Math380: Stats&Machine Learning - Spring 2018

Meeting Times:
MWF 11:30 AM - 12:20 PM, Sturges 223

Instructor:
Dr. Yusuf Bilgic, South Hall 324A, 245-5484
bilgic@geneseo.edu

My Website for Course:
http://go.geneseo.edu/bilgic (click on Math380-Stat&ML)

Office Hours:
MW 10:30 - 11:20 am, Tue 11:00 - 1:00 pm, and Thursday with appointment

Course Description:
This is a 300-level math topic course in statistics that counts toward electives for math majors. Emphasis is placed on the mathematics behind the theories, methods, algorithms and techniques in Statistics and Machine Learning. Application will be implemented in R/Python. This course serves as an applied learning capstone math course in statistical data and algorithmic modeling as well. The course initially covers estimation theory, statistical and machine learning techniques and algorithms, and then is followed by project meetings in the last third portion of the semester. In the final project, students will work on an individually-chosen real-world data project/problem in group. The project will cultivate a report/manuscript and conference/public talk. Credits: 3(3-0).

Learning Outcomes:
Upon successful completion of Math 380 – Topics in Math (Statistics and Machine Learning), a student will be able to:

- know the basics of estimation theory,
- develop skills for the practices of data statistical and machine learning techniques and algorithms,
- understand the mathematics behind statistical and machine learning techniques and algorithms,
- understand, apply and extend statistical and machine learning techniques and algorithms including linear and nonlinear modeling, supervised and unsupervised learning (classification, LDA, tree-based regression, SVM, K-means, NN), parametric and nonparametric methods, estimation methods, multivariate data visualization, sampling methods, model averaging (bagging, boosting), feature extraction and selection, data reduction, model improvement, and
- perform real-world data analysis in a topic of interest with a technology and programing language, design and write data analysis report, use professional writing tools (like R+Sweave) and present the results.

Texts and Lecture Resources:

- [**ISLR**] (**Main text 1**) An Introduction to Statistical Learning, with Applications in R by James, Witten, Hastie and Tibshirani (Springer, 2013). Download it from http://www-bcf.usc.edu/~gareth/ISL/. Main resource webpage is http://www-bcf.usc.edu/~gareth/ISL/index.html (data, r code etc). For exercise solutions, visit https://github.com/asadoughi/stat-learning. See slides and videos at https://www.r-bloggers.com/in-depth-introduction-to-machine-learning-in-15-hours-of-expert-videos/
- [**ESL**] (**Main text 2** - an advanced treatment of the topics found in ISLR) The Elements of Statistical Learning by Hastie, T., Tibshirani, R., and Friedman, J., 2009, Second Ed. Download from https://web.stanford.edu/~hastie/ElemStatLearn/
- [**PT**] (text 3 on Stat and Probability Theory) [Miller] John E. Freund's Mathematical Statistics with Applications (8th edition) by Miller and Miller; or [Walpole] Probability and Statistics for Scientists and Engineers`, 9th Ed., by Walpole, Myers, et.al.; [**SI**] Statistical Inference by Casella & Berger, Springer, any version;

- [**HSFML**] (summary of stat and math theory behind ML) Handbook of Statistical Foundations of Machine Learning. G. Bontempi, online free version
- **Yusuf`s module outlines**, **board work**, handouts and posted notes.

Also:

- [PSwR] (text for prerequisite topics) Probability and Statistics with R, Second Edition, Maria Dolores Ugarte, Ana F. Militino, Alan T. Arnholt. CRC Press. Equivalently: [**IPSR**] Introduction to Probability and Statistics Using R. G. Jay Kerns.
- [**MLR**] (text for simpler version of Stat and ML modeling with R) Machine Learning with R. Second Ed, Brett Lantz.
- [**FCML**] (useful math and code for ML) A First Course in Machine Learning, 2016, Simon Rogers and Mark Girolami. http://www.dcs.gla.ac.uk/~srogers/firstcourseml/
- [MLAP] (intro to ML practices with R) Machine Learning: An Algorithmic Perspective, 2nd Ed, by Marsland 2014
- Linear Algebra Review (e.g. http://www.maths.gla.ac.uk/~ajb/dvi-ps/2w-notes.pdf)
- Journal articles and specialty topics

Others and Materials:

- Purchase a 1.5 inch binder to keep the notes, module outlines and tasks
- Have a nice printing balance
- Expect in-class labs, bring laptops

Assessment of Learning Outcomes:

The assessment of students' mastery of the above mentioned outcomes will occur on the basis of:

- Attendance with active participation, topic presentation and contribution (15%)
- Assignments (HW, lab) and random tasks (20%)
- One midterm project (20%)
- Binder and digital folders organization (5%)
- Group project with data analysis report and conference presentation (20%), and
- Final in-class test on concepts and theory (20%) – **May 3rd, Thursday, 12 to 3:00pm**

A numerical average will be computed using the weights indicated above, and converted to a letter grade according to the scale shown below:

A 93-100%	A- 90-92%	B+ 87-89%	B 83-86%	B- 80-82%
C+ 77-79%	C 73-76%	C- 70-72%	D 60-69%	E 0-59%

Homework, Tasks, Labs, Quizzes, Midterms:

- For each module, a homework will be assigned. I call it Module HW. Expect 9 HWs. See the google doc for deadlines.
- Some tasks randomly can be assigned. Expect 4 random tasks.
- Midterms are take-home tests, individually on applied and conceptual topics.
- All submissions require using the professional report tool, R+Sweave.
- A hard copy is submitted in class; once I give the feedback, keep the revised copy in the self-binder; and the latest digital copy is kept under your digital self-folder.
- Quizzes are pop assessments that may not be announced.
- Expected study hour per week is 6hr. In April, this would increase up to 10hr per week

Final Data Project with Conference Presentation:

This is a required group project (two to three) with a report and presentation at UPSTAT Conference on April 20-21, 2018 at UR. The project proposal is submitted by Week 9. A full report is prepared and submitted on the last day of the semester. Last one third portion of the semester sessions is dedicated for this purpose as research meeting. You will need to register for the conference.

Binder and digital folders organization:

- Self-binder: I give handouts and do board work on the theory whenever topics come. I expect you to keep notes and the handouts in the binder. Bring and hand in the self-binder in the final test.
- Digital self-folder: You keep all the submitted assignments under a google folder module by module. Also keep the report and resources under for projects. First, create a google folder, give a name to it (my example is YusufBilgic_Math380_Spring_2018) and share it with me with the edit access option.
- In the final test, I will grade these based on the criterion: organization of the materials.

My Web Page and Resources:

- Our course website that I share materials and resources can be found in the link http://go.geneseo.edu/bilgic (click on Math380-Stat&ML) or click. You will find helpful resources.
- For reaching the shared materials folder, click. This is a google folder that I put all under it.
- Also frequently visit the announcement page, click.
- Bookmark the three links above.
- Need your input periodically on the Track Topic google sheet, click. You find semester plan and deadlines for tasks in this doc.

R with RStudio and Python with editor:

- Install Jupyter Distribution that combine all what we need including use of the Geneseo server; Install R (statistical software) for free from http://cran.us.r-project.org/ along with RStudio (interface for R) from http://www.rstudio.com; Install Python 3; Install useful packages.

Course Outline, Lesson Plan and Tasks:

- Check the google link I shared for the lesson plan and deadlines of the tasks and project. I will remind these in the announcement site and class.

Incompletes, Disability, Academic Integrity and Dishonesty:

- University & Departmental policy will be followed.

Your learning, communication, eye contact, office visit, feedback, suggestion, contribution, positivity and creativity are all my motivation as an instructor.

WELCOME TO MY CLASS!

Subject to revision - For an updated version of the syllabus, visit the course website.

YB, 1/16/2018