CS 780/880 Introduction to Machine Learning

Spring 2017

Website: http://cs.unh.edu/~mpetrik/teaching/intro_ml_17/

Overview

This class will teach you how to use machine learning to understand data and make predictions. The class will focus on understanding the fundamental concepts and algorithms that underlie modern machine learning and data science algorithms.

Contact Information

Marek Petrik	
Email:	mailto:mpetrik@cs.unh.edu
Office:	Kingsbury W233
Office hours:	Thursday: 2pm - 3:30pm
Discussion forum:	https://piazza.com/unh/spring2017/cs780cs880

The fastest way to get your question about course material answered is to post it on piazza (see the link above).

Textbooks

ISL James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning [online pdf]
ELS Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics (2nd ed.) [online pdf]

Other references

DL Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. [online pdf on github]
LA Strang, G. Introduction to Linear Algebra. (2016)
CO Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. [online pdf]
RL Sutton, R. S., & Barto, A. (2012). Reinforcement learning. 2nd edition [online pdf draft]

Evaluation

The weights of individual components in computing the grade are as follows:

50%	6 Assignments
15%	Midterm exam
30%	Final exam
15%	Class project

Note that the weights sum to 110%. Each assignment will have slightly different questions for the graduate and undergraduate version of the class.

Disability If you are registered with the student disability office, please let me know right away so that I can provide proper accommodation.

Academic honesty

- University Academic Honesty Policy:
  https://www.unh.edu/student-life/handbook/academic/academic-honesty
- Tutorial on Plagiarism:
  http://cola.unh.edu/plagiarism-tutorial-0

Assignments

General policies

- Assignments are due on the day indicated **by 12:40pm**
- Assignments should be turned in as a PDF on myCourses, or printed and turned in at class; if other methods fail, email to mailto:mpetrik@cs.unh.edu with Subject that contains the string `[CS780880HW]`
- There is no credit for assignments turned in late; This is so we can discuss solutions in timely manner.
- Collaboration when solving assignments is *encouraged* but all of the writing must be yours. You may collaborate via Piazza or in person.
- Please do not copy solutions from online sources; it defeats the purpose of the assignments and you will not do well in the exams.

Programming Assignments The class will involve hands-on data analysis using machine learning methods. The recommended language for programming assignments is R which is an excellent tool for statistical analysis and machine learning. No prior knowledge of R is needed or expected; the book and lectures will include a gentle introduction to the language.

Exams

The final exam will be take-home over a period of several days. The exam will require about 3 hours to complete. No collaboration on the exams.

Tentative Syllabus

Please check for updates on the website

Date	Day	Topic	Reading	Assignment
Jan 24	Tue	*Snow day*		
Jan 26	Thu	Statistical learning and R language	ISL 1,2	
Jan 31	Tue	Linear regression basics	ISL 3.1-2	
Feb 02	Thu	*No class*		
Feb 07	Tue	Linear regression advanced	ISL 3.3-6	
Feb 09	Thu	Classification and logistic regression	ISL 4.1-3	
Feb 14	Tue	Classification, naive Bayes and LDA	ISL 4.4-6	1
Mar 16	Thu	Linear algebra for machine learning: review	LA 3,4,6	
Feb 21	Tue	Linear algebra and optimization	CO 2,3	project description, data
Feb 23	Thu	Overfitting and resampling methods	ISL 5.1	2
Feb 28	Tue	Cross-validation and bootstrapping	ISL 5.2	
Mar 02	Thu	Linear model selection, priors	ISL 6.1	evaluation
Mar 07	Tue	Midterm review		3
Mar 09	Thu	**Midterm exam**; material until 2/23		
Mar 14	Tue	*Spring break, no class*		
Mar 16	Thu	*Spring break, no class*		
Mar 21	Tue	Linear model selection and regularization	ISL 6.2	
Mar 23	Thu	Building nonlinear features	ISL 7.1-3	lit. review and method
Mar 28	Tue	Nearest neighbor methods and GAMs	ISL 7.4-7	4
Mar 30	Thu	Tree-based methods and boosting	ISL 8	
Apr 04	Tue	Support vector machines and other techniques	ISL 9	
Apr 06	Thu	Unsupervised learning, PCA	ISL 10	preliminary results
Apr 11	Tue	Reinforcement learning	RL 1,3,4	5
Apr 13	Thu	Neural networks and deep learning	DL	
Apr 18	Tue	Neural networks and deep learning	DL	
Apr 20	Thu	Big data and machine learning		
Apr 25	Tue	Machine learning in practice		6
Apr 27	Thu	Project presentations		final report
May 02	Tue	Guest speaker		
May 04	Thu	Final exam review		
May 11-17	?	**Final exam**		