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

Course #:	STAT 5494 (CRN 16961) DS 6494 (CRN 16856)
Course Title:	Statistical Machine Learning II
Credit Hours:	4 (Lecture plus Lab)
Term:	Fall 2024 (Instruction 08/26/2024 Monday – 12/05/2024 Thursday)
Course Meetings & Location:	The course will be held asynchronously online. Course materials can be accessed via Blackboard, and I am available on Zoom during class/office hours: https://utep-edu.zoom.us/j/89893259003
Prerequisite Courses:	STAT 5380, STAT 5474 or Instructor Approval.
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:	1:00 – 2:00 pm MW or by appointment
Class Web page:	Blackboard
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. Naïve Bayes Learning
18. Hidden Markov Models (HMM)
19. Reinforcement Learning

Course Schedule:

09/02	Labor Day
11/01	Course Drop Date
11/28 - 11/29	Thanksgiving Holiday
12/09 - 12/13	Final Exam Week

Course Activities/Assignments: Presentations and projects will be assigned throughout the semester.

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 several 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.

To promote a strong work ethic, late unexcused submissions will be penalized by 10% deduction per day (e.g., 10% for 0-24 hours late, 20% for 24-48 hours late, etc.). 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: The course will be held online asynchronously, without any attendance checks. However, it's your responsibility to consistently review and study the course materials.

Academic Integrity Policy: Please see http://academics.utep.edu/Default.aspx?tabid=23785

Civility Statement: 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.
COVID-19 Precaution Statement	Please stay home if you have been diagnosed with COVID-19 or are experiencing COVID-19 symptoms. If you are feeling unwell, please let me know as soon as possible, so that we can work on appropriate accommodations. If you have tested positive for COVID-19, you are encouraged to report your results to covidaction@utep.edu, so that the Dean of Students Office can provide you with support and help with communication with your professors. The Student Health Center is equipped to provide COVID-19 testing. The Center for Disease Control and Prevention recommends that people in areas of substantial or high COVID-19 transmission wear face masks when indoors in groups of people. The best way that Miners can take care of Miners is to get the vaccine. If you still need the vaccine, it is widely available in the El Paso area, and will be available at no charge on campus during the first week of classes. For more information about the current rates, testing, and vaccinations, please visit epstrong.org.