the course, students will have acquired enough knowledge of the underlying machinery to intuit and implement solutions to non-trivial data science problems arising in biology and medicine. Rudimentary knowledge of probability theory, and basic exposure to scripting languages such as python/R is required.

Planned Assessments

Assignments will generally require modeling problems in R or Python. There will be occasional homeworks, one mid-term assignment, and one final project. The mid-term will be a Kaggle In-Class Competition (https://www.kaggle.com/about/inclass/overview). Final assessment will depend on the performance on the assignments.

Tentative Syllabus

Alternating Colors indicate breakdown in weeks (10 WEEKS)

1) Introduction to Automated Inference, Machine Learning, Probability Theory, & Statistical Modeling of Data
2) Linear Regression with One Variable
3) Review of Linear Algebra Basics
4) Linear Regression with Multiple Variables
5) Logistic Regression
6) The MCMCglmm R Library
7) Concepts of Overfitting, & Regularization
8) LASSO and Ridge Regression Analysis
9) Introduction to Problem Archetypes: Classification & Regression
10) The SkLearn Python Library
11) Support Vector Machines & Support Vector Regression
12) Decision Trees
13) Random Forests & Extremely Randomized Trees in Classification & Regression
14) Boosting Algorithms: The CatBoost Implementation
15) Neural Nets: Multi Layer Perceptrons
16) Convolutional Neural Nets (CNN) in Image Classification
17) Introduction to the Tensorflow Library for CNNs
18) Recurrent Nets (RNN) and Long Short-term Memory (LSTM) Architectures
19) Modeling Insights In Practical Problem Solving in Biosystems
20) Introduction to Stochastic Processes in Biological and Clinical Context
21) Time Series Modeling for Biological & Clinical Signal Intelligence
22) Unsupervised Inference & Zero-knowledge Learning for Data-driven Model Discovery
23) Clustering & Dimensionality Reduction Algorithms
24) Anomaly Detection In the Clinical Context
25) ML-enabled Tools In Precision & Personalized Medicine

TEXTS

[1] C. BISHOP, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, 2006.
[2] I. GOODFELLOW, Y. BENGIO, A. COURVILLE, AND F. BACH, *Deep Learning*, MIT Press, 2016.
[3] K. MURPHY AND F. BACH, *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machi, MIT Press, 2012.
[4] G. VAROQUAUX, *Scikit-learn tutorial: statistical-learning for sientific data processing*. http://gael-varoquaux.info/scikit-learn-tutorial/, 2010.