CSE 572: Data Mining

General Information

Time and Location

Tuesdays and Thursdays 12:00pm – 1:15 pm @ CAVC 359

Contact Information

Instructor: Dr. Shayok Chakraborty

Office: BYENG 376

Office Hours: Tuesdays and Thursdays 10:00 am – 11:00 am or by appointment

[Cancellation of office hours due to travel or other urgent university business will be announced on Blackboard in advance].

Email: shayok.chakraborty@asu.edu

Teaching Assistant: Ze Gong (zgong11@asu.edu)

Office: BYENG M1 – 30

Office Hours: Mondays and Wednesdays 2:00 pm – 3:00 pm

Textbooks

Introduction to Data Mining

Authors: Pang-Ning Tan, Michael Steinbach and Vipin Kumar

Website: http://www-users.cs.umn.edu/~kumar/dmbook/index.php

Software

We will use MATLAB as the software for this course. Please download the latest version from the ASU My Apps website.

Topics to be Covered

This course will introduce the fundamentals of data mining and pattern recognition. Topics to be covered include:

Basic Data Analysis

- Introduction to data mining
- Attribute and types of attribute
- Types of data
- Data Quality
- Data Preprocessing
- Data Transformation
- Measures of similarity and dissimilarity
- Measures of central tendency and dispersion
- Data Exploration

Classification

- Nearest Neighbor
- Decision Tree
- Bayesian Classifiers
- Artificial Neural Networks
- Support Vector Machines
- Kernel Learning
- Ensemble Methods – Bagging and Boosting
- Adaboost and Random Forests
- Class Imbalance Problems
- ROC Curves
- Model Selection, Evaluation and Cross Validation
- Model Overfitting and Bias Variance Trade-off

Clustering

- Motivation
- k-means clustering
- Hierarchical clustering
- DBSCAN

Association Rule Mining

- Problem Definition
- Confidence and Support
- Apriori Principle
- Rule Generation

Advanced Topics (Time Permitting)

- Dimensionality Reduction
- Semi-Supervised Learning

- Active Learning
- Transfer Learning

Prerequisites

Students are expected to have a working knowledge of basic probability theory and linear algebra. In addition, proficiency in MATLAB programming is required for the mini-projects.

Assessment

Below is a breakdown of percentages needed for the corresponding letter grades.

Grade	Percentage
A+	95-100
A	90-94
A-	85-89
B+	80-84
B	75-79
B-	70-74
C+	65-69
C	60-64
C-	55-59
D	50-54
E	Below 50

Below is a breakdown of assignments and exams, their assigned weights and a brief description of each.

Assignments (60%): *Students will need to complete 5 or 6 assignments, each will have equal weightage. Some of the assignments will be based on implementation.*

Midterm Exam (20%): *The midterm exam (in class) will be held in March. The date is not yet finalized. The details will be posted in Blackboard.*

Final Exam (20%): *The final exam (in class) will be held on Tuesday, May 2, 2017. It will be comprehensive. More details will be posted in Blackboard.*

We will not have a separate course project for this course; rather, some of the assignments will be based on implementations of concepts taught in class and will be termed mini-projects. The mini-

projects will have at least 30% weightage, so that MCS students can use this course for their MCS project portfolio.

Absence & Make-Up Policies

You are expected to submit all assignments on the due date. Late submissions will not be accepted. Students who expect to miss class due to officially university-sanctioned activities should inform the instructor early in the semester. Alternative arrangements will generally be made for any examinations and assignments affected by such absences.

Classroom Behavior

Cell phones must be turned off or set to vibrate during class to avoid causing distractions. Any violent or threatening conduct by an ASU student in this class will be reported to the ASU Police Department and the Office of the Dean of Students.

Academic Integrity

All students in this class are subject to ASU's Academic Integrity Policy (available at http://provost.asu.edu/academicintegrity) and should acquaint themselves with its content and requirements, including a strict prohibition against plagiarism. All violations will be reported to the Dean's office, who maintain records of all offenses. Specific rules for this class are as follows. All assignments and projects must be your own individual work, unless specified as team efforts. You are encouraged to learn from each other but copying is strongly discouraged. All solutions turned in for credit are to be your individual work and should demonstrate your problem solving skills. The instructor reserves the right to question a student orally or in writing and to use his evaluation of the student's understanding of the assignment and of the submitted solution as evidence of cheating. Violators of this policy may be faced with penalties.

Disability Accommodations

Suitable accommodations will be made for students having disabilities and students should notify the instructor as early as possible if they will require same. Such students must be registered with the Disability Resource Center and provide documentation to that effect.

Note: Any information in this syllabus may be subject to change with reasonable advance notice.