



## schedule

### WEEK 1

#### **Introduction and Applications**

September 12

- Topics
  - A Course Overview
  - Data Vocabulary
- [Suggested Reading](#)
- [Assignment 1 is assigned](#) - *Review & Exploring Data*

### WEEK 2

#### **Mining for Association Rules**

September 19

- Topics
  - Definitions of Frequent Itemsets
  - Determining Frequent Itemsets
  - Creating Association Rules
- [Suggested Reading](#)
- [Submissions](#)
  - [Assignment 1 is due](#)
  - [Assignment 2 is assigned](#) - *Association Rules*

### WEEK 3

#### **Accessing, Storing, and Computing with "Big" Data**

September 26

- Topics
  - Distributed Filesystems and Storage
  - Introducing the MapReduce Paradigm
  - Distributed Computation
- [Suggested Reading](#) - [Chapter 2, Sections 2.1-2.4](#)
- [Submissions](#)
  - [Assignment 2 is due](#)
  - [Assignment 3 is assigned](#) - *Map Reduce Problem*

### WEEK 4

#### **Large Scale Data (Pre)-Processing**

October 3

- Topics
  - Basics of Linear Algebra and Probability Theory
  - The Multiple Places Where Data Lives & Multi-source Joins
  - Covariance, Correlation, and Cosine Similarity
  - Dimensionality Reduction and Feature Selection
- Suggested Reading
  - [Linear Algebra Review](#)
  - [Dimensionality Reduction](#)
  - [Map Reduce, Sections 2.5-2.7](#)

### WEEK 5

#### **Mining Data without Labels**

October 10

- Topics
  - Introducing the Gaussian Distribution
  - Parameter Estimation of a Distribution
  - Anomaly and Outlier Detection
  - Unsupervised Modeling with k-Means and Clustering

- [Maximum Likelihood](#)
- [Unsupervised Clustering, Chapter 7 - 7.2](#)
- [Submissions](#)
  - [Assignment 3 is due](#)
  - [Assignment 4 is assigned](#) - *Parameter Estimation & Clustering*

---

**WEEK 6****[Mining Small-ish Data - Statistical Learning](#)**

October 17

- Topics
  - The Bayesian Framework
  - Naive Bayes Classification
  - Tree-based Algorithms - Random Forests
- Suggested Reading
  - [Naïve Bayes](#)
  - [Tree Algorithms - Chapters 3.1 - 3.3](#)
- [Submissions](#)
  - [Assignment 4 is due](#)
  - [Assignment 5 is assigned](#) - *Bayesian Framework & ML Libraries*

---

**WEEK 7****Midterm Exam**

October 24

- Topics
  - Linear Algebra Review
  - MapReduce Problems
  - Principle Component Analysis
  - Parameter Estimation
  - Unsupervised Clustering
  - Bayesian Framework

---

**WEEK 8****No Instruction This Week**

October 31

- Happy Halloween

---

**WEEK 9****[Mining Big Data - Foundations of Machine Learning](#)**

November 7

- Topics
  - The Classification Framework
  - The Objective Function, Regularization, and Constraints
  - Logistic Regression - Precursor to Modern Data Mining
  - Batch Data Processing - Gradient Descent
  - The Bias and Variance Tradeoff
- **In-Class Colabs:** Logistic Regression with MNIST
- Suggested Reading
  - [Evaluation Metrics](#)
  - [Logistic Regression \(\[1\], \[2\]\)](#)
- [Submissions](#)
  - [Assignment 5 is due](#)
  - [Assignment 6 is assigned](#) - *Gradient Descent*

---

**WEEK 10****[Mining Images with Deep Learning](#)**

November 14

- Topics
  - Working with Tensors - Reviewing Multivariate Calculus
  - Deep Learning - A Historical Perspective
  - The Backpropation Algorithm
  - Convolutional Neural Networks
- [Suggested Reading](#)

---

**WEEK 11****[Mining Text with Self Supervision](#)**

November 21

- Some Basic Approaches
- Semi-Supervised Learning
- The Concept of an Embedding Space
- The Attention Mechanism
- Large Language Models - From BERT to ChatGPT
- Suggested Reading
- [Submissions](#)
  - [Assignment 6 is due](#)
  - [Project proposals](#) are **due**

**WEEK 12**

### **Data Mining Applications**

November 28

- Topics
  - Social Network Data Mining
  - Recommendation Sciences
  - Time Series Analysis
- Suggested Reading

**WEEK 13**

### **Project Presentations and Industry Day**

December 5

- Data Mining in Industry
  - [Mining for Anomalous Behavior](#)
  - [Mining in Operational Logistics](#)
  - [Mining to Notify and Alert](#)
- [Submissions](#)
  - [Final projects](#) are **due**, including [presentation slides](#) and [writeup](#)

**WEEK 14**

### **Final Exam**

December 12

- Topics
  - Objective Functions
  - Logistic Regression
  - Association Rule Mining
  - Evaluation Metrics
  - Backpropagation
  - Convolutions and Recurrence

## grading criterion

<b>Labs &amp; Participation</b>	10%
<b>Data Mining Project</b>	10%
<b>Assignments</b>	20%
<b>Midterm Exam</b>	30%
<b>Final Exam</b>	30%

## course meeting times

### *Lectures*

- Tues, 6pm-9:20pm
- Room TBD

### *Office Hours*

- Professor, Thurs, 8:30-9:30pm
- TA, Date/Time TBD

## suggested textbooks

[Introduction to Data Mining, 2nd Edition](#) Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2018

---

[Mining of Massive Data Sets, 3rd Edition](#) Jure Leskovec, Anand Rajaraman, and Jeff Ullman, 2014

---

[Deep Learning](#) Ian Goodfellow, Yoshua Bengio, and Aaron Courville, 2016