STAT 6240.10 - Statistical Data Mining (Spring, 2019)

Course and Contact Information

- **Course:** Statistics, Statistical Data Mining, STAT 6240.10
- **Semester:** Spring, 2019
- **Meeting Time:** Thursday 6:10 - 8:40 PM
- **Location:** Duques Hall (2201 G St NW), Room 152

Instructor

- Name: Emre Barut
- Campus Address: Rome Hall #763 (801 22nd St NW)
- Phone: (202) 994-9025
- Email: barut@gwu.edu

- Instructor Office Hours: Thursday 3:00-5:00 PM
- Teaching Assistant (TA): <TBD>
- TA Office Hours: <TBD>

Course Description

Introduction to basic data mining concepts and techniques for discovering interesting patterns hidden in large-scale data sets, focusing on issues relating to effectiveness and efficiency.

Course Prerequisites

- Advanced Knowledge of R Programming
- STAT 6201 and 6202 (Mathematical Statistics I and II)
- STAT 6214 (Applied Linear Models) or equivalent

Learning Outcomes

As a result of completing this course, students will be able to:

1. Wrangle data,
2. Perform data analysis using machine learning,
3. Interpret complex statistical/machine learning models.

Average minimum amount of out-of-class or independent learning expected per week: 15 hours.

Required Textbooks, Materials and Recommended Readings

Suggested References

- **An Introduction to Statistical Learning** by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. Available online for free at www-bcf.usc.edu/∼gareth/ISL/.
- **The Elements of Statistical Learning** by Trevor Hastie, Robert Tibshirani and Jerome Friedman. Available online for free at https://web.stanford.edu/∼hastie/ElemStatLearn/.
- **Data Mining and Business Analytics with R** by Johannes Ledolter.
- **Advanced Data Analysis from an Elementary Point of View** by Cosma Rohilla Shalizi. Available at https://www.stat.cmu.edu/∼cshalizi/ADAfaEPoV/.
- **Deep Learning with R** by François Chollet and J. J. Allaire.
- **Applied Predictive Modeling** by Max Kuhn and Kjell Johnson.

Course Websites

We will be using two different websites. Please make sure that you register on Datacamp.com.

- Blackboard: Please check Blackboard frequently, as there may be assignments, announcements, and material passed to the class during the week. You can find it at blackboard.gwu.edu. You'll need to login using your GW user ID and password.
- Datacamp: This is an interactive website that offers classes on various fields of data science. We will be using a few of their classes. They normally require a premium subscription but you will have free access to the website during the Spring 2019 semester.

Week-by-week schedule of topics to be presented:

Week	Date	HW Due	Content
1	Jan-17		Overview
2	Jan-24	HW 0	PCA, Factor Models, Dimensionality Reduction
3	Jan-31	HW 1	Mixture Models, Clustering
4	Feb-07	HW 2	Classification Overview
5	Feb-14	HW 3	Model Building, Diagnosis, Decision Trees
6	Feb-21	HW 4	Bagging, Boosting, Random Forests
7	Feb-28	HW 5	Gradient Boosting, Evaluating Complex Classifiers
8	Mar-07	HW 6	Support Vector Machines, Ensemble Models
	Mar-14		*<Spring Break>*
9	Mar-21	<Midterm Project>	Neural Networks I - Basics
10	Mar-28	HW 7	Neural Networks II - Convolutional Neural Networks (CNN)
11	Apr-04		Neural Networks III - Variations on CNNs, NN Interpretation
12	Apr-11	HW 8	Text Analysis
13	Apr-18	HW 9	Complex Regression Methods
14	Apr-25	<Final Project>	Final Project Presentations

Grading

- Homework (45%): 8-10 assignments, due on Thursday before class, unless noted otherwise. Your lowest HW grade will be dropped. Late work will not be accepted; no exceptions will be made to this rule.

- Homework Presentation (5%): Each week 3 to 4 students will be quasi-randomly selected and they will present some portion of their homework solutions to the class. You will be graded on the quality of your communication, responses and overall approach. The random selection will be done when the class starts - you will not know in advance if you are supposed to present this week.
- Midterm Project (15%): Individual project that involves explorative data analysis and classification. More details will be provided later.
- Final Project (35%): You will form teams of 3 or 4, and enter a Kaggle.com competition. The competition will be decided later. You will be graded on your approach and your performance on Kaggle. You will present your results on the last day of class (April 25).

University policies

University policy on observance of religious holidays

In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance. For details and policy, see: students.gwu.edu/accommodations-religious-holidays.

Academic integrity code

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For details and complete code, see: studentconduct.gwu.edu/code-academic-integrity

Safety and security

In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.

Support for students outside the classroom

- **Disability Support Services (DSS)**: Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Rome Hall, Suite 102, to establish eligibility and to coordinate reasonable accommodations. For additional information see: disabilitysupport.gwu.edu/
- **Mental Health Services**: The University's Mental Health Services offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include: crisis and emergency mental health consultations confidential assessment, counseling services (individual and small group), and referrals. For additional information see counselingcenter.gwu.edu/ or call (202)-994-5300.

Security

- In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.