CS3332 (4-0) - Applied Machine Learning

Catalog description:

Survey of machine-learning techniques of artificial intelligence with a particular focus on military applications. Topics include types of machine learning, training and testing of machine learning, data preparation, decision trees, Bayesian reasoning, linear models, neural networks, case-based reasoning, and reinforcement learning. Each method will be related to important military and government applications. This course is intended for students who are not computer-science majors.

Prerequisites: CS3331, CS3310, or equivalent

Syllabus:

Grading will be three homework assignments and three tests. Weekly topics are:

- Data types, data formatting, and data transformation
- Categories of machine learning, general principles of training and testing (applications to sensor data)
- Decision trees (applications to lethal autonomous systems)
- Bayesian reasoning (applications to sensor fusion)
- Linear models (applications to anomaly detection)
- Support-vector machines and principal-components analysis (applications to digital forensics)
- Basics of neural networks (applications to computer vision, applications to predictive maintenance)
- Case-based reasoning (applications to real-time systems)
- Clustering (applications to network-traffic analysis)
- Reinforcement learning (application to robots)

Note: The course will not have designated laboratory hours since those are difficult to do with distance learning. However, the homework assignments will include exercises with software and data that students will download.

Grading

Three homework assignments, a test, and a term project. They are weighted approximately equally. Usually averages on the homework and test are 70%, so the questions are difficult. However, grades are assigned relative to the rest of the class, so you are not penalized when a question is difficult for everyone. Overall, a median grade in this course is between an A- and B+. You will only get below a B if you do not appear to be doing the required work.

Homework should be submitted as a paper (hard) copy of a single document with your name at the top of the document, all material for each problem together (no appendices), and without the text of the

problems (just your answers). Generally homework questions are written so quoting or paraphrasing something written somewhere else will not answer the question. Homework submitted after the due date but before solutions are made available is subject to a 15% penalty; homework will not be accepted after solutions are made available.

Homework 1 can be done in two-person teams without communicating with any other members of the class. Homeworks 2 and 3 must be done individually without communicating with anyone except the instructor. You can look up or discuss general knowledge of the subject, but you cannot discuss the homework questions with anyone besides the instructor. Spreadsheet programs like Excel are not allowed for the assignments; you can use a calculator, but all the steps of your calculations need to be written out on the homework.

Since homework and test averages run around 70%, so expect them to be reasonably difficult and require some time to do. It is important to pace yourself on the homework and start working on the homework as soon as it is handed out. If you have trouble figuring out what to do for a homework question, discuss with the instructor well before it is due. Do not expect to be able to only work on an assignment the night before it is due.

The term project can be done individually, in a two-person team, or a three-person team. More details will be given.

Textbook and notes

The required textbook is I. Witten, E. Frank, M. Pal, and C. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, fourth edition, Morgan Kaufmann, 2016. The third edition is sufficient, but the fourth edition is better. Chapters will also be assigned from V. Schonberger and K. Cukier, *Big Data*, Mariner Books, 2013; and G. Smith, *Standard Deviations*, Overlook Press, 2014. This course is focused on algorithms, not training to use software.

We will use a Sakai site for course information, slide revisions, and additional readings. We will also email you copies of the slides before the first class; please report if you do not receive them.

Week	Topics	Homework and Tests	Readings
1	Introduction and overview (1 hour) Data types and formatting (2 hours) Data transformations (1 hour)		Schonberger reading
2	Data transformations (1 hour) Categories of machine learning and general training methods (2 hours) Testing of machine learning (1 hour)		Chapter 1 and sections 5.1, 5.2, 5.3, and 5.8 in the textbook
3	Decision trees (2 hours) Applications to lethal autonomous systems (1 hour) Introduction to Bayesian reasoning (1 hour)		Chapters 2 and 3
4	Methods of Bayesian reasoning (2 hours)	Homework 1 due, covering the first three weeks of topics.	4.4 and 4.5

	Applications to sensor fusion (1 hour) Applications to digital forensics (1 hour)	One problem on formatting and transforming data, one on measuring learning performance, one on building a decision tree, one on Bayesian reasoning	
5	Linear models (2 hours) Applications to anomaly detection (2 hours)		4.1, 4.3,
6	Support-vector machines and principal-components analysis (2 hours) Applications to digital forensics (1 hour) Applications to natural-language processing (1 hour)	Midterm test, covering the topics in weeks 1-5	6.1, 6.2
7	Basics of neural networks (2 hours) Applications to computer vision (2 hours)	Homework 2 due, covering topics in weeks 4-6; several problems requiring using a machine-learning software tool, one problem on linear models, one problem on neural networks	4.2, 4.6
8	Case-based reasoning (2 hours) Applications to real-time systems (1 hour) Clustering methods (1 hour)		4.7, 4.8
9	Applications of clustering to network-traffic analysis (1 hour) Evolutionary algorithms (2 hours) Applications to cyber-attack recognition (1 hour)		7.1
10	Reinforcement learning (2 hours) Applications to robots (2 hours)	Homework 3 due, covering topics in weeks 7-9; one problem on neural networks, one problem on case-based reasoning or clustering, and one problem on evolutionary algorithms, one problem using the machine-learning tool of the second homework assignment	Chapters 1, 2, and 7 in Smith reading
11	Review for final test	Final test, covering topics in weeks 6-10 Term project due at end of exam week	

Learning outcomes:

- Identify and distinguish the major methods of machine learning.
- Be able to prepare data for effective use of machine-learning methods.
- Be able to run machine-learning methods on a standard tool.
- Be able to interpret output of machine-learning methods and draw valid conclusions from it.
- Understand the major terms used to describe machine learning.
- Be able to use machine-learning tools.
- Be able to draw conclusions from results of machine learning.