

# MATH 6365 - AUTOMATIC LEARNING AND DATA MINING

Spring 2019

---

<b>Instructor:</b> Andrey Skripnikov	<b>Time:</b> MW 4 - 5:30 PM
<b>Email:</b> <a href="mailto:andreysk@math.uh.edu">andreysk@math.uh.edu</a>	<b>Classroom:</b> CBB 214

---

**Contact:** Mostly over email, during office hours (by appointment) or using *UH Blackboard*.

**Main References:** While lecture notes will serve as the main source of material for the course, the following books constitute plausible reference (second book not as much though):

- "An Introduction to Statistical Learning (with applications in  $R$ )" by James, Witten et al.
- "Neural Networks with  $R$ " by G. Ciaburro

**Description:** Course will predominantly deal with applications for such automatic learning and data mining techniques as decision trees, random forests and neural networks. Moreover, the topics of model validation and statistical resampling methods will be addressed.  $R$  Statistical programming will be used throughout the course.

**Learning Objectives:** By the end of the course a successful student should:

- Have a solid conceptual grasp on the described automatic learning methods.
- Be able to correctly identify the appropriate techniques to deal with particular data sets. Realize the advantages and disadvantages of using a particular learning technique.
- Have a working knowledge of  $R$  programming software in order to apply those techniques and subsequently assess the quality of fitted models.
- Demonstrate the ability to clearly communicate the results of applying selected statistical learning methods to the data.

**Software:** Make sure to download  $R$  and  $RStudio$  (which can't be installed without  $R$ ) before the course starts. Use the link <https://www.rstudio.com/products/rstudio/download/> to download it from the mirror appropriate for your platform. Let me know via email in case you encounter difficulties.

**Tentative Course Outline:**

- **Introduction: What is Statistical Learning?** Supervised and unsupervised learning. Regression and classification.
- **Tree-based Models.** Decision and regression trees: splitting algorithm, tree pruning. Random forests: bootstrap, bagging, random splitting.
- **Model Validation.** Validation set approach. Cross-validation.
- **Neural Networks: Intro.** Single-layer perceptron: neuron model, learning weights. Simple & Multi-Layer Perceptron. Deep Learning.
- **Neural Networks: Advanced Topics.** Recurrent Neural Networks for sequential data, e.g. time series. Convolutional Neural Networks for image data.

**Grading Policy:**

- Attendance and In-Class Quizzes - 10%.
- Midterms 1 & 2 - 20% each.
- Homework - 25%.
- Final Group Project - 25%.

**Midterm Exams:** Held during class times.

**Homework:** Assigned on bi-weekly basis.

**Final Group Project:** A group assignment, consisting of students teaming up, deciding on the data set of interest, posing research questions and applying ML techniques to address those questions. Each group will eventually submit a report of research findings and member contributions.

**Academic Honesty:** Please click the link below for the full explanation of the Academic Honesty policy and procedure:

[http://www.uh.edu/provost/policies/honesty/\\_documents-honesty/academic-honesty-policy.pdf](http://www.uh.edu/provost/policies/honesty/_documents-honesty/academic-honesty-policy.pdf)

Definitions:

”Academic dishonesty” means employing a method or technique or engaging in conduct in an academic endeavor that contravenes the standards of ethical integrity expected at the University of Houston or by a course instructor to fulfill any and all academic requirements. Academic dishonesty includes but is not limited to, the following:

*Plagiarism*

- a. Representing as ones own work the work of another without acknowledging the source (plagiarism). Plagiarism includes copying verbatim text from the literature, whether printed or electronic, in all assignments including field.

*Cheating and Unauthorized Group Work*

- b. Openly cheating in an examination, as copying from anothers paper;
- c. Being able to view during an examination, quiz or any in-class assignment an electronic device that allows communication with another person, access to unauthorized material, access to the internet, or the ability to capture an image, unless expressly permitted by the instructor;
- d. Using and/or possessing crib notes, as unauthorized use of notes or the like to aid in answering questions during an examination;
- e. Giving or receiving unauthorized aid during an examination, such as trading examinations, whispering answers, and passing notes, and using electronic devices to transmit or receive information;
- f. Securing another to take a test in the students place. Both the student taking the test for another and the student registered in the course are at fault;

*Fabrication, Falsification, and Misrepresentation*

- g. Changing answers or grades on a test that has been returned to a student in an attempt to claim instructor error;
- h. Using anothers laboratory results as ones own, whether with or without the permission of the owner;

- i. Falsifying results in laboratory experiments;
- j. Misrepresenting academic records or achievements as they pertain to course prerequisites or corequisites for the purpose of enrolling or remaining in a course for which one is not eligible;
- k. Representing oneself as a person who has earned a degree without having earned that particular degree

#### *Stealing and Abuse of Academic Materials*

- l. Stealing, as theft of tests or grade books, from faculty offices or elsewhere, or knowingly using stolen tests or materials in satisfaction of exams, papers, or other assignments; this includes the removal of items posted for use by the students;
- m. Mutilating or stealing library materimaterials; misshelving materials with the intent to reduce accessibility to other students;

#### *Complicity in Academic Dishonesty*

- n. Failing to report to the instructor or departmental hearing officer an incident which the student believes to be a violation of the academic honesty policy;

#### *Academic Misconduct*

- o. Any other conduct which a reasonable person in the same or similar circumstances would recognize as dishonest or improper in an academic setting.

#### Process:

Students shall have the responsibility of reporting incidents of alleged academic dishonesty to the instructor of record involved or to the appropriate authority if the alleged act is not associated with a specific class within 5 class days of the incident. Faculty or instructor of record shall have the responsibility of reporting incidents of alleged academic dishonesty through their college hearing officer within 5 class days of the incident. The faculty should include the recommended sanction in the report. The college hearing officer will notify the student of the report and recommended sanction. The student can accept the sanction and waive a hearing or request a college hearing. A hearing shall be set within 10 days and would be consist of two faculty and three students chosen by the hearing officer.

**UH CAPS Statement:** Counseling and Psychological Services (CAPS) can help students who are having difficulties managing stress, adjusting to the demands of a professional program, or feeling sad and hopeless. You can reach CAPS ([www.uh.edu/caps](http://www.uh.edu/caps)) by calling 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis. No appointment is necessary for the Let's Talk program, a drop-in consultation service at convenient locations and hours around campus. [http://www.uh.edu/caps/outreach/lets\\_talk.html](http://www.uh.edu/caps/outreach/lets_talk.html)