Course:			
Course type:	Elective	EE, CE, and IT*	Credit: 3
Level:	Undergraduate - Graduate		
Co-requisite(s):			
Prerequisite(s):	At least one of the following courses: Machine Learning, Pattern Recognition, Neural Networks		
Prerequisite by topic:			
Textbook(s):			
Coordinator:	Babak Arabi, Mohammad Amin Sadeghi		
Goals:	The amount of available data is increasing every day. This comes with an opportunity and a threat. The opportunity is that more and more data help us gain insights about our environment and make better decisions. The threat is that gaining insight is increasingly becoming more difficult. The goal of this course is to prepare students to derive insights from real-world data. The students will 1- use the latest tools and 2- analytical techniques, to 3- interpret data, 4- and communicate their new findings to inform others. This course is supposed to help students connect their mathematics and machine learning skills to real-world applications. This course prepares students to apply the latest statistical and machine learning methods for identifying patterns in and extracting insights from complex datasets, as well as for making predictions. Finally, students will also learn how to communicate the findings of data analysis effectively using visualization tools.		
Outcome:	Students after the course are expected to look at data and see things that other's don't see. In machine learning courses, students are expected to gain a good knowledge machine learning techniques. However, in		

	machine learning courses students are not expected to solve real-world problems using machine learning. In this course students are expected to gain skills to connect real-world problems to machine learning techniques. The students in this course are expected to be able to extract insights from real-world data science problems.
Topics:	Data science process - Philosophy of Data Science - Data Science Applications - Common mistakes in understanding data - data scraping/sampling/cleaning to prepare data - Optimal Design of experiments - Intuition and Hypothesis Learn to see things that others don't see - Introduction to Exploratory Data Analysis - Visualize to explore data - Histograms - A/B testing and Comparison - correlations and causality - associations, tradeoffs - time-series data and Trends Visualization and communication - Good vs Bad Visualization - Visualize to communicate findings - Visualize to persuade audience - Storytelling with data Statistical Inference - The problem of a-priori distribution in Bayesian learning - A Bayesian approach to classification - A Bayesian approach to regression - Statistical Modeling - Graphical Models - Bayesian networks - Temporal Models, Markov Chains - Hidden Markov models Predictive Analytics - Examples of Predictive Analytics Problems - Regression - Recommender Systems Deep learning - Intro: Neocognitron (Fukushima), HMax (Poggio) - Convolutional Neural Networks (LeCun) - Deep Belief Networks (Hinton) - Boltzmann Machine & Restricted Boltzmann Machine

	- ImageNet dataset & AlexNet breakthrough - On RNN, and RL in deep learning, - Deep Learning applications - Distributed Learning Machine learning with less supervision Fuzzy clustering, Possibilistic clustering - Cluster prototypes and related clusterings - Fuzzy C-means, Gustafson-Kessel, and Gath-Geva clustering - Cluster validity for fuzzy clusters - Information theoretic clustering - semi-supervised learning - domain adaptation and transfer learning - gradual learning and Incremental learning - Recent approaches to the unsupervised problem - Decision fusion - Combination of scores - Combination of rankings - Combination of Week classifiers and boosting Privacy - Ethics ad privacy - Legal aspects of privacy - Data Encryption and Anonymization
Computer usage:	The students will use computer to explore and visualize data. They will use R
Assignments:	Students will have practical/programming assignments to experience a number of big data technologies.
Projects:	The students need to explore real-world problems and extract insight. This course will have a number of assignments on kaggle.com
Grading:	Assignments 60% Midterm 15% Final Exam 25%
Further readings:	
Prepared by:	Mohammad Amin Sadeghi
Date:	Jan 5, 2017

*EE: Electrical EngineeringCE: Computer EngineeringIT: Information Technology