CX 4240 Spring 2026

Course Information

- **Instructor**: Bo Dai (bodai@cc.gatech.edu)
- **Course Prefix and Number**: CSE 4240 A
- **Term**: Spring 2026

Course Description

This is a preliminary course for machine learning techniques, which extracts useful knowledge from data in real-world applications. On the technique side, we will cover key machine learning methods (supervised learning, representation learning, generative models, and foundation models). On the application side, it will introduce various applications of these techniques, including images/text generation and robotics. It will introduce how to formulate real-world tasks as data analysis problems, key methods for solving these problems, and their advantages and disadvantages.

Target Audience:

- **Practitioners:** This course is beneficial for individuals seeking to apply machine learning and data science techniques to solve real-world challenges in their professional or personal lives.
- **Researchers:** This course provides valuable foundational knowledge and advanced techniques for students aspiring to conduct cutting-edge research in areas such as data mining, machine learning, natural language processing, and other related fields.

Prerequisites:

- solid knowledge of probability, statistics, calculus, and linear algebra;
- basic knowledge of machine learning;
- solid programming skills, preferably in Python.

These topics will be covered in **Four Modules**:

- **Module I: Background Knowledge**
  - Linear algebra
  - Probability and statistics
  - Optimization
- **Module II: Supervised Learning**
  - Linear Regression and Classification
    - Ridge regression, logistic regression, naive Bayes
  - Neural Network
    - CNN, RNN
- **Module III: Unsupervised Learning**
  - Clustering
    - K-means, Gaussian Mixture Models
  - Dimension Reduction and Representation Learning
    - PCA, SimCLR,
  - Generative Models.
    - VAE
- **Module IV: LLM**
  - Attention, Transformer
  - Supervised Fine-tuning
  - RLHF

Course Learning Outcomes

The learning objective is that by the end of this course, the students are able to formulate their real-world problems at hand, choose appropriate machine learning methods to acquire insights from data automatically, and even come up with innovative solutions for solving open problems in this field.

Required Course Materials

The required textbook for this class is (note that the material of the class goes beyond this book):

- Bishop. Pattern Recognition and Machine Learning. Springer. 2006
- Wainwright & Jordan. Graphical Models, Exponential Families, and Variational Inference. Now Publishers, Inc. 2008
- Deep Learning, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Grading Policy

The grading breakdown is as follows:

- **Homework (30%)**: There will be three assignments, each account for 10% towards your final score. Each assignment includes written analysis and/or programming for testing your understanding of the taught content.
  - Late policy: Assignments are due at 11:59PM of the due date. You will be allowed 2 total late days (48 hours) without penalty for the entire semester (for homework only, not applicable to exams or projects). Once those days are used, you will be penalized according to the following policy:
    - Homework is worth full credit before the due time.
    - It is worth 75% credit for the next 24 hours.
    - It is worth 50% credit for the second next 24 hours.
    - It is worth zero credit after that.
  - Follow the Georgia Tech Academic Honor Code.
  - Regrade Requests: If you believe that the course staff made an objective error in grading, you may submit a regrade request on Gradescope within **3 days** of the grade release. Your request should briefly summarize why the original grading was incorrect. Note that staff may regrade the entire submission, so it is possible for you to lose more points than you gain if a mistake was overlooked in the first time.

  You are required to use Latex (OverLeaf Latex Example in the Video), or a word processing software to generate your solutions to the written questions.

Handwritten solutions WILL NOT BE ACCEPTED. You can easily export your Jupyter Notebook to a Python file and import that to your desired python IDE to debug your code for assignments.

- **Project (40%)**:
  - Team Size: Each project must be completed in a team of 3-5. Once you have formed your group, please send one email per team to the class instructor list with the names of all team members. If you have trouble forming a group, please send us an email and we will help you find project partners.

    The team formation email will be due at **11:59 PM on Feb 16th**.

  - Projects Topics:
    - reproduce classic papers, include but not limited to:
      - Deep Residual Learning for Image Recognition
      - Auto-Encoding Variational Bayes.
      - A Simple Framework for Contrastive Learning of Visual Representations.
      - Sequence to Sequence Learning with Neural Networks
      - etc
    - You may also refer to the Stanford Project Examples.

  - 2 Deliverables:
    - Presentation (15%)
    - Final Report (25%): *All write-ups should use the NeurIPS style.*

      *Your final report is expected to be **5 pages excluding references**. It should have roughly the following format:*

      - *Introduction: problem definition and motivation*
      - *Background & Related Work: background info and literature survey*
      - *Methods – Overview of your proposed method – Intuition on why should it be better than the state of the art – Details of models and algorithms that you developed*
      - *Experiments – Description of your testbed and a list of questions your experiments are designed to answer – Details of the experiments and results*

- *Conclusion: discussion and future work*

The project final report will be due at **11:59 PM on May 4th**

- Criteria:
  - 30% for proposed method (soundness and originality)
  - 30% for correctness, completeness, and difficulty of experiments and figures
  - 20% for empirical and theoretical analysis of results and methods
  - 20% for quality of writing (clarity, organization, flow, etc.)

- Computing Resources: Google Colaboratory allows free access to run Jupyter Notebooks using GPU resources. The Google Cloud Platform and AWS Educate are also good resources. The GitHub Student Developer Pack also offers free Microsoft Azure and Digital Ocean credits. This semester, we are also offering PACE ICE, Georgia Tech's in-home cluster to students.

- **Exam (30%) :** One exam will be held on March 18 in lieu of the regular class:
  - It will be a closed-book exam, so no notes or communication with peers is allowed.
  - There will be no make-up exams, so be sure to attend on the scheduled date. Missing the exam will result in zero credit.

- **Participation (5% extra credit)**:We appreciate student participation in the class! We will be awarding, on a case-by-case basis, up to 5% in extra credit to the top Ed contributors based on the number of (meaningful) instructor-endorsed answers or other significant contributions that assist the teaching staff or other students in the course. The most helpful

contributor will receive the greatest amount of extra credit, and other students with significant contributions will receive a percentage of that.

Attendance Policy

All lectures will be in-person. The materials will be posted after class.

Attendance is not required but expected. The lectures are a chance to directly interact with your instructor and classmates.

Academic and Research Honesty/Integrity Statement

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. Review the Student Code of Conduct and the Academic Honor Code, especially Appendix A: Graduate Addendum to the Academic Honor Code.

Students are expected to perform research in an ethical and responsible manner. All Doctoral and Master's Thesis students are required to take the Responsible Conduct of Research training, and it is expected that students abide by the principles taught in that training while performing research for this thesis course.

Allegations of scientific or scholarly misconduct are handled in accordance with the procedures outlined by the Policy for Responding to Allegations of Scientific or Other Scholarly Misconduct.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services as soon as possible to make an appointment to discuss your special needs and to obtain an accommodations

letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.