DSCI353-353m-453: Class 00a

2001-353-353m-453-00a-f-

Roger H. French, Peitian Wang

14 January, 2020

Contents

1.1.0.1 Reading, Homeworks, Projects, SemProjects

-

1.2 Readings for next class:

   –

-

1.3 Laboratory Exercises:

- SemProjects:

  - SemProjects: have 4 parts, we'll have reports on Sect. 1,2 then 3,4
    * SemProj Report Out #1 in Class w07a,07b, Tues/Thurs, Feb. 25,27
    * SemProj Report Out #2 in Class w10a,10b, Tues/Thurs, March 24,26
    * SemProj Report Out #3 in Class w14a,14b, Tues/Thurs April 21,23
    * SemProj Report #4 is the full, comprehensive project due at final exam.
  - These are Peer Graded, using the Rubric
  - Assistance on SemProjects is done in DSCI352-352m-452 Class
    * DSCI352 meetings during Friday Community Hour, 12:45 to 1:45pm in Olin 303
    * Is taught by Prof. Laura Bruckman (lsh41@case.edu)

- Office Hours: Mondays, Wednesdays 4pm to 5pm in White 540

- Final Exam

  - Thursday March 30th, 2020, 12 noon to 3pm

1.3.0.1 Textbooks

1.3.0.1.1 Introduction to R and Data Science

For students new to R, Coding, Inferential Statistics

- Peng: R Programming for Data Science
- Peng: Exploratory Data Analysis with R
- OIS = Diez, Barr, Çetinkaya-Runde: Open Intro Stat v4

1.3.0.1.2 Textbooks for this class

- R4DS = Wickham, Grolemund: R for Data Science
- ISLR = James, Witten, Hastie, Tibshirani: Intro to Statistical Learning with R
- ESL = Trevor Hastie, Tibshirani, Friedman: Elements of Statistical Learning
- DLwR = Chollet, Allaire: Deep Learning with R

1.3.0.2 Syllabus

1.3.0.3 Topic

Day:Date	Foundation	Practicum	Readings (optional)	Due (optional)
w01a:Tu:1/14/20	Open Data Science	R, Rstudio IDE, Git		(LE0)
w01b:Th:1/16/20	Intro R Markdown	Forking Class Repo	(R4DS-1,2,3)	
w02a:Tu:1/21/20	Statistical Learning	Pred. Analytics	ISLR2	(LE0)
w02b:Th:1/23/20	Data Analytic Style	Tidy Data Manip.	(R4DS-4,5,6)	LE1
w03a:Tu:1/28/20	Lin. Regr.	Pairs Plots	(OIS7)	
w03b:Th:1/30/20	Mult. Lin. Regr.	Test Stats	ISLR3 (R4DS-7,8)	
w04a:Tu:2/4/20	Logistic Regr.	Interaction Terms	ISLR4	**LE1:Due**, LE2
w04b:Th:2/6/20	Classification		(OIS8)	
w05a:Tu:2/11/20*	Resample Cross-Valid.	Cluster Analysis	ISLR5	
w05b:Th:2/13/20	Bootstrap	Steps of Data Analysis	DL1,2 (R4DS9-16)	**LE2:Due**
w06a:Tu:2/18/20	LMS: Subset		ISLR6	LE3
w06b:Th:2/20/20	LMS: Feature Selec.	Coeff. Uncertainties	DL3,4 (R4DS17-21)	
w07a:Tu:2/25/20*	BeyondL: Spline, GAM	**SemProj1-453**	ISLR7	
w07b:Th:2/27/20	MidTerm Review	**SemProj1-3/452**	DL5,6 (R4DS22-25)	**LE3:Due**
w08a:Tu:3/3/20	**MIDTERM EXAM**			
w08b:Th:3/5/20	Dim. Reduc. & PCA		ISLR8 (R4DS26-30)	
Tu:Th:3/9-13/20	**SPRING BREAK**			
w09a:Tu:3/17/20	Regr. Trees	Dec. Trees	ISLR9	LE4
w09b:Th:3/19/20	Bagging, Boosting	How DT Work	ISLR10	
w10a:Tu:3/24/20*	Support Vector Mach.	**SemProj2-453**	ESL11	**LE4:Due**
w10b:Th:3/26/20*	ML Overview, Caret	**SemProj2-3/452**	DLR1	LE5
w11a:Tu:3/31/20*	Neural Networks	MNIST digits	DLR2	
w11b:Th:4/2/20*	NN Topo., Types, Train	ImageNet	DLR3	**LE5:Due**
w12a:Tu:4/7/20*	R-Keras/TensorFlow	CNN w TF	DLR4	LE6
w12b:Th:4/9/20	CNN w/TF	EL Image Sup. ML		
w13a:Tu:4/14/20	CNNs w/small data	DLwR 2.1		**LE6:Due** LE7
w13b:Th:4/16/20	Tboard, TFestimators	pretrained CNNS		
w14a:Tu:4/21/20	R Packaging	**SemProj3-453**		
w14b:Th:4/23/20	Final Exam Review	**SemProj3-3/452**		**LE7:Due**
	FINAL EXAM	**Th. 3/30/2020, 12-3pm**	Nord 356	

Figure 1: IT Fundamentals: Applied Data Science with R, Syllabus