MIDS: Introduction to Machine Learning

Dan Gillick
Joshua Blumenstock

Course Summary

Machine learning is a rapidly growing field at the intersection of computer science and statistics and concerned with finding patterns in data. It is responsible for tremendous advances in technology, from personalized product recommendations to speech recognition in cell phones. The goal of this course is to provide a broad introduction to the key ideas in machine learning. The emphasis will be on intuition and practical examples rather than theoretical results, though some experience with probability, statistics, and linear algebra will be important. Through a variety of lecture examples and programming projects, students will learn how to apply powerful machine-learning techniques to new problems, how to run evaluations and interpret results, and how to think about scaling up from thousands of data points to billions.

Prerequisites

1. Students must have completed the following core data science courses prior to enrollment:
    a. Research Design
    b. Storing and Retrieving Data
    c. Exploring and Analyzing Data
2. Undergraduate-level probability and statistics. Linear algebra is recommended.
3. Programming experience, preferably in Python.

Assignments and Grading

Course grades will be based mostly on four projects designed to synthesize concepts introduced in the lectures. Each project will require some programming (in Python) to set up experiments and a write-up summarizing results. Grading will be based on correctness and the ability to meet certain experimental milestones (e.g., reducing classification error rate by some percent). Each project will include at least one open-ended question to encourage more research. Course grades will be assessed according to four programming projects and accompanying write-ups.

Course Resources

Most textbooks on machine learning are written with considerable technical detail. As a result, there is no one textbook that aligns with this course. We will list readings that correspond to each week, including some general philosophy and landmark research papers, as well as few chapters from [Hal Daume's unfinished textbook].

Week 1: Introduction

- Overview of machine-learning applications
- Brief history
- Fundamentals of machine learning

Read Halevy, Norvig, Pereira, The unreasonable effectiveness of data

Week 2: Problem Setup and Nearest Neighbors

- Why prediction?
- Training and test data; cross-validation
- Evaluation and baselines
- Generalization and overfitting: linear models vs. nearest neighbors
- K-nearest neighbors, distance metrics
- Case study: real estate value, digit classification

Read Feynman, R. (1974, June). Cargo cult science. *Engineering and Science 37*(7).
Read Domingos, A few useful things to know about machine learning
Optional Hawkins, The problem of overfitting

Week 3: Supervised Learning I: Naive Bayes

- Probability review: Random variables, Independence, Bayes rule
- Generative models and Naive Bayes
- Maximum likelihood estimation and smoothing
- Case study: spam classification

Read Paul Graham on Naive Bayes (2002)
Optional Michael Collins tutorial on Naive Bayes (with math), see pages 1–4
Optional Norvig, How to write a spelling corrector

Week 4: Supervised Learning II: Decision Trees

- Decision Trees
- Information Gain
- Overfitting and pruning
- Ensemble methods
- Case study: customer churn, fuel efficiency

Read blog post from yhat about predicting churn
Optional Carter, An introduction to information theory and entropy
Optional Freund and Schapire, A Short introduction to Adaboost
Optional Delgado, et al, Do we need hundreds of classifiers to solve real-world problems?

Week 5: Supervised Learning III: Regression

- Review of linear regression
- Inference and prediction
- Logistic regression and classification
- Extensions and advanced topics

Optional: Breiman, Statistical Modeling: The Two Cultures
Optional: Deriving Least Squares
Optional: Freedman, Statistical Models and Shoe Leather

Week 6: Supervised Learning IV: More Linear Models

- Gradient descent for regression
- Regularization

Read Hal Daume, Gradient Descent (chapter 6)
Optional Bottou, Stochastic Gradient Descent Tricks

Week 7: Supervised Learning V: Neural Networks

- The perceptron
- State of the art: neural networks for speech recognition

Read Hal Daume, Neural Networks (chapter 8)
Optional LeCun, et al, Deep Learning

Week 8: Supervised Learning VI: SVMs, Choosing Classifiers, Speech Recognition

- Support Vector Machines
- Comparing classifiers: performance, training speed, model size, interpretability
- Feature engineering tips
- Speech recognition overview

Read An empirical comparison of supervised learning algorithms.
Optional Cosma Shalizi SVM lecture notes.
Optional On comparing classifiers: Pitfalls to avoid and a recommended approach.
Optional SKLearn classifier comparisons for toy problems.

Week 9: Unsupervised Learning I: Cluster Analysis

- What if our data don't have labels?
- Distance metrics (Hamming, Euclidean, Cosine, Mahalanobis)
- K-means clustering
- Hierarchical clustering

Optional Zhao and Karypis, Hierarchical clustering algorithms for document datasets
Optional Eisen et al., Cluster analysis and display of genome-wide expression patterns

Week 10: Unsupervised Learning II: Expectation Maximization

- Expectation-Maximization and the idea of hidden variables
- Basics of Gaussian Mixture Models
- Case study: speaker identification

Read Tibshirani lecture notes on EM
Optional Doug Reynolds original paper on GMMs for speaker identification

Week 11: Unsupervised Learning III: Dimensionality Reduction
- Motivation
- Dimensionality reduction
- Principal Component Analysis
- Case study: Eigenfaces
- Other methods for dimensionality reduction: SVD, NNMF, LDA

Optional Hal Daume, Unsupervised Learning (chapter 13)
Optional Turk and Pentland, Eigenfaces for Recognition

Week 12: Network Analysis
- Graph algorithms (pagerank).
- Network link predictions.
- Scaling and other challenges.

Read Page and Brin, The Anatomy of a Large-Scale Hypertextual Web Search Engine
Optional Barabasi, The Scale-Free Property

Week 13: Recommender systems
- Motivation
- The Netflix challenge
- Content-based methods
- Learning features and parameters
- Nearest-neighbor CF

Optional Gomez and Hunt, The Netflix Recommender System: Algorithms, Business Value, and Innovation
Optional Koren, Collaborative Filtering with Temporal Dynamics

Week 14: Wrap-Up
- Topics beyond the scope of this course
- What your instructors do