Fall 2020
Department of Statistics
3470: 484 Introduction to Machine Learning

Time and Place: www 100% asynchronous
Instructor: Dr. Nao Mimoto
Office: CAS 413
Phone: (330) 972 8011
Email: nmimoto@uakron.edu
Office Hours: TTh 2:00-3:00 and by appointment
Class Web Page: https://nmimoto.github.io/483
WebEx page: https://uakron.webex.com/meet/nmimoto
Prerequisite: 3470:461 Applied Statistics or equivalent

Textbook: James, G. et. al. / Introduction to Statistical Learning with Application in R. Springer. (Available on web)

Course Description: Prerequisite: 3470:461 or equivalent. Statistical Learning. Bias-Variance trade-off. Generalized, Logistic, Ridge Regression. Lasso. Linear Discriminant Analysis. K-fold Cross-Validation. Neural Network. Support Vector Machine. Principal Component Analysis. K-means and Hierarchical Clustering.

Learning Objectives: Students who successfully completing this course will:

(1) Understand the basic concepts of popular machine learning methods.
(2) Understand the basic concepts behind model improvement.
(3) Understand the basic concepts of modern model assessment and validation techniques.
(4) Be able to implement above task to real-life dataset using software R.

Office Hours (WebEX): Tuesdays and Thursdays 2-3pm, and by appointment.
My WebEX room link: https://uakron.webex.com/meet/nmimoto

Statistical Software R: We will use a statistical software called Rstudio. You can download for free, and install on your Windows, Mac or Linux machine. See instruction posted on class

web page for more details. Students are not assumed to have prior knowledge of the software.

Online Lecture Videos: Lecture videos will be uploaded on Brightspace.

In-class Assignments: There will be three in-class assignments toward the beginning of the semester. Students will model and analyze the dataset using the software R. This is intended to bring all students up to pace with R coding necessary. Instructor's help will be available on site. If students can not complete the assignment during the class time, it will be due on next lecture.

Modeling Assignments: There are two modeling assignment toward the end of the semester. Students will be given real-life datasets and asked to perform various model fitting and diagnostic techniques, as well as to show understanding of their theoretical implications and limitations within a practical context. Students will submit a written report with R code and figures.

Grading Scale: Grading is based on homework and exam scores according to the following schemes:

On-line quizzes (9)	40%
In-class Assignments (3)	30%
Modeling Assignments (2)	30%

A	B	C	D
90%	80%	70%	60%

A plus/minus grade may be assigned if a student's total fall just below one of the cut-offs.

Academic Honesty: Academic honesty is fundamental to the activities and principles of any university. All members of an academic community must be confident that each person's work has been responsible and honorably acquired, developed and presented. The assumption that your work is a fair representation of your actual ability, knowledge and skills form the basis of institutional quality and the quality of its graduates.

The University of Akron regards academic dishonesty as a serious matter and an act of academic misconduct, which can cause sanctions to be imposed, such as a failing grade, disciplinary probation, suspension, or dismissal from the University. The University community is governed by the policies and regulations contained within the Code of Student Conduct available at www.uakron.edu/sja. Any form of cheating on exams, projects, cases, or homework assignments can result in any of the actions listed above.

ADA: Any student who feels he/she may need an accommodation based on the impact of a disability should contact the Office of Accessibility at (330) 972-7928.
The office is located in Simmons Hall room 105. Their web page can be viewed at http://www.uakron.edu/access/.

Disclaimer: The instructor reserves the right to make any changes he considers academically advisable. Changes will be announced in class and on the course web site. It is your responsibility to keep up with any changed policies and assignments.

Topics Covered

1. Introduction to R
2. Introduction to Statistical Learning
   Overfitting, Bias-Variance trade-off, Reducible and Irreducible error
3. Linear Regression
   SLR, GLM, Best Subset, Stepwise Selection
4. Classification Models
   Logistic Regression Linear Discriminant Analysis
5. Resampling and Cross-Validation
   Leave-One-Out-Cross-Validation, K-fold Cross Validation, Jackknife
6. Regularization
   Ridge Regression, Lasso
7. Support Vector Machine
   Kernels, tuning parameter
8. Neural Network
   Hidden Layers, Forward and Backward propagation
9. Unsupervised Learning
   Principal Component Analysis, K-means Clustering, Hierarchical Clustering.

Tentative Schedule

Week	Week Of	Tues	Thurs
1	Aug 24	Intro to R	Stats Learning
2	Aug 31	Regression	Regression
3	Sep 7	Classification	**In-Class Assignment 1**
4	Sep 14	Classification	Classification
5	Sep 21	Cross-Validation	Cross-Validation
6	Sep 28	*[President's Day]*	**In-Class Assignment 2**
7	Oct 5	Regularization	Regularization
8	Oct 12	SVM	SVM
9	Oct 19	SVM	SVM
10	Oct 26	Neural Network	Neural Network
11	Nov 2	Neural Network	Neural Network
12	Nov 9	Neural Network	Clustering Analysis
13	Nov 16	Clustering Analysis	Clustering Analysis Lab **(Assignment 4 due)**
14	Nov 23	PCA	*[Thanksgiving]*
15	Nov 30	PCA	PCA
16	Dec 7	**(Assignment 5 due on Wed Dec 9)**	

Important Dates:

- Aug 30: Last day to add without signatures

- Sep 06: Last day to add
- Sep 06: Last day to drop without WD appearing
- Oct 11: Last day to drop