16.711 (203) Special Topics: Computational Data Modeling is a new interdisciplinary graduate course offered to students from the Colleges of Engineering, Science and Business. The course objectives are to provide the student analytical and computational skills for deciphering information in large data sets with application towards developing discipline specific prediction, forecasting and decision models. Statistical methods that serve as a foundation for exploratory data analysis, characterization, hypotheses and inference will be presented using an interactive R-programming platform. A particular focus of the course is on learning algorithms and machinery that can process big data sets efficiently with respect to time and computing resources. To this end, students will learn to map problems to graphical processor unit (GPU) platforms for parallel processing and apply message-passing interface (MPI) for computation on a high-performance computing cluster (HPCC). Case studies, with data drawn from the disciplines of systems biology, communications, air-traffic and social networks, text analysis, plastics manufacturing, education and medicine will be presented by faculty undertaking research in these areas.

Date: FALL 2013, UML N. Campus, Wednesdays, 6:30 : 9:20 pm

Room: Ball 323

Lead Instructors: K. Chandra (ECE), K. Daniels (CS), C. Thompson (ECE)

Supporting Faculty: D. Kazmer (Plastics), L. Motiwalla (Business), V. Mehta (MIT Lincoln Labs), A. Oztekin (Business), S. Yoon (Chemical Eng.), X. Zhang (Civil)

Text Book: Pattern Recognition and Machine Learning , C.M. Bishop, Springer 2009.

References: Course notes, papers to be read and discussed and R – related resources will be provided.

9/4/13	Ch1: Data Modeling Overview; Supervised & Unsupervised, Learning	**R:** Introduction and interactive tutorial	Introduction to Probability concepts: Bayes theorem; Gaussian distribution; MLE Concepts	**R:** Linear regression example; Error Analysis; Overfitting; Regularization; Pearson correlation; P-values
9/11/13	Ch 2: Fitting Probability density functions (pdf) to data; Design of prior distributions for Bayesian Estimation	**R:** Generate random variables from model pdf and compare with estimated pdf;	Development of predictive posterior distributions through sequential learning;	**R:** Algorithm for sequential estimation(Robbins-Monro)
9/18/13	**Case Study A**			
9/25/13	Ch.3: Supervised Learning: Linear Models for Regression; Basis functions	**R :** Sequential learning algorithm: Stochastic gradient descent (LMS)	Ch.3: Bayesian Linear Regression	**R:** Analyzing multiple ensembles of data: parallel processing on GPU's : Rpu Library
10/02/13	Ch. 4: Linear Models for Classification; Discriminant Functions: 2 class, multiple class (Fisher's Discriminant)	**R:** Analyzing classification models using network data	Logistic Regression for Classification	**R:** Comparing performance of classifiers using least-squares and logistic regression
10/09/13	**Case Study B:**			
10/23/13	**Ch. 6 & 7 :** Kernel Methods &	**R:** Examples of constrained optimization	Discussion of problem definition	**R:** Exploratory data

	Support Vector Machines (SVM's)	and quadratic programming	and data sets for analysis in group projects	analysis: Group design and discussion of results
10/30/13	**Ch.7:** SVM's continued	**R:** Performance of a two class classifier using linear models and SVM's	**R:** Group Projects continued	
11/06/13	**Case Study C:**			
11/13/13	**Ch. 9 :** Mixture Models and Expectation Maximization	**R:** K-means clustering application	**R:** Group Project Presentations	
11/20/13	**Case Study D:**			
11/27/13	**Ch.11:** Sampling Methods: Monte Carlo and Gibbs	**R:** Application of Monte Carlo methods: Sampling from empirical pdfs;	**R:** Group Project Presentations	
12/04/13	**Ch. 12:** Principal component analysis	**R:** PCA application for high dimensional data		
12/11/13	WRAP UP …..			