

CS 419 Informatics. A.K.A. Machine Learning

Objective: This course presents students with basic, yet seminal techniques for machine learning and data mining. At the end of the course the student should be able to understand how these, and other more sophisticated techniques work and when it is appropriate to apply them. This is a valuable skill in a job market where information management and data mining are key. This is a project based class.

Instructor: Francisco Iacobelli

email: f-iacobelli@neiu.edu

At: Lech Walesa Hall 31060

When: W: 7:05p - 9:45p.

Textbook: Machine Learning by Tom Mitchell, 1997

Office Hours: Check the course calendar.

- Conditions to take this course:**
- You must be able to save documents as PDF and pictures as PNG or JPEG.
 - You must be comfortable using a terminal window and running python and java from the command line.
 - You should be prepared to spend 10-15 hours a week on this course on average. b

Note: This syllabus is a living syllabus and future topics are subject to change.

| Week | To be covered | Assignment |
|------|---|--|
| 8/30 | <ul style="list-style-type: none">• Administrivia/Introduction: Machine Learning and Data Mining• Objectives: Introduce students to the problem of data mining and understand why it is difficult. | Opinion Mining with Movie Reviews. Use the reviews in here and turn in your program and a paragraph or two describing your training and test sets, your reasoning and your evaluation. |
| 1 | General to Specific Learning: Ch. 2 | <ul style="list-style-type: none">• Slides• Assignment 2: Exercise 2.5 (a) and (b) in the book |
| 2 | Decision Trees: Ch. 3 | <ul style="list-style-type: none">• Slides• Assignment 3: Exercise 3.4 (a), (b) and (c) |
| 3 | Neural Networks: Ch. 4 | <ul style="list-style-type: none">• Slides• Download Weka and play with it (not graded, but needed for the next assignment) |

| Week | To be covered | Assignment |
|-------|---|--|
| 4 | Basic Statistics: Ch. 5 | <ul style="list-style-type: none"> • Slides • Homework: download this file (txt_sentoken.arff) and load it in Weka. This file is the movie review data adapted to work with Weka. • Using Weka, evaluate the performance of Neural Networks (with a couple of different number of hidden nodes), Decision Trees (with and without pruning), Naïve Bayes, and one other method, not covered in class (do basic research to know how it works, roughly). • Repeat the last step about 10 times, obtain measures of accuracy, F-measure, precision and recall and compare which one performs the best. • Write one or two paragraph explaining why is it that one method works better than another. What makes it better suited for this problem. |
| 5 | Bayesian Learning I Ch. 6; 6.1 – 6.10. | Slides |
| 6 | Bayesian Networks Ch. 6; 6.11 – 6.13 and attend the talk by professor Neapolitan in LWH 3044 at 7pm. | |
| 7 | Bayesian Networks Ch. 6; 6.11 – 6.13 | |
| 8-9 | Markov Models | |
| 10 | Clustering: K-Nearest Neighbor Ch. 8 | <ul style="list-style-type: none"> • Compute, using the Naïve Bayes approach, the probability of playing tennis given $\langle \text{Overcast}, \text{Cool}, \text{Low}, \text{Weak} \rangle$. Use Table 3.2 (p.59) in the book for training. • In the slides for Bayesian Learning above, Slide 43 contains a Bayesian Network. Based on that network, answer the question: if John and Mary call, has there been a burglary? how probable is that? In other words, compute $P(\text{Burglary} \text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true})$. • Choose Teams. I have many corpora available |
| 11 | A-Priori Associations Reading to be assigned | Read and Blog about assigned Papers |
| 12-14 | Reinforcement Learning | |
| 15-16 | Advanced Topics I TBD | |
| 17 | Final Presentations | |

Grading

| Item | Weight |
|--------------------------------|--------|
| Assignment and Quizzes Average | 60% |
| Midterm Project + Test | 20% |
| Final Project | 20% |

| Weighted Average | Course Grade |
|------------------|--------------|
| 90% or higher | A |
| 80% – 89% | B |
| 70% – 79% | C |
| 60% – 69% | D |
| 0% – 59% | F |

Every Class

Come prepared, google the topics, read the book and, of course, do the homework on your own.

Late Work Policy

No late submissions will be accepted.

Academic Integrity

Students are required to abide by Northeastern Illinois University’s academic integrity policy. Failure to adhere to this policy will likely result in a failing grade in the class and / or expulsion from the University.

Web Link to Emergency Information

It is recognized that a safe university environment is a shared responsibility of faculty, staff, and students, all of whom are expected to familiarize themselves with and cooperate with emergency procedures. Emergency Procedures and Safety Information can be found on NEIUport on the MyNEIU tab or at:

Main campus: http://www.neiu.edu/~neiutemp/Emergency_Procedures/MainCampus/

El Centro(English version): http://www.neiu.edu/~neiutemp/Emergency_Procedures/ElCentro/

El Centro(Spanish version): http://www.neiu.edu/~neiutemp/Emergency_Procedures/ElCentro_Spanish/

CCICS: http://www.neiu.edu/~neiutemp/Emergency_Procedures/CCICS/

Chicago Teachers Center(CTC): http://www.neiu.edu/~neiutemp/Emergency_Procedures/CTC/

University Center at Grayslake: <http://ucenter.ehclients.com/pdfs/UCLCEmergencyPreparedness.pdf>