Department of Information Sciences and Technology

AIT 590: MACHINE LEARNING

Course Syllabus

Fall 2019

Instructor: Dr. Mahdi Hashemi
Email: mhashemi@gmu.edu
Website: http://mason.gmu.edu/~mhashemi/
Office hours: Monday 2-4 PM

Description

One of the most common tasks performed by data scientists and data analysts is prediction and machine learning. Machine learning combines advanced topics in statistics, probabilities, linear algebra, and calculus to design mathematical models that learn from data or experience to solve new problems. This course concentrates on the theories, mathematics, logics, and algorithms behind machine learning, in addition to their applications. Home works, exams, in-class presentations, and final projects evaluate students’ understanding of theories, strengths, weaknesses, and characteristics of machine learning techniques, in addition to their comparison, evaluation, and implementation.

Requirements

An open-book closed-phone Qualifier’s Exam will be held during the second session, so come prepared. It covers topics in statistics, probabilities, linear algebra, calculus, and algorithms. It forms 5% of your final grade. More details about this exam are provided in the Grading Section. The final project and some of the homeworks require programming. You will experience some in-class Python programming with the instructor’s help. However, I encourage you to start learning programming if you have no experience in any language, e.g. Matlab, R, Python, or similar toolkits. The choice of the programming language for homeworks and projects is left to students and supported by the instructor. Yet, Python is recommended.

Objectives

On successful completion of this course, students will be able to:

- understand basic topics in machine learning, including classification, regression, clustering, feature selection and generation,
- implement and evaluate major machine learning models,
- overcome the challenges of working with real-world datasets.
Textbook


References


Grading

<table>
<thead>
<tr>
<th>Score</th>
<th>A+</th>
<th>A</th>
<th>A-</th>
<th>B+</th>
<th>B</th>
<th>B-</th>
<th>C</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>97-100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93-96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90-92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87-89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83-86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77-82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Qualifier exam given at the second session (open-book, questions on linear algebra, probabilities, statistics, and algorithms): 5%
- Five homeworks (due on midterm exam day and final exam day, upload in blackboard): 5%
- Midterm Exam (closed book, no electronic devices except a basic calculator): 25%
- Final Exam (not cumulative, closed book, no electronic devices except a basic calculator): 25%
- Python class presentation: 10%
  - Individual or a group of 2.
  - You need to communicate with the instructor about your topic.
  - You pick one of the topics from the syllabus, find the relevant commands in Python, and implement it using a dataset of yours.
- Theoretical Final project: 15%
  - Individual.
  - The report pdf and power point files must be digitally submitted before the deadline.
  - Presentation to the teacher during office hours (10~15 min) required. You might be asked to present in class as well.
  - The project involves picking a topic of interest from the syllabus, reading the relevant material and summarizing the key theoretical results. The emphasis should be on theory. You are not required to do new research. Proof of concept with implementations and experiments with real data are appreciated but not required.
  - Plagiarism in writing the report will be penalized with a failing grade in the course and referral to the Dean of Students.
- Experimental Final project: 15%
  - Individual.
  - Language of your choice.
  - The report pdf and the code (zipped project folder) must be submitted before the deadline
  - The project involves training different machine learning models using this dataset and evaluating it. Both training and test data contain missing values shown with a question mark (?). More details will be discussed by the instructor in class.
  - Plagiarism in coding will be penalized with a failing grade in the course and referral to the Dean of Students.
Topics

- Classifiers
  - Bayesian classifiers
    - Normal pdf
    - Nonparametric estimation
    - Naïve Bayes
    - Bayesian belief networks (or Bayesian networks for short)
  - Content-dependent classification
    - Markov chain model
  - Linear classifiers
    - Perceptron
    - SVM
    - Least squares
    - Logistic regression
  - Non-linear classifiers
    - Kernel SVM
    - Decision trees
- Evaluating classifiers and regressors
  - Generalization accuracy
  - Setting the value of hyper-parameters
  - Evaluation metrics for regression models (RMSE, coefficient of determination R²)
  - Evaluation metrics for classification models
- Processing the data before training
- Feature selection and generation
  - Linear regression
    - Feature selection
    - Shrinking the coefficients
    - Feature generation
  - Classification
    - Feature selection
    - Feature generation
- Clustering
  - Proximity measures
    - Proximity measures between two points
    - Proximity functions between a point and a set/cluster
    - Proximity functions between two sets/clusters
  - Sequential algorithms
    - Basic Sequential Algorithmic Scheme
    - What is the appropriate number of clusters?
    - A modification of Basic Sequential Algorithmic Scheme
    - MaxMin Sequential Algorithmic Scheme
- Refinement of the clusters
  - Hierarchical algorithms
    - Agglomerative algorithms
    - Divisive algorithms
    - Choice of the best number of clusters
  - Schemes based on function optimization
    - Mixture decomposition or probabilistic clustering algorithms
    - Fuzzy clustering algorithms
    - Possibilistic clustering algorithms
    - Hard clustering algorithms
  - Evaluating clusters
    - Why evaluating clusters?
    - Monte Carlo technique
    - Internal criteria
Important Dates

Dates for dropping, adding the course etc. are available via http://registrar.gmu.edu/calendars/

Religious Holidays

A list of religious holidays is available on the University Life Calendar page. Any student whose religious observance conflicts with a scheduled course activity must contact the Instructor at least 2 weeks in advance of the conflict date in order to make alternative arrangements.

Attendance Policy

Students in in-class sections are expected to attend each class, to complete any required preparatory work and to participate actively in lectures, discussions and exercises. As members of the academic community, all students are expected to contribute regardless of their proficiency with the subject matter.

Students are expected to make prior arrangements with Instructor if they know in advance that they will miss any class and to consult with the Instructor as soon as possible if they miss any class without prior notice. Any student who expects to miss more than one class session is strongly advised to drop the course and take it in a later semester when he/she can attend every class.

Departmental policy requires students to take exams at the scheduled time and place, unless there are truly compelling circumstances supported by appropriate documentation. Except in such circumstances, failure to attend a scheduled exam will result in a score of zero (0) for that exam, in accordance with Mason policy on final exams. Students should not make travel plans or other discretionary arrangements that conflict with scheduled classes and/or exams. If the University is closed due to weather or other unforeseen conditions, final exams may be rescheduled – students are strongly advised not to make plans that would prevent them from attending exams that may be rescheduled during the entire exam period.

Classroom conduct

Students are expected to conduct themselves in a manner that is conducive to learning, as directed by the Instructor. Any student who negatively impacts the opportunity for other students to learn will be warned – if disruptive behavior continues, the student will be asked to leave the classroom.

Electronic devices are potential distractions in the classroom environment. Cell phones, pagers and other handheld devices must be turned off or set to "silent" mode and not used while class is in session. Laptop computers and similar devices may be used only if such use is directly related to the classroom activity in progress – for some activities the Instructor may require that such devices not be used in order to maximize student engagement.
Communications

Registered students will be given access to a section of the Blackboard Learning System for this course. Blackboard will be used as the primary mechanism (outside of lectures) to disseminate course information, including announcements, lecture slides, homework and other assignments, and scores for homework and exams.

Communication with the Instructor on issues relating to the individual student should be conducted using Mason email, via telephone, or in person - not in the public forums on Blackboard. Mason Mail is the preferred method – for urgent messages, you should also attempt to contact the Instructor via telephone. Federal privacy law and Mason policy require that any communication with a student related in any way to a student's status be conducted using secure Mason systems – if you use email to communicate with the Instructor you MUST send messages from your Mason email account.

When sending an e-mail to the instructor, please include the following:

- Course number
- Your full name

Privacy

Instructors respect and protect the privacy of information related to individual students.

As described above, issues relating to an individual student will discussed via email, telephone or in person. Instructors will not discuss issues relating to an individual student with other students (or anyone without a need to know) without prior permission of the student.

Assessable work other than final exams will be returned to individual students directly by the Instructor (or by a faculty or staff member or a Teaching Assistant designated by the Instructor, or via another secure method). Under no circumstances will a student's graded work be returned to another student.

Faculty and staff will take care to protect the privacy of each student's scores and grades.

Disability Accommodations

The Office of Disability Services (ODS) works with disabled students to arrange for appropriate accommodations to ensure equal access to university services. Any student with a disability of any kind is strongly encouraged to register with ODS as soon as possible and take advantage of the services offered.

Accommodations for disabled students must be made in advance – ODS cannot assist students retroactively, and at least one week's notice is required for special accommodations related to exams. Any student who needs accommodation should contact the Instructor during the first week of the semester so the sufficient time is allowed to make arrangements.
Honor Code

All members of the Mason community are expected to uphold the principles of scholarly ethics. Similarly, graduating students are bound by the ethical requirements of the professional communities they join. The ethics requirements for some of the communities relevant to IST graduates are available via the following links:

- ACM Code of Ethics and Professional Conduct
- IEEE Code of Ethics
- EC-Council Code of Ethics

On admission to George Mason University, students agree to comply with the requirements of the Mason Honor System and Code. The Honor Code will be strictly enforced in this course. Honor Code cases are heard by a panel consisting of students – students who meet the requirements are encouraged to nominate themselves to serve on the Honor Committee.

Any use of the words or ideas of another person(s), without explicit attribution that clearly identifies the material used and its source in an appropriate manner, is plagiarism and will not be tolerated. The Instructor reserves the right to use manual and/or automated means (including such services as SafeAssign.com) to detect plagiarism in any work submitted by students for this course, and to direct Teaching Assistants and/or other faculty and/or staff members to do likewise in support of this course. For this course, the following requirements are specified:

- All assessable work is to be prepared by the individual student, unless the Instructor explicitly directs otherwise.
- All work must be newly created by the individual student for this course for this semester.

Any usage of work developed for another course, or for this course in a prior semester, is strictly prohibited without prior approval from the instructor.

Students may seek assistance with assigned work (and are encouraged to do so if they feel the need), provided:

- Such assistance is acknowledged in the submitted work, clearly identifying the person(s) giving assistance and the nature of the assistance given.
- Any work to be submitted is prepared entirely an exclusively by the student submitting it. Students are expressly prohibited from sharing any assessable work for this course in any manner with other students (except students assigned as Teaching Assistants to this course and the student's section), unless all students involved have had their work graded and returned by the Instructor, or the Instructor has explicitly approved such sharing.

Counseling and Psychological Services

Phone: 703-993-2380
https://caps.gmu.edu/
Title IX Coordinator

Angela Nastase
Compliance, Diversity and Ethics
373 Aquia Building, MS 2C2, Fairfax Campus
Phone: 703-993-8730
titleix@gmu.edu
https://diversity.gmu.edu/sexual-misconduct