MACHINE LEARNING

http://kogan.rutgers.edu/ml-phd

26:198:622:01 Index: 20535

A Ph.D. course offered in the spring of 2018

Accounting and Information Systems
Rutgers Business School - Newark and New Brunswick
Rutgers University

Prof. Alexander Kogan
One Washington Park #924 (Newark), (973) 353-1064
kogan@rutgers.edu

Location:	Day:	Time:
Newark Campus; 1WP-204	Thursday	1:00 p.m. - 3:50 p.m.

Overview: Many successful applications of machine learning exist already, including systems that analyze past sales data to predict customer behavior, identify fraudulent credit card transactions, and recognize faces or spoken speech. This course will cover supervised learning, Bayesian decision theory, parametric methods, multivariate methods, dimensionality reduction, clustering, nonparametric methods, decision trees, linear discrimination, multilayer perceptrons, hidden Markov models, assessing and comparing classification algorithms, and combining multiple learners.

Required Textbook:

- Ethem ALPAYDIN. ***Introduction to Machine Learning, Third Edition***. The MIT Press, August 2014, ISBN 978-0-262-028189.
  https://mitpress.mit.edu/books/introduction-machine-learning-0
- Author's Web site for this textbook (providing, in particular, errata and the PPT presentations to be used in class) can be found at:
  http://www.cmpe.boun.edu.tr/~ethem/i2ml3e/

The course will utilize the Blackboard online facilities, which can be found at:

https://blackboard.rutgers.edu/

For a student to gain access to our Blackboard system, they must be enrolled and they must have a NETID. Once an enrolled student obtains a NETID, they will be added to the roster within 2 business days. Students should also check their email account in the system and if it is not correct, they need to update their official student record. Students who do not have a NETID, can create one online using this link:

https://netid.rutgers.edu/

Coursework: The coursework includes attendance of lectures and participation in class discussions, writing a research paper and presenting it in class, completing a computational project in machine learning and submitting its results for evaluation, and taking the final exam that will be a review of a published machine

learning paper.

Research Paper: Every student is required to write a research paper devoted to an important topic in machine learning. While it is expected that most students will choose a topic devoted to applications of machine learning in business, other topics may work as well. Every student is required to prepare a three page long proposal for the research paper, and submit this proposal for instructor's evaluation by **March 1, 2018**. The Word document must be posted to the appropriate Blackboard discussion forum, and its printout should be submitted in class. The research paper should be presented during the last meeting of the class on **April 26, 2018**. Both the research paper and the presentation should also be posted to the appropriate Blackboard forum before the last meeting of the class. The research articles to be covered in the research paper can be found in the following theoretical and applied journals publishing relevant articles:

- Machine Learning: http://link.springer.com/journal/10994
- Journal of Machine Learning Research: http://jmlr.org/
- Data Mining and Knowledge Discovery: http://link.springer.com/journal/10618
- IEEE Transactions on Knowledge and Data Engineering: http://www.computer.org/portal/web/tkde
- Intelligent Systems in Accounting, Finance & Management: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1174
- Journal of Emerging Technologies in Accounting: http://aaapubs.org/loi/jeta
- Decision Support Systems: http://www.sciencedirect.com/science/journal/01679236

Most of these journals are available through the Rutgers University Library subscriptions, and can be accessed from campus computers or from home through the library Web site:

http://www.libraries.rutgers.edu/find_articles

The following online research tools can be useful in conducting bibliographic searches for your research paper:

- DBLP Computer Science Bibliography: http://dblp.uni-trier.de/db/
- CiteSeer - Scientific Literature Digital Library and Search Engine: http://citeseer.ist.psu.edu/
- Google Scholar: http://scholar.google.com/

Important guidance on writing machine learning papers can be found in the following manuscript:

- Pat Langley, Crafting Papers on Machine Learning: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.5877

Computational Project: Every student is required to carry out a computational project focused on experimental comparison of several machine learning methods on different datasets. The computational experiment should compare at least **FIVE** different machine learning methods on at least **FOUR** different datasets. Every student is required to prepare a computational project proposal and submit this proposal for instructor's evaluation by **March 29, 2018**. The Word document describing the machine learning methods and the datasets to be utilized in the experiment, as well as the proposed experimental methodology, must be posted to the appropriate Blackboard discussion forum, and its printout should be submitted in class. The description of results of the computational experiment should also be posted to the appropriate Blackboard forum and its printout should be submitted during the last meeting of the class on **April 26, 2018**. It is recommended that the computational experiments be conducted using the public domain Machine Learning software package called **Weka** (using the most recent book version - currently 3-6-13). Here are some useful Weka links:

- Weka Homepage: http://www.cs.waikato.ac.nz/~ml/weka/

- WekaWiki: http://weka.wikispaces.com/
- Weka Documentation: http://www.cs.waikato.ac.nz/ml/weka/index_documentation.html
- Weka Frequently Asked Questions: http://weka.wikispaces.com/Frequently+Asked+Questions

Another good public domain Machine learning software package that can be use in experiments in addition to (or instead of) Weka is RapidMiner:

http://rapidminer.com/

The datatsets for use in computational experiments can be obtained from the **UCI Machine Learning Repository**:

http://mlearn.ics.uci.edu/MLRepository.html

It is absolutely essential to start working on the research paper and the computational project as soon as possible.

Final Exam: You will be given 24 hours starting at **1 PM on May 3, 2018** to write a critical review of a published machine learning article. The article will be made available in the Assigments section of the Blackboard. The completed review has to be submitted through the Assignments section of the Blackboard by **1 PM on May 4, 2018**.

Grading: The evaluations of your research paper, the computational project, and the final exam article review will be the basis for the course grade:

40%	Research Paper
30%	Computational Project
30%	Final Exam Article Review

Communications and Course E-Mailing List:

The best way to contact me is via email. The course is supported by the RAMS e-mailing list ***ml-phd-list***. The list membership is automatically synchronized with the current class roster. Make sure that your current e-mail address is available in the Rutgers online directory. To post a message to the list, e-mail it to

ml-phd-list@rams.rutgers.edu

All the postings to this list are permanently archived and available from

https://rams.rutgers.edu/archive/archive.cgi

Please note that your postings should be appropriate for this course.

Every student is responsible for maintaining the current e-mail address in the Rutgers Online Directory. You can maintain your personal information at:

https://personalinfo.rutgers.edu/pi/

Preliminary Schedule:

1. **01/18/2018**
   Chapter 1 - Introduction to Machine learning
   Chapter 2 - Supervised Learning
2. **01/25/2018**
   Chapter 2 - Supervised Learning
   Chapter 3 - Bayesian Decision Theory
3. **02/01/2018**
   Chapter 4 - Parametric Methods
4. **02/08/2018**
   Chapter 5 - Multivariate Methods
5. **02/15/2018**
   Chapter 6 - Dimensionality Reduction
6. **02/22/2018**
   Chapter 7 - Clustering
   Chapter 8 - Nonparametric Methods
7. **03/01/2018**
   Chapter 8 - Nonparametric Methods
   Chapter 9 - Decision Trees
   - **Research Paper Proposal is due**
8. **03/08/2018**
   Chapter 19 - Design and Analysis of Machine Learning Experiments
9. **03/22/2018**
   Chapter 10 - Linear Discrimination
10. **03/29/2018**
    Chapter 11 - Multilayer Perceptrons
    - **Computational Project Proposal is due**
11. **04/05/2018**
    Chapter 13 - Kernel Machines
    Chapter 14 - Graphical Models
12. **04/12/2018**
    Chapter 14 - Graphical Models
    Chapter 16 - Bayesian Estimation
13. **04/19/2018**
    Chapter 16 - Bayesian Estimation
    Chapter 17 - Combining Multiple Learners
14. **04/26/2018**
    Research Paper Presentations
    Computational Projects are due
15. **05/03/2018**
    Final Exam starts at 1 PM
    Article review is due at 1 PM on May 4, 2018