INF 552: Machine Learning for Data Informatics

Units :   4
Location :   ZHS 252

Instructor :   Ion Muslea
Office Hours :   by appointment
Contact Info :   muslea@usc.edu

Teaching Assistant: -

Course Description and Learning Objectives

The learning objectives for students in this course are:
- acquire a ***hands-on understanding of applying machine learning in the real world***
- understand major (types of) concepts, algorithms, approaches used in Machine Learning
- understand main supervised and unsupervised learning techniques
- understand Bayesian decision theory and nonparametric methods
- understand decision trees, support vector machines, regression, and clustering

This is a foundational course with primary application in data analytics. It is intended to be accessible to students with technical backgrounds as well as to students with less technical backgrounds. The textbooks for INF-552 are:
- "Introduction to Machine Learning", 3rd edition, MIT Press, 2015, by Ethem Alpaydin
- "Machine Learning Yearning", by Andrew Ng (free copy @ http://www.mlyearning.org)

Recommended Preparation
A basic mathematical background in probability, statistics, and linear algebra, as well as basic programming skills and a basic understanding of engineering principles are strongly encouraged.

Description and Assessment of Assignments
Grading will be based on students' understanding of lecture material, the thoroughness of their projects, the clarity of their project reports and presentations, and their ability to explore related areas. Students can work on the project in groups of 2-3, but should mention their individual contributions.

Grading Breakdown
- Project: 70%
  - See project details below
- Quizzes: 10%
  - scope: all the reading materials from ***last 2 weeks***
  - the lowest score of all quizzes will be ignored
- Final exam: 20%
  - ***<u>December 11</u>***, 4:30-6:30 pm

Course project: the purpose of the class project is for you to acquire the hands-on experience of identifying a data problem and applying machine learning approaches to solve it. Students are encouraged to identify unique applications for ML and develop novel approaches. Working as a group is permitted if the project is large enough to justify it; teams can consist of up to 3 persons.

Project Timeline (changes might apply):
- Weeks 1-3:   identifying team members and project topic
- Week 4:   **9/11**   draft proposal due (team member, topics, and milestones)
  - feedback will be provided to modify proposal as/if needed
- Week 6:   **9/25**   in-class presentation of proposal (about 7' per project)
- Week 9:   **10/16**   mid-term report & presentation due
  - emphasize each team member's contribution!
- Week 10:   **10/23**   in-class presentation of mid-term progress
- Week 14:   **11/20**   final report & presentation due
  - emphasize each team member's contribution!
- Week 15:   **11/27**   in-class presentation of final results

Sample projects: google "***ideas machine learning class projects***"
- http://www.hlt.utdallas.edu/~vgogate/ml/2012s/projects.html
  - http://cs229.stanford.edu/projects2011.html
  - http://www.cs.cmu.edu/~guestrin/Class/10701/projects.html
- https://elitedatascience.com/machine-learning-projects-for-beginners
- http://web.cs.hacettepe.edu.tr/~ilyas/Courses/BIL712/possibleprojecttopics.pdf

Bootstrap your project work by using the tools you are most proficient with, such as:
- Java   & Weka
- Python & ML libraries
- Matlab & ML libraries
- R   & ML libraries
- C++   & MLC++

Grading breakdown of the course project:
- Proposal: 5%
- Mid-term report & presentation: 15% = 10% + 5%
  - **10%:** PDF document with the written report (prose @ level of a workshop paper)
  - **5%:**
    - PPTX/Keynote document with the actual presentation
    - a 10' presentation in front of the class
- Final report & presentation: 80% = 70% + 10%
  - **70%**: PDF doc with the written report (@ level of an accepted workshop paper)
  - **10%**:
    - a PPTX/Keynote document with the actual presentation
    - a 10-15' presentation in front of the class

NOTE: for both mid-term & final report & presentation, a draft of the PDF & PPTX documents is due ***one week before the actual presentation***, so that it can be reviewed in advance. Both the mi-term and final report should use as a template the paper [Caruana & Mizil, 2006]: https://www.cs.cornell.edu/~caruana/ctp/ct.papers/caruana.icml06.pdf

The reports and presentations will be graded based on their thoroughness and the clarity of the prose, ideas, explanations, and discussion. It is paramount that you offer crystal-clear descriptions of:
-    the problem definition (what are you doing?)
-    the problem motivation (why are you doing it)
-    the proposed approach (how are you doing it @ macro-level?)
-    data collection & preparation (based on what are you doing it?)
-    the experimental details (how are you doing it @ micro-level?)
-    the in-depth discussions of the results (what did you learn?)
The reports and presentations will be also graded based on the originality, effort, and challenges of both the data and algorithmic issues.

Assignment Submission Policy
All submission are due before the beginning of the class (i.e., before 5:30 pm on the corresponding Tuesday). Students are required to submit the assignments on USC's BlackBoard portal or via email (TBD). One submission per team is sufficient. All submitted files should be named as follows:
-    last+first names of authors (in consistent, alphabetic order)
-    the document type
-    a brief version of the title
For example, the dues dates and the names of the **7 files** to be submitted for the Caruana & Mizil paper would be:
-    09/11:   CaruanaRich_MizilAlex_DraftProposal-Report_LargeScaleMLEval.pdf
-    09/25:   CaruanaRich_MizilAlex_FinalProposal-Report_LargeScaleMLEval.pptx
-    09/25:   CaruanaRich_MizilAlex_FinalProposal-Presentation_LargeScaleMLEval.pptx
-    10/16:   CaruanaRich_MizilAlex_MidTerm-Report_LargeScaleMLEval.pdf
-    10/16:   CaruanaRich_MizilAlex_MidTerm-Presentation_LargeScaleMLEval.pptx
-    11/20:   CaruanaRich_MizilAlex_Final-Report_LargeScaleMLEval.pdf
-    11/20:   CaruanaRich_MizilAlex_Final-Presentation_LargeScaleMLEval.pptx

Academic Conduct
Plagiarism—presenting someone else's ideas as your own, either verbatim or recast in your own words—is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" https://policy.usc.edu/student/scampus/part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct http://policy.usc.edu/scientific-misconduct.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute http://ali.usc.edu, which sponsors courses and workshops specifically for international graduate students. The Office of Disability Services and Programs http://dsp.usc.edu provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, USC Emergency Information http://emergency.usc.edu will provide safety and other updates, including ways in which instruction will be continued by means of BlackBoard, teleconferencing, and other technology.