PQHS 471: Machine Learning /Data Mining (Spring 2018)

Instructor: Chun Li

Time	Tuesday/Thursday 2:30 – 3:45 pm
Location	Wood Building WG-73
Office hour	Available through contact; Wolstein Research Building 2528; cxl791@case.edu
Course site ?	https://sites.google.com/a/case.edu/epbi471 ??

General description: This course aims to introduce concepts and major methods in statistical learning, machine learning, and data mining, emphasizing on the statistical aspects of various approaches and on biomedical applications. Specifically, we will cover prediction model building, model regularization (shrinkage, lasso), classification (logistic regression, discriminant analysis, k-nearest neighbors), trees; ensemble methods (random forests, boosting), support vector machines, artificial neural networks (backpropagation, deep learning, CNN, RNN); association rules, k-means and hierarchical clustering, GANs. Basic techniques that are applicable to many of the areas, such as cross-validation, the bootstrap, dimensionality reduction, and splines, will be explained and used repeatedly. Minimum prerequisites are calculus, linear algebra, and some exposure to statistics (PQHG 431).

Books:

book	book title and webpage
ISLR	James et al. (2013) *An Introduction to Statistical Learning, with Applications in R*. Springer. (8th printing at https://link.springer.com/) http://www-bcf.usc.edu/~gareth/ISL/
HOML	Géron (2017) *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly. http://proquest.safaribooksonline.com/9781491962282?uicode=ohlink
NNDL	Nielsen (2015) *Neural Networks and Deep Learning*. http://neuralnetworksanddeeplearning.com/
ESL	Hastie et al. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer. (12th printing) http://www.stanford.edu/~hastie/ElemStatLearn/
CASI	Efron and Hastie (2016) *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press. https://web.stanford.edu/~hastie/CASI/
DL	Goodfellow et al. (2016) *Deep Learning*. MIT Press. http://www.deeplearningbook.org/
MMDS	Leskovec et al. (2014) *Mining of Massive Datasets*, 2nd ed. Cambridge University Press. http://www.mmds.org/
R4DS	Grolemund and Wickham (2017) *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly. http://r4ds.had.co.nz

Course style: Lecture + Discussion

1. Students should read the material to be covered before each lecture. I will randomly call on students to briefly (<1 minute) summarize the material: what is this section about (big picture, methods in general). It is okay if you do not understand some technical details.
2. Students are strongly encouraged to raise questions and participate in discussions.

Course grade: 25% each for

1) homework,
2) midterm (on March 8),
3) final exam (in the week of April 30), and
4) participation (summarize materials and participate in discussions)

PQHS 471 tentative schedule (Spring 2018):

Week	Date	HW/exam	Chapters	Topic
			ISLR	HOML
1	1/16		ISLR 1-2, HOML 1	Introduction; data science (AI, big data); R/git/Python; statistical learning and machine learning in general
2	1/22		ISLR 3, HOML 4?	linear regression; curse of dimensionality
3	1/29		ISLR 4, HOML 3	classification, LDA/QDA, ROC, etc.
4	2/5	HW1 due	ISLR 5-6	cross-validation, bootstrap, subset selection
5	2/12		ISLR 6-7, HOML 4	ridge, lasso, splines
6	2/19		ISLR 7-8, HOML 6	local regression, GAMs, trees
7	2/26	HW2 due	ISLR 8, HOML 7	random forests, boosting
8	3/5	Midterm	ISLR 9, HOML 5	support vector machines
-	3/12	Spring break		
9	3/19		NNDL 1-3, HOML 10	neural networks, backpropagation, model tuning
10	3/26		NNDL 6, HOML 11–14	deep learning, CNN, RNN
11	4/2		ISLR 10, HOML 8	unsupervised learning, PCA, clustering
12	4/9	HW3 due	ESL 14.2, 14.4, 14.8-9	association rules, SOM, MDS
13	4/16			GANs, additional topics
14	4/23			Examples and review
15	4/30	Final exam		

HOML 2 (as HW?), 9 (TensorFlow)

Midterm take home?