

Machine Learning Boot Camp

Analyzing Biomedical and Health Data



Day 1: Thursday June 6, 2019

7:45 – 8:00	Check-in & Breakfast for those attending Optional Session	
8:00 – 9:00	Optional Session: Intro to R	Cody Chiuзан, PhD
9:00 – 9:15	Check-in & Breakfast for full group, and Room Transition	
9:15 – 9:30	Welcome and Introduction	Cody Chiuзан, PhD
9:30 – 10:00	Session 1: Low vs High-Dimensional Data (Lecture) <ul style="list-style-type: none"> Problems Challenges Approaches 	Cody Chiuзан, PhD
10:00 – 10:30	Session 2: Classical vs. New Methodologies (Lecture) <ul style="list-style-type: none"> Differences, assumptions, applications 	Cody Chiuзан, PhD
10:30 – 10:45	Break / One-on-one questions	
10:45 – 12:00	Session 3: Penalized Regression Methods (Lab) <ul style="list-style-type: none"> Ridge, Lasso, Elastic Net Regression 	Cody Chiuзан, PhD
12:00 – 1:00	Networking Lunch	11 th Floor, Room 1101
1:00 – 1:45	Session 4: Penalized Regression Methods (Lab) <ul style="list-style-type: none"> Model fitting Feature selection, validation, overcome overfitting 	Cody Chiuзан, PhD
1:45 – 2:45	Session 5: Non-Linear Regression (Lecture + Lab) <ul style="list-style-type: none"> Splines and local regression GAM and MARS 	Yifei Sun, PhD
2:45 – 3:00	Break / One-on-one questions	
3:00 – 4:45	Session 6: Classification Models (Lecture + Lab) <ul style="list-style-type: none"> Loss-based Classification (e.g. Logistic Regression and Support Vector Machines (SVM)) Validation and Calibration 	Noah Simon, PhD
4:45 – 5:00	Questions and Wrap-Up	



COLUMBIA

MAILMAN SCHOOL
OF PUBLIC HEALTH

COLUMBIA UNIVERSITY
IRVING INSTITUTE FOR CLINICAL
AND TRANSLATIONAL RESEARCH

722 W. 168th Street, New York, NY 10032 5th Floor, Room 532 A/B - June 6-7, 2019

The Machine Learning Boot Camp is hosted by the SHARP Training Program (Skills for Health and Research Professionals) and Department of Biostatistics at the Columbia University Mailman School of Public Health, and the Irving Institute for Clinical and Translational Research: Biostatistics, Epidemiology, and Research Design (BERD) Educational Resource.

Machine Learning Boot Camp

Analyzing Biomedical and Health Data



Day 2: Friday June 7, 2019

8:00 – 8:30	Breakfast	
8:30 – 8:45	Recap from Day 1: Group Questions & Answers	
8:45 – 10:00	Session 7: Tree Based Methods (Lecture) <ul style="list-style-type: none">• Decision/Regression Trees• Ensembling (e.g. Bagging, Random Forests, and Boosting)	Noah Simon, PhD
10:00 – 10:15	Break / One-on-one questions	
10:15 – 11:00	Session 8: Tree Based Methods (Lab)	Noah Simon, PhD
11:00 – 12:00	Session 9: Model Interpretation (Lecture + Lab) <ul style="list-style-type: none">• Visualization• Local Interpretable Model-Agnostic Explanations	Yifei Sun, PhD
12:00 – 1:00	Networking Lunch	11 th Floor, Room 1101
1:00 – 2:15	Session 10: Clustering Algorithms (Lecture + Lab) <ul style="list-style-type: none">• K means• Hierarchical clustering	Yifei Sun, PhD
2:15 – 3:30	Session 11: Principal Component Analysis (PCA) (Lecture + Lab) <ul style="list-style-type: none">• PCA• Principal component regression and partial least squares	Yifei Sun, PhD
3:30 – 3:45	Break / One-on-one questions	
3:45 – 4:45	Session 12: Deep Learning (Lecture) <ul style="list-style-type: none">• Brief Tutorial• Pros/Cons of Popular Approaches (e.g., neural, convolutional, recurrent networks)	Noah Simon, PhD
4:45 – 5:00	Evaluation, Questions, Wrap-Up	



COLUMBIA

MAILMAN SCHOOL
OF PUBLIC HEALTH

COLUMBIA UNIVERSITY
IRVING INSTITUTE FOR CLINICAL
AND TRANSLATIONAL RESEARCH

722 W. 168th Street, New York, NY 10032 5th Floor, Room 532 A/B - June 6-7, 2019

The Machine Learning Boot Camp is hosted by the SHARP Training Program (Skills for Health and Research Professionals) and Department of Biostatistics at the Columbia University Mailman School of Public Health, and the Irving Institute for Clinical and Translational Research: Biostatistics, Epidemiology, and Research Design (BERD) Educational Resource.