IOWA STATE UNIVERSITY
Department of Computer Science

COM S 573: Machine Learning

Hongyang Gao
CS@ISU

Course Information

- Instructor: Hongyang Gao, Assistant Professor
  - Research on machine learning and deep learning
  - Office: Atanasoff Hall 207
  - https://faculty.sites.iastate.edu/hygao/
  - E-mail: hygao@iastate.edu
- Lecture
  - Meeting Time: TT 3:40 PM - 4:55 PM
- Logistics
  - Class materials are distributed via Canvas

Teaching Assistant

- Zhaoning Yu
- Email: znyu@iastate.edu
- Office Hours: Friday 3:00 PM to 4:00 PM

- Siyuan Sun
- Email: sxs14473@iastate.edu
- Office Hours: Friday 4:00 PM to 5:00 PM
- Office Location: on Canvas

About this class

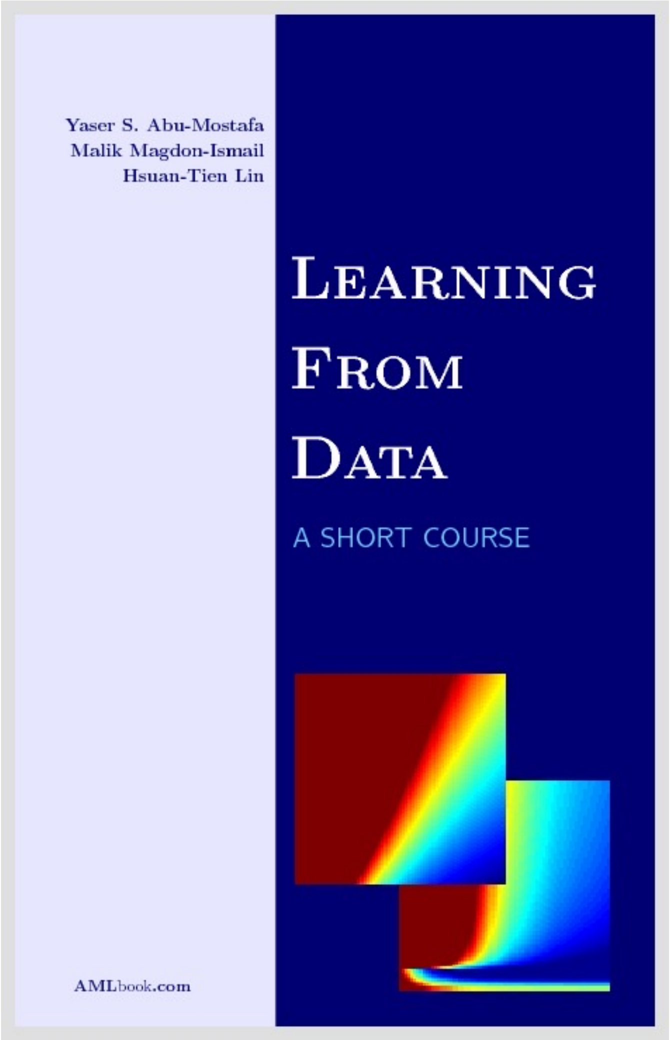
- We are encouraged to wear mask
- No audit is allowed
- **Extra credit** will be award for attending the class
- Will remain in a face-to-face way

Prerequisites etc.

- ***Linear algebra*** (vectors, matrices, matrix-vector computations, vector and matrix norms, linear independence …)
- Multivariate calculus (derivatives of univariate functions, derivatives of multivariate functions, chain rule, Taylor expansion)
- Basic probability and statistics (discrete and continuous probability distributions, sum rule, product rule, marginal probability distributions, conditional probability distributions, joint probability distributions, independence and conditional independence, Bayes Theorem, variance and covariance, expectation)
- ***Python programming***, data and sample code will be provided
- Parts of the class are technical; some level of math is required

Required Textbook

- Learning from Data, by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien, Lin, AMLBook
- Chapters 1-5: http://www.amazon.com/gp/product/1600490069
- e-Chapters (6-9): http://amlbook.com/



COM S 573 Roadmap

Week	Topic
1	Introduction, Linear Regression
2	Linear Perceptron, Logistic Regression
3	SVM, Gradient Descent, Batch Training
4	Overfitting and Regularization
5	Boosting & Ensemble Learning
6	Decision Trees
7	Clustering, Expectation Maximization
8	Dimensional Reduction, PCA
9,10,11	Neural network, Convolution, Pooling, Attention
12-13	DNN Architectures
14	Project Presentation

Assignments and exams

- Assignments (5 × 7% = 35%)
  - A written component + a programming component
  - Homework submission should include a report and code.
  - Homework requires Python programming
    - Data and skeleton code will be provided in Python format. **No code instructions during class!**
- Three midterm exams (45%, 15% each)
- Project (20%)
- **Extra credit** for attendance
- **NO FINAL**

Project (20%)

- There will be one semester-long, team project. Students are required to form teams of 2-3 students and carry out a project related to machine learning and applications to computer vision or natural language processing.
- Example projects include (1) implementation and comparison of several existing methods on benchmark data sets and gain some insights, (2) extension of existing methods by incorporating more functions and features, (3) improvements to current models and algorithms with experimental evaluation.
- The minimal requirement is that each project must have an experimental section with results.

More on project

Project milestones are as follows:

- Project proposal: Each team is required to discuss their project proposal with the instructor and submit a one-page proposal.
- Mid-term report: Each team is required to submit a mid-term report of 3 pages to report the results of project. Preliminary experimental results and plan for the remaining parts of the project are required.

More on project

- Final report and presentation: At the end of semester, each team will be asked to submit a report (minimum of 6 pages excluding references) and do a presentation on their research project. The presentation should be done by all team members and the contributions of each team member should be made clear in both presentation and in report.
- Example projects from Stanford Deep Learning classes can be found at: http://cs231n.stanford.edu/project.html
- https://www.kaggle.com/datasets

Use Latex

- Please use **Latex** for your HWs and Project reports
- https://www.overleaf.com/learn/latex/Tutorials
- **20%** will be deducted if not using latex

Grading based on absolute percentage

Name:	Range:	
A	100 %	to 93.0%
A-	< 93.0 %	to 90.0%
B+	< 90.0 %	to 87.0%
B	< 87.0 %	to 83.0%
B-	< 83.0 %	to 80.0%
C+	< 80.0 %	to 77.0%
C	< 77.0 %	to 73.0%
C-	< 73.0 %	to 70.0%
D+	< 70.0 %	to 67.0%
D	< 67.0 %	to 63.0%
D-	< 63.0 %	to 60.0%
F	< 60.0 %	to 0.0%

Attendance and other rules

- All absences will be handled according to Iowa State Academic Conduct: https://catalog.iastate.edu/academic_conduct/. It is your responsibility to keep up with the class, even when unexpected events interfere.
- Missed exams will only be rescheduled for university excused absences. Note that if advanced notice is not feasible, you have 2 business days to provide notification. A zero will be assigned for exams due to an unexcused absence. Documentation must be submitted prior to making up a missed exam.

A few more remarks

- Slides are used for showing some pictures and save time
- Attend lectures, ask/answer questions
- Machine learning is an important field, a key field of AI
- This class will take a foundational approach, and tries to understand all concepts in great details
- All HW and exam questions will be based on **understanding, not memorization**
- HWs will also focus on hands-on programming and implementation

Introduce Yourself

- Name
- Major (CS, CE, …)
- Program (PhD, Master, …)
- Advisor
- Research Area

Linear Algebra Review

- Basic Notations:
  - By $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ we denote a **matrix** with *m* rows and *n* columns
  - By $\boldsymbol{x} \in \mathbb{R}^{n}$, we denote a **vector** with *n* entries.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Linear Algebra Review

A   m x n
B   n x p
A·B   m·n·p
m x p

- Matrix Multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is the matrix

$$C = AB \in \mathbb{R}^{m \times p},$$

where

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

- Dot Product

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$

$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

n x 1   n x 1   1xn · nx1 = 1

Linear Algebra Review

- The Transpose

$$(A^T)_{ij} = A_{ji}$$

$$(A^T)^T = A$$

$$(AB)^T = B^T A^T$$

$$(A+B)^T = A^T + B^T$$

- Norms

$$\|x\|_2 = \ ?$$

Linear Algebra Review

$\|x\|_2^2 = x^T x$

- The Transpose

x^T $\quad$ x

$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$(A^T)_{ij} = A_{ji}$$

$$(A^T)^T = A$$
$$(AB)^T = B^T A^T$$
$$(A+B)^T = A^T + B^T$$

$x_1^2 + x_2^2 + \cdots + x_n^2$

- Norms

A ***norm*** of a vector $\|x\|$ is informally a measure of the "length" of the vector. For example, we have the commonly-used Euclidean or ℓ_2 norm,

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

$= \quad = \sum_{i=1}^{n} x_i^2$

Note that $\|x\|_2^2 = x^T x$.

Linear Algebra Review

- The derivatives of vector functions

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors of orders n and m respectively:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \tag{D.1}$$

where each component y_i may be a function of all the x_j, a fact represented by saying that $\mathbf{y}$ is a function of $\mathbf{x}$, or

$$\mathbf{y} = \mathbf{y}(\mathbf{x}). \tag{D.2}$$

If $n = 1$, $\mathbf{x}$ reduces to a scalar, which we call x. If $m = 1$, $\mathbf{y}$ reduces to a scalar, which we call y. Various applications are studied in the following subsections.

Linear Algebra Review

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \longrightarrow y$

- The derivatives of vector functions

The derivative of the vector $\mathbf{y}$ with respect to vector $\mathbf{x}$ is the $n \times m$ matrix

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{D.3}$$

$\frac{dy}{dx}$

If y is a scalar,

$$\frac{\partial y}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}. \tag{D.4}$$

Linear Algebra Review

- The derivatives of vector functions
- The following table collects several useful vector derivative formulas.

$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{Ax} + \mathbf{A}^T\mathbf{x}$

$x_1^2 + x_2^2 + x_3^2$

$\frac{\partial y}{x_1} = 2x_1$

$\frac{\partial y}{x_2} = 2x_2$

$2\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2x$