syllabus

schedule

WEEK 1

Introduction and Applications

September 12

- Lecture Notes
- Lecture Recording
- Topics
  - A Course Overview
  - Data Vocabulary
- Suggested Reading
- Assignment 1

WEEK 2

Exploring, Storing, and Working with "Big" Data

September 19

- Lecture Notes
- Lecture Recording
- Topics
  - Distributed Filesystems and Storage
  - Introducing the MapReduce Paradigm
  - Distributed Computation
- Suggested Reading - Chapter 2, Sections 2.1-2.4
- Submissions
  - Assignment 1 is **due**
  - Assignment 2

WEEK 3

Large Scale Data Preprocessing

September 26

- Lecture Notes
- Lecture Recording
- Topics
  - The Multiple Places Where Data Lives & Multi-source Joins
  - Using MapReduce to Create Training Data
  - Covariance, Correlation, and Cosine Similarity
  - Dimensionality Reduction and Feature Selection
- Suggested Reading
  - Map Reduce, Sections 2.5-2.7
  - Evaluation Metrics

WEEK 4

Mining for Association Rules

October 3

- Lecture Notes
- Lecture Recording
- Topics
  - Definitions of Frequent Itemsets
  - Determining Frequent Itemsets
  - Creating Association Rules
- Suggested Reading
- Submissions
  - Assignment 2 is **due**
  - Assignment 3 is assigned

WEEK 5

Mining with Distributions

- Lecture Notes
- Lecture Recording
- Topics
  - Introducing the Gaussian Distribution
  - Parameter Estimation for Distributions
  - Detecting Anomalies with Naive Bayes
  - Unsupervised Modeling with k-Means and Clustering
  - Evaluation Metrics and Expected Precision/Recall
- Suggested Reading

WEEK 6

Mining Structured Data

October 17

- Topics
  - Reviewing Some Linear Algebra
  - An Introduction to Naive Bayes Classification
  - The Bayesian Framework - Posterior Probabilities
  - Tree-based Algorithms - Random Forests
  - Logistic Regression - A Precursor to AI Innovation
- In-Class Colabs
  - Logistic Regression
  - Fine-tuning MNIST
- Suggested Reading
- Submissions
  - Assignment 3 is **due**
  - Assignment 4 is assigned

WEEK 7

Mining Images

October 24

- Topics
  - Deep Learning - A Historical Perspective
  - The Backpropation Algorithm
  - Convolutional Neural Networks
  - Batch Data Processing
  - Evaluation Metrics and Expected Precision / Recall
- In-Class Colabs
  - Image Classification
- Suggested Reading

WEEK 8

No Instruction This Week

October 31

- Happy Halloween
- Assignment 4 is **due**

WEEK 9

Midterm Exam

November 7

- Topics
  - Linear Algebra Review
  - MapReduce Problems
  - Principle Component Analysis
  - Association Rule Mining
  - Parameter Estimation
  - Unsupervised Clustering
  - Bayesian Framework
  - Supervised Logistic Regression

WEEK 10

Mining Text

November 14

- Topics
  - Some Basic Approaches
  - Semi-Supervised Learning

        - The Attention Mechanism
        - Large Language Models - From BERT to ChatGPT
    - Suggested Reading
    - Assignment 5 is assigned

WEEK 11

Mining Graphs and Social Networks

November 21

- Lecture Notes
- Lecture Recording
- Topics
    - Some Basic Approaches
- Suggested Reading
- Submissions
    - Assignment 5 is **due**
    - Project submissions - slides and writeup

WEEK 12

Special Topics in Data Mining

November 28

- Lecture Notes
- Lecture Recording
- Topics
    - Recommendation Sciences
    - Time Series Analysis
    - Mining Structured Data
- Suggested Readings

WEEK 13

Project Presentations

December 5

- Lecture Notes
- Submissions
    - Project slides and writeup are **due**

WEEK 14

Final Exams

December 12

grading criterion

Labs & Participation	10%
Assignments	20%
Final Project	20%
Midterm Exam	25%
Final Exam	25%

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