Artificial Intelligence

Instructor: Jie Shen

Dept. of Computer Science

Sept. 6, 2022

AI applications...

- auto driving
- robotics
- deepfake
- ...

But...

Theoretical Foundation of AI

About the Course

Tough

- Not for introductory purpose
- Research oriented
  - Emphasize on theoretical aspects
  - Analyze computational cost
  - Understand statistical accuracy
- Strong background in calculus, linear algebra, and probability
  - If cannot do Quiz 0, consider dropping the course

Syllabus

- Review of calculus, probability, linear algebra
- random projection
- singular value decomposition, principal component analysis
- k-means clustering, subspace clustering
- dictionary learning and sparse coding
- low-rank matrix estimation, with applications to recommender systems
- computational social science
- robust mean estimation, robust classification
- algorithmic fairness

Course Staff

Instructor: Jie Shen (jie.shen@stevens.edu)

TA:

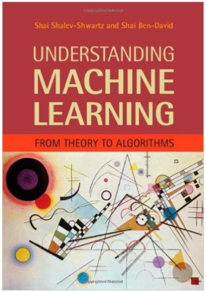
- Section A: Shiwei Zeng (szeng4@stevens.edu)
- Section B: Nan Cui (ncui@stevens.edu)

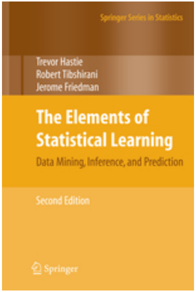
CA: may need 1 for each section (TBD after drop/add period)

Office Hours: TBD

Textbook & Reference

- No Required Textbook
- Recommended (available online)





- Research Papers in NeurIPS, ICML, COLT

Grading

- Midterm Exam (50%)
  - open book
  - electronic device does not help
  - **tentatively on Oct 25**
- Final Exam (50%)
  - closed-book, on Dec. 13
- Final Grade

90 - 100	85 - 89	80 - 84	75 - 79	70 - 74	<70
A	A-	B+	B	B-	Fail

Quiz 0 (20 min)

1. Let $\boldsymbol{x} = (1\ 2\ 3)$, $\boldsymbol{y} = (1\ 1\ 1)$. Calculate $\boldsymbol{x}\boldsymbol{y}^\top$ and $\boldsymbol{x}^\top\boldsymbol{y}$.

2. Show that $\frac{1}{2}(e^x + e^{-x}) \leq e^{x^2/2}$ for all $x \in \mathbb{R}$, where e is the base of the natural logarithm. [Hint: Taylor expansion]

3. Show that for all $x > 0$, $\log(1 + x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}$.

Linear Algebra Overview

A d-dimensional column vector $\boldsymbol{x}$ is a set of d numbers

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

- Bold lowercase letters for vectors
- Almost all the data is vector

"All the News That's Fit to Print"

The New York Times

LATE CITY EDITION

10 CENTS

MEN WALK ON MOON

ASTRONAUTS LAND ON PLAIN; COLLECT ROCKS, PLANT FLAG

Voice From Moon: 'Eagle Has Landed'

A Powdery Surface Is Closely Explored

Vector Operations

Suppose $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ are column vectors, $a, b \in \mathbb{R}$

- $\boldsymbol{x}^\top \stackrel{\text{def}}{=} \begin{pmatrix} x_1 & x_2 & \dots & x_d \end{pmatrix}$
- $a\boldsymbol{x} \stackrel{\text{def}}{=} \begin{pmatrix} ax_1 & ax_2 & \dots & ax_d \end{pmatrix}^\top$
- $\boldsymbol{x} + \boldsymbol{y} \stackrel{\text{def}}{=} \begin{pmatrix} x_1 + y_1 & x_2 + y_2 & \dots & x_d + y_d \end{pmatrix}$
- $a\boldsymbol{x} + b\boldsymbol{y}$
- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \stackrel{\text{def}}{=} \sum_{i=1}^d x_i y_i \in \mathbb{R}$
  - Sometimes use $\boldsymbol{x}^\top \boldsymbol{y}$, $\boldsymbol{x} \cdot \boldsymbol{y}$

Vector Norms

- $\|\boldsymbol{x}\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{d} x_i^2}$
    - Broadly used
    - $\|\boldsymbol{x} - \boldsymbol{y}\|_2$
- $\|\boldsymbol{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{d} |x_i|$
- $\|\boldsymbol{x}\|_\infty \stackrel{\text{def}}{=} \max_{1 \le i \le d} |x_i|$

Matrix

Vector: a set of numbers
Matrix: a set of vectors

- Bold capital letters $\boldsymbol{X} \in \mathbb{R}^{d \times n}$
- $\boldsymbol{X} = (x_{ij})_{1 \le i \le d, 1 \le j \le n} = (\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \ldots \ \boldsymbol{x}_n)$
- $a\boldsymbol{X}$ for $a \in \mathbb{R}$
- $a\boldsymbol{X} + b\boldsymbol{Y}$ when $\boldsymbol{X}, \boldsymbol{Y}$ have the same size
- Multiplication: $\boldsymbol{X} \in \mathbb{R}^{d \times n}$, $\boldsymbol{Y} \in \mathbb{R}^{p \times m}$
  - Can do $\boldsymbol{X}\boldsymbol{Y}$ only when $n = p$
  - $\boldsymbol{X}\boldsymbol{Y} \in \mathbb{R}^{d \times m}$
  - For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{x}^\top \boldsymbol{y} \in \mathbb{R}$, $\boldsymbol{x}\boldsymbol{y}^\top \in \mathbb{R}^{d \times d}$
- Transpose
- Symmetric matrix, diagonal
- Inverse of a square matrix

Probability Overview

Probability: measure of likelihood that an event will occur.

- From 0 to 1
- Coin tossing (heads or tails)

- Random variable X
- Events $= \{0, 1\}$
- X has distribution $\mathcal{D}$

Probability Overview

- If X is discrete, probability mass function $p(x) = P(X = x)$
  - Takes value from a countable set
  - $\{0, 1\}$
  - $\{1, 2, 3, \dots\}$
- If X is continuous, probability density function (PDF) $p(x)$

$$P(X \leq x) = \int_{-\infty}^{x} p(z)dz$$

  - Uniform distribution
  - Normal distribution
- $P(X \leq x)$: cumulative density function

Expected Value

Expected value

- Discrete: $\mathbb{E}[X] = \sum xp(x)$
- Continuous: $\mathbb{E}[X] = \int xp(x)dx$
- Practice: Play a game for money. Each time

$$\Pr(X = 1) = 0.6, \quad \Pr(X = -1) = 0.4.$$

When can we win 100 dollars?

Expectation

- Average of multiple outcomes
- Not quite useful in practice
  - gambling
  - weather forecasting (rainy, sunny, dry)
  - in expectation = I guess
- But, $\mathbb{E}[X]$ implies $P(X)$

Markov's Inequality

Theorem. If $X > 0$, $P(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ for all $t > 0$.

- Proof of correctness
- Proof of tightness
- Negative random variables
    - moment-generating function

Variance

- $\mathrm{Var}(X) \overset{\text{def}}{=} \mathbb{E}[X - \mathbb{E}X]^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$
- $X_1, X_2, \ldots X_n$ are independent, then $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i)$

Chebyshev's Inequality

Hoeffding's Inequality

Symmetric Bernoulli distribution: $P(X = 1) = P(X = -1) = 1/2$

Theorem. Let $X_1, X_2, \ldots, X_n$ be independent symmetric Bernoulli random variables. Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Then, for any $t \geq 0$,

$$P\left(\sum_{i=1}^{n} a_i X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\left\|\boldsymbol{a}\right\|_2^2}\right)$$

- Proof of correctness
- Generalize to non-symmetric distribution

$$P(X = 1) = p \in [0, 1], \quad P(X = -1) = 1 - p$$