Rutgers, The State University of New Jersey
Ph.D. Program – IT Major
Syllabus
Machine Learning for Data Science
Spring 2017
Tuesday- 1:00-3:50pm

Instructor Prof. Nabil Adam
TA: Ms. Debopriya Ghosh
Office : 1029, 1 Washington Park
Office Hours :TBA
E-mail : adam@andromeda.rutgers.edu
Homepage : http://idsla.newark.rutgers.edu/

Purpose: A huge amount of data has been, and continues to be collected from disparate sources and technology advances allow storing and processing greater quantities of data. The term "Big Data", which has been commonly used in the research community, industry, and government, is typically characterized by not only volume but also variety, and velocity. This ever growing data in volume, granularity, variety, presents major challenges with respect to learning from these massive amount of data. Machine learning techniques help understand the process and make prediction for the future. Specifically, machine learning techniques enable identifying patterns, trends, and anomalies in these datasets, and summarizing them with underlying models, thus help turn data into information, turn information into knowledge, and turn knowledge into action.

The purpose of this course is to present advanced topics of machine learning and algorithms. We will explore various approaches currently available in the literature and their application to real-world problems such as personalized healthcare, sentiment analysis, and community detection.

Prerequisite: 26:198:644 Data Mining (or equivalent); 22:198:603 Database Systems (or equivalent); 26:960:577:01 Linear Statistical Models (or equivalent); Working knowledge of R/ Python and SQL**.**

Reading Material: There is no text assigned to this course at the moment. However, the following books either cover one topic in depth or cover some of the preliminary concepts of the topics. In addition to the books listed below, the reading list includes a number of research papers.

Text Book / Supplementary Readings

1. **Peter Flach. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge: Cambridge University Press, 2012.**
2. **Brett Lantz. *Machine Learning with R - Second Edition*. 2nd Edition, Packt Publishing, 2015.**
3. **Alpaydin, Ethem. *Machine Learning: The New AI*. MIT Press, 2016.**
4. **James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning: With Applications in R. New York: Springer, 2013.**

Expected Work:
1. Class Participation: 5%.

2. Home Work Assignments: 15%.
3. Mid-term Exam: 20% **(March 21, 2017)**
4. Research Paper/Project and Presentation 30% **(April 25 and May 2, 2017)**
5. Final Examinations 30% **(May 9, 2017)**

Date	Topic
January 17, 2017	**Lecture 1: Introduction** • What is data mining? • Assigned Reading: 50 years of data science • http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf • Moore's Law • How large is large? • Exploratory Data Analysis and Visualization (Examples) • Supervised vs Unsupervised Learning • Examples of applications to Big Data • Quick Review of Basic R programming
January 24, 2017	**Lecture 2: Computational Analysis** **Lecture 3: ML for Big Data Analytics**

January 31, 2017
February 7, 2017
	Lecture 4: Supervised Learning
	• Classification
	• Linear Models
	  ○ Linear Regression
	  ○ Logistic Regression (binomial, multinomial, ordinal)
	• Naïve Bayes
	• Support Vector Machines (SVDs)

February 14, 2017	**Lecture 5: Feature Selection** • Penalized models (elasticnet and glmnet) • SVD • PCA • Enriched PCA
February 21, 2017	**Lecture 6: Decision Trees** • CART, Bump Hunting • Random Forest (bagging and boosting)
February 28, 2017 March 7, 2017	**Lecture 7a: Unsupervised Learning** • Clustering • Different Types of Clustering • K-means • Hierarchical Clustering • PAM • Model Based Clustering

	• Weighted K-means • Examples of applications to Big Data
March 14, 2017	**Spring Recess**
March 21, 2017	**Midterm Examination**
March 28, 2017	**Lecture 7a: Unsupervised Learning (Continue)** **Lecture 7b: Recommender System**
April 4, 2017	**Lecture 8: Text Mining** • Feature Extraction • Classification • NLP: Topic Modeling • Examples of Application of Text Mining
April 11, 2017	**Lecture 9: Graphical Models** • Bayesian networks • PDN • Examples of applications to Big Data
April 18, 2017	**Demonstration of Research Projects**
April 25, 2017 May 2, 2017	**Student Presentations**
May 9, 2017	**Final Examination**

References

[1] Domingos, P., "A few useful things to know about machine learning." CACM 55, 2012

[2] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 3-24.

[3] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data Mining with Big Data", IEEE Transactions on Knowledge and Data Engineering, Volume. 26, NO. 1, January 2014

[4] Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, Graham J. Williams, "Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives", Computational Intelligence Magazine, IEEE (Volume: 9, Issue: 4)

[5] Huang, Guang-Bin, et al. "Extreme learning machine for regression and multiclass classification." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42.2 (2012): 513-529.

[6] Havens, Timothy C., et al. "Fuzzy c-means algorithms for very large data."Fuzzy Systems, IEEE Transactions on 20.6 (2012): 1130-1146.

[7] Wang, Jingdong, et al. "Fast approximate k-means via cluster closures."Multimedia Data Mining and Analytics. Springer International Publishing, 2015. 373-395.

[8] Smola, Alex, and Vladimir Vapnik. "Support vector regression machines."Advances in neural information processing systems 9 (1997): 155-161.

[9] P. Indyk and R. Motwani. “Approximate nearest neighbor: towards removing the curse of dimensionality,” ACM Symposium on Theory of Computing, pp. 604–613, 1998