THE UNIVERSITY OF TEXAS AT EL PASO
COLLEGE OF SCIENCE
DEPARTMENT OF MATHEMATICAL SCIENCES

Course #:	STAT 5494 (CRN 16606)
Course Title:	Statistical Data Mining
Credit Hours:	4 (Lecture plus Lab)
Term:	Fall 2020 (Instruction 08/24/2020 – 12/03/2020)
Course Meetings & Location:	MW 9:00 am – 10:20 am, Hudspeth Hall 213
	MW 10:30 am – 11:20 am, Hudspeth Hall 213
Prerequisite Courses:	STAT 4380 and STAT 5380 or Instructor Approval. Linear regression and generalized linear models (GLM) and some programming experiences would be a plus.
Instructor:	Xiaogang Su
Office Location:	Bell Hall 320
Contact Info:	Phone: (915) 747-6860 [O]
	Email: xsu@utep.edu
	Fax: (915) 744-6502
Office Hours:	MW 4:00 pm – 5:00 pm
Class Web page:	https://sites.google.com/site/utepstat5494/

Textbook(s), Materials:	Required:	Hastie, T., Tibshirani, R., and Friedman, J. H. (2008), *Elements of Statistical Learning*, 2nd Edition Chapman and Hall. ISBN-13: 978-0387848570 http://www-stat.stanford.edu/~tibs/ElemStatLearn/
	Suggested:	• Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley
		• Clarke, B., Fokoue, E., and Zhang, H. (2009). *Principles and Theory for Data Mining and Machine Learning*. Springer. ISBN-13: 978-0387981345
		• Bishop, C. M. (1996). *Neural Networks for Pattern Recognition*. Oxford University Press: USA. ISBN-13: 978-0198538646.

Course Description and Learning Outcomes:	Data Mining emerges as an interdisciplinary field with joint inputs from statistics, computer science, machine learning, and artificial intelligence. This course is intended to cover some commonly-used data mining techniques, with more focus on the most technical part - statistical learning algorithms. The materials are arranged in two main categories: unsupervised learning and supervised learning. A tentative outline of the specific topics is provided below.

Topic Outline   1. An Overview of Data Mining;
2. Intro to Optimization;
3. Intro to Reproducing Kernel Hilbert Space (RKHS);

Part I: Unsupervised Learning
4. Association Rules: Market Basket Analysis; the Apriori Algorithm; Generalized Association Rules.
5. Outlier Detection: Statistical Discordance Tests, Nearest Neighbor and Distance-Based; Partitioning-Based; Influential Points.
6. Non-Negative Matrix Factorization: Archetypal Analysis.
7. Web Mining: Google PageRank

Part II: Supervised Learning
8. Parametric Nonlinear Regression: Numerical Optimization methods;
9. Nonparametric Smoothing I - the Kernel Method: K-nearest neighbor (KNN); kernel regression;
10. Nonparametric Smoothing II - the Spline Method: Polynomial Regression and Piecewise Polynomials; (Natural) Cubic Splines; B-Splines; Smoothing Splines - the penalized likelihood.
11. Generalized Additive Models (GAM): Additive Models and GAM.
12. Multivariate Adaptive Regression Splines (MARS); Generalization to different types of responses.
13. Projection Pursuit Regression (PPR): the backfitting algorithm
14. Artificial Neural Networks - I: Multilayer Perceptron (MLP): ANN history and architectures; Single-Layer Perceptron and MLP; Back Propagation (Backprop) and other optimization methods; deep learning
15. Artificial Neural Networks - II: Radial Basis Function Networks (RBF): Ordinary RBF and Normalized RBF.
16. Support Vector Machines (SVM)

Part II: Additional Topics (if time allows)
17. Hidden Markov Models (HMM)
18. Reinforcement Learning

Course Schedule: 09/07   Labor Day
10/30   Course Drop Date
11/26 - 11/27   Thanksgiving
12/07 -12/11   Final Exam Week

Course Activities/Assignments: Presentations and projects will be assigned throughout the semester. NO LATE COURSEWORK WILL BE ACCEPTED, EXCEPT EXTREME SCENARIOS.

Final Presentations: TBA

Assessment of Course Objectives: Each student will be evaluated by the quality of his/her own assigned presentations as well as their contribution to the discussions during other students' presentations.

Grading Policy:	There will be a number of computer assignments and a final project. The assignments make up 70% and the final project makes up 30% to your final score. For the final project, students are given the freedom to select data from whatever field they are interested in. Students should make their own plans to collect data, raise interesting research questions, and consult the instructor for the adequacy of the project. Also, each student will have the opportunity to present their work in class. There will also be a few in-class quizzes or exams, which make up 5% extra credit; however, these five extra credits are only applicable to those who complete all assignments and exams without ANY unexcused absence from class attendance. No make-up exam will be given and no late project submission is accepted without justifiable reasons.

Letter grades are determined according to the following scale:
Grade Score

A	90 +
B	80-89
C	70-79
D	60-69
F	<60

Make-up Policy:	All other assignments must be turned in on time.
Attendance Policy:	Class attendance is REQUIRED and helpful to decide borderline grades. If a student has to be absent from a particular class, he/she will be responsible for catching up with course material. A late arrival of 15 minutes or more will be considered as an absence. Students will be dropped for four or more unjustified absences from class or lab session. Your academic advisor will be consulted before final action is decided and taken. Any unjustified absences from class or lab session will cause loss of eligibility of receiving extra credits. If you expect to miss up to 10 class hours for ANY REASON, then please do not consider taking this course.
Academic Integrity Policy:	Please see http://academics.utep.edu/Default.aspx?tabid=23785
Civility Statement:	This is a class where participation is required. You will be participating in classroom discussions. All students will be treated with respect.
Disability Statement:	If you have a disability and need classroom accommodations, please contact The Center for Accommodations and Support Services (CASS) at 747-5148, or by email to cass@utep.edu, or visit their office located in UTEP Union East, Room 106. For additional information, please visit the CASS website at www.sa.utep.edu/cass.
Military Statement:	If you are a military student with the potential of being called to military service and /or training during the course of the semester, you are encouraged to contact me as soon as possible.

| UTEP College of Science Policies | Watch out for the UTEP drop/withdraws deadline for the semester.   The College of Science will remain aligned with the University and not approve any drop requests after that date. |

All grades of Incomplete must be accompanied by an Incomplete Contract that has been signed by the instructor of record, student, departmental chair, and the dean.   Although UTEP will allow a maximum of one year to complete this contract, the College of Science requests it be limited to month based upon completion data.   A grade of Incomplete is only used in extraordinary circumstances confined to a limited event such as a missed exam, project, or lab.   If the student has missed a significant amount of work (e.g. multiple assignments or tasks), a grade of Incomplete is not appropriate or warranted.