Course title: Pattern Recognition and Data Mining

Quarter/Year: Fall 2016

Course Number: COEN 281

Instructor: Dr. Manish Marwah

Email: mmarwah@scu.edu

Class times: Tue (7:10pm-9:00pm) and Sat (11:10am-1:00pm).

Location: Kenna Hall 105

Course description:
This is a graduate level introductory data mining / machine learning course. The basic question we will try to address in this course is: How to *understand* data and use it to derive *insights* and make *predictions*? The focus of the course will be on data mining *algorithms* and *techniques*, and their *applications.* We will study: 1) how to explore and visualize data; 2) supervised learning techniques, such as, k-NN, logistic regression, decision trees, ensemble methods, support vector machines, naïve Bayes, neural networks; 3) unsupervised learning techniques, such as, different clustering methods, principal component analysis; 4) deep learning methods such as convolutional neural networks; 5) recommender systems and text analytics. The prerequisites of the course are: multivariate calculus, linear algebra, probability and statistics, and a good amount of programming experience, preferably in Python or R.

Course objectives:
The main objective of the course are:
- To learn and understand the basic data mining and machine learning methods
- To be able to implement these methods
- To be able to apply these methods to real data
- To be able to evaluate a particular method
- To be able to compare different methods/techniques
- To be able to pose a data mining question, use a data set and the methods learned to try to answer that that question
- To be able to read about data mining in the popular press with a critical eye

Projects, assignments, and exams in this course will evaluate the student's ability to meet the course objectives.

Course material:
There is no textbook required for this course. Camino will be used for all course material. The following books are good references.
- Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Pearson/Addison Wesley.
- Data mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers
- Pattern Recognition and Machine Learning, C. Bishop

- Deep Learning, Ian Goodfellow et al. https://www.deeplearningbook.org/

Homework:
There will be three HWs which will contain both problems and programming assignments. You will typically have about two weeks to complete a HW. The programming assignments with be in R or python.

Late policy:
There are 5 late days that you can use as you wish without any penalty. Beyond that there is a 25% penalty per day.

Exams:
The course has a midterm and a final examination. Students must take the examination on the scheduled day/time. If there is a valid reason you cannot make it to the exam, please talk to me *prior* to the exam.

Group project:
Each student group (3-4 people) is required to propose a project that uses data mining on real dataset to solve an interesting problem. More details on the project will be provided in the lecture.

Class participation:
Participation in classroom discussion and on the Camino discussion board are encouraged. In fact, 5% of the grade is reserved to class/discussion board participation.

Grading:

Homework	25%
Class participation	5%
Scribe notes	5%
Project	25%
Mid-term	20%
Final	20%

Academic Policy:
Student responsibility: By enrolling in this class, a student is bound by the requirements stated in this syllabus and given by the instructor. Students must operate with integrity in their dealings with faculty and other students; engage the learning materials with appropriate attention and dedication; maintain their engagement when challenged by difficult learning activities; contribute to the learning of others; and perform to standards set by the faculty.
Classroom behavior: Students should behave properly during class time so as to facilitate the teaching and learning. You should turn off all electronic devices that may affect the class.
Academic integrity (Honor Code): Academic dishonesty is prohibited and is considered a violation of the university rules. It includes, but is not limited to, cheating, plagiarism, and collusion. All students taking courses in the school of Engineering agree, individually and collectively, that they will neither give nor receive unpermitted aid in examination or other course work that is to be used by the instructor as a basis of grading. Violation will be taken seriously and will be referred to the Office of Student Life and other senior university officials for possible disciplinary action. This information will become a part of

the academic records of the students who violated the policy. You are responsible for understanding the policy available at: http://www.scu.edu/studentlife/resources/academicintegrity/index.cfm.

To request academic accommodations for a disability, students must contact Disabilities Resources located on the second floor of Benson. Phone numbers are (408) 554-4111; TTY (408) 554-5445. Students must register and provide documentation of a disability to Disabilities Resources prior to receiving academic accommodations.

Tentative Course Schedule

	Topic	Comment
Week 1	Introduction, Data, Exploratory data analysis, Visualization	HW1 out
Week 2	Supervised learning basics, cross validation, overfitting, evaluation, linear regression, k-NN, Trees	
Week 3	Logistic regression, SVM, Bias-Variance trade-off, ROC curve, Ensembles	HW2 out
Week 4	Probability review, naïve Bayes	
Week 5	Neural Networks, Debugging learning algorithms	Project Proposal due
Week 6	Review and Midterm	
Week 7	Unsupervised Methods – PCA, Clustering – k-means, hierarchical, spectral	HW 3 out
Week 8	Deep learning, recommender systems	Project checkpoint report due
Week 9	Project Presentations	
Week 10	Review and Exam	Project report due

Final: Dec 6, 7:10 pm		