Course Syllabus

Jump to Today

Arizona State University

CSE 572: Data Mining

Course and Faculty Information

Course Description: Advanced data mining techniques: classification, clustering, association, preprocessing; performance evaluation; information assurance, Web mining, security and privacy issues, and other applications. Students must have a solid background in database management systems, search, learning, and statistics to be successful in this course.

Credits: 3

Prerequisites: Computer Engineering or Computer Science or Software Engineering graduate student or Online CS nondegree-seeking graduate student. 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.

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)

- Reinforcement Learning
- Deep Learning
- Transfer Learning
- Deep Mind for Google

Course Learning Outcomes

At the completion of this course, students will be able to:

1.1 Explain the history and Purpose of data mining across multiple disciplines
1.2 Differentiate what is and what is not data mining

1.3 Describe different data mining tasks
1.4 Recognize attributes of data needed for data mining
1.5 Review and summarize data exploration techniques for use in initial data analysis

2.1 Define classification and classification applications
2.2 Compare and contrast common classification techniques
2.3 Apply common algorithms used in data mining
3.1 Define Instance Based Classifiers
3.2 Use the basics of probability theory to calculate the Bayes Classifier
3.3 Use the probability estimation to calculate Naive Bayes Classifier
3.4 Recognize the basic structure of Neural Networks
3.5 Identify the Perceptron learning algorithm
3.6 Recall the Artificial Neural Networks learning model
3.7 Support vector machines
4.1 Define cluster analysis
4.2 Differentiate what is and what is not cluster analysis
4.3 Categorize different types of clusters
4.4 Use common algorithmic measures to evaluate clusters
4.5 Analyze DB Scan in relation to other clustering methods

5.1 Apply the Mining Association Rules to discover relationships in large datasets
5.2 Use inferencing techniques to analyze association rule analysis results
5.3 Identify ways to reduce the computational complexity of frequent itemset generation
5.4 Describe how to efficiently generate rules from frequent datasets

6.1 Describe components that comprise deep learning
6.2 Implement a deep neural network using common tools such as Keras or Theano
6.3 Describe the structure and usage of Restricted Boltzmann Machines
6.4 Design Restricted Boltzmann Machine algorithm to create a movie recommendation application
6.5 Describe the structure of deep autoencoders and describe different application scenarios where they can be used
6.6 Apply deep autoencoders to on a sample data to derive low dimensional representations
6.7 Compare open source tools that allow for the fast implementation of data mining tasks

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 (http://www-**

users.cs.umn.edu/~kumar/dmbook/index.php)

Course Access

Your ASU courses can be accessed by both **my.asu.edu** (http://my.asu.edu) and **myasucourses.asu.edu** (http://myasucourses.asu.edu); bookmark both in the event that one site is down.

Computer Requirements

This is a fully online course; therefore, it requires a computer with internet access and the following technologies:

- Web browsers (**Chrome** (https://www.google.com/chrome), **Mozilla Firefox** (http://www.mozilla.org/en-US/firefox/new/), or **Safari** (http://www.apple.com/safari/))
- **Adobe Acrobat Reader** (http://get.adobe.com/reader/) (free)
- **Adobe Flash Player** (http://get.adobe.com/flashplayer/) (free)
- Webcam, microphone, headset/earbuds, and speaker
- Microsoft Office (**Microsoft 365 is free** (https://myapps.asu.edu/app/microsoft-office-2016-home-usage) for all currently-enrolled ASU students)
- Reliable broadband internet connection (DSL or cable) to stream videos.

Note: A smartphone, iPad, Chromebook, etc. will not be sufficient for completing your work in ASU Online courses. While you will be able to access course content with mobile devices, you must use a computer for all assignments, quizzes, and virtual labs.

Software

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

Student Success

To be successful:

- check the course daily
- read announcements
- read and respond to course email messages as needed
- complete assignments by the due dates specified
- communicate regularly with your instructor and peers
- create a study and/or assignment schedule to stay on track
- access **ASU Online Student Resources** (http://goto.asuonline.asu.edu/success/online-

resources.html)

Grading

Your grade will be determined based on the following grading schema:

Grade	Percentage
A+	100% - 97%
A	<97-94%
A-	<94-90%
B+	<90-87%
B	<87-84%
B-	<84-80%
C+	<80-77%
C	<77-70%
D	<70-60%
E	<60%

Submitting Assignments

All assignments, unless otherwise announced, MUST be submitted to the designated area of Canvas. Do not submit an assignment via email.

Assignment due dates follow Arizona Standard time. Click the following link to access the **Time Converter** (http://www.thetimezoneconverter.com/) to ensure you account for the difference in Time Zones. Note: Arizona does not observe daylight savings time.

Grading Procedure

Grades reflect your performance on assignments and adherence to deadlines. Grades on assignments will be available within 72 hours of the due date in the Gradebook.

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

Assignments (60%):

Students will need to complete 3 assignments. The assignments have equal weights of 20% for each assignment.

Midterm Exam (20 %)

Online midterm exam upto Module 4

Final Exam (20%):

The final exam (online) will be comprehensive. More details will be posted on Canvas. We will not have a separate course project for this course.

If any student wants to have this course as a portfolio, then two extra tasks are required. We can discuss those tasks after the first two assignments are done.

Late or Missed Assignments

Notify the instructor **BEFORE** an assignment is due if an urgent situation arises and you are unable to submit the assignment on time.

Follow the appropriate University policies to request an **accommodation for religious practices (http://www.asu.edu/aad/manuals/acd/acd304-04.html)** or to accommodate a missed assignment **due to University-sanctioned activities (http://www.asu.edu/aad/manuals/acd/acd304-02.html)**.

Communicating With the Instructor

Community Forum

This course uses a discussion topic called "Community Forum" for general questions and comments about the course. Prior to posting a question or comment, check the syllabus, announcements, and existing posts to ensure it's not redundant. You are encouraged to respond to the questions of your classmates.

Email questions of a personal nature to your instructor. You can expect a response within 72 hours.

Chat

The Chat tool in Canvas allows students and teachers to interact in real time. Use Chat only for informal course-related conversations unless your instructor informs you otherwise. Chat is not ideal for questions about assignments; instructors are not required to monitor it and conversations may be buried or lost.

Email

ASU email is an **official means of communication (http://www.asu.edu/aad/manuals/ssm/ssm107-03.html)** among students, faculty, and staff. Students are expected to read and act upon an email in a timely fashion. Students bear the responsibility of missed messages and should check their ASU-assigned email regularly.

All instructor correspondence will be sent to your ASU email account.

ASU Online Course Policies

View the **ASU Online Course Policies (https://asuonline-dev.asu.edu/qm-template/CanvasQM/qm-policies.html)**

Accessibility Statements

View the Accessibility section to review accessibility statements for common tools and resources used in ASU Online courses.

If any other tools are used in this course, links to the accessibility statements will be listed below this sentence.

Syllabus Disclaimer

The syllabus is a statement of intent and serves as an implicit agreement between the instructor and the student. Every effort will be made to avoid changing the course schedule but the possibility exists that unforeseen events will make syllabus changes necessary. Remember to check your ASU email and the course site often.

Academic Integrity

All students in this class are subject to ASU's Academic Integrity Policy (available at **http://provost.asu.edu/academicintegrity (https://provost.asu.edu/academic-integrity)**) 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 offences. 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.

Course Summary:

Date	Details	Due
	Total Grade Points (https://canvas.asu.edu/courses/133393/assignments/3548490)	